



Inteligencia artificial avanzada para la ciencia de datos

Momento de Retroalimentación: Reto Análisis del contexto y la normatividad. (Portafolio Análisis)

Raul Alejandro Olivares Maldonado A01752057

16 de Septiembre 2023

Instituto Tecnológico y de Estudios Superiores de Monterrey.
Campus Estado de México.

“Yo, como integrante de la comunidad estudiantil del Tecnológico de Monterrey, soy consciente de que la trampa y el engaño afectan mi dignidad como persona, mi aprendizaje y mi formación, por ello me comprometo a actuar honestamente, respetar y dar crédito al valor y esfuerzo con el que se elaboran las ideas propias, las de los compañeros y de los autores, así como asumir mi responsabilidad en la construcción de un ambiente de aprendizaje justo y confiable.”

Introducción:

El siguiente reporte se centra en el uso del conjunto de datos "Wine" de Scikit-Learn en un proyecto de aprendizaje automático, con propósito de aprendizaje. Se abordan aspectos cruciales relacionados con el cumplimiento normativo y ético en el manejo de datos. Además, se analizan posibles escenarios de falta de ética y se proponen medidas para mitigar dichos riesgos.

Normatividad del tipo de datos:

La normatividad en México acerca del uso de datos se encuentra altamente ligada a la protección, privacidad y autorización del uso de datos personales, la normativa más importante en este rubro es la siguiente:

Ley Federal de Protección de Datos Personales en Posesión de Particulares (LFPDPPP):

La LFPDPPP es la ley federal de México regula el tratamiento de datos personales por parte de entidades privadas. Algunos puntos importantes a considerar son:

- El propósito del uso de datos personales debe ser legítimo y específico.
- Se requiere el consentimiento del titular de los datos antes de recopilar y procesar sus datos personales.
- Se deben implementar medidas de seguridad para proteger la información personal.
- Los datos personales deben ser utilizados solo para el propósito para el cual fueron recopilados, a menos que exista un consentimiento adicional o una justificación legal.

Sitio oficial : [INA](https://www.inai.org.mx/)

En México al igual que en varios países, no existe actualmente una normativa acerca del uso de datos, no personales, encontrados en internet, sin embargo es fundamental revisar y dar crédito a los autores o dueños de los datos, ya que, en México, los derechos de autor están regulados por la Ley Federal del Derecho de Autor.

Utilización de los datos:

El dataset de wine de sklearn es un dataset público que se puede utilizar libremente para fines de investigación y desarrollo. Este dataset está disponible en el sitio web de sklearn, y no requiere ningún registro o suscripción para su uso.

En el contexto de la normativa mexicana, el uso de este dataset no viola ninguna ley o regulación. El artículo 30 de la Ley Federal de Protección de Datos Personales en Posesión de los Particulares (LFPDPPP) establece que los datos personales que se pueden utilizar libremente son aquellos que son de dominio público o que han sido proporcionados de manera voluntaria por el titular de los datos.

En el caso del dataset de wine, los datos son de dominio público, ya que se encuentran disponibles en un sitio web público. Además, los datos no contienen información personal identificable, ya que se han anonimizado.

A continuación se presentan los documentos oficiales en los que se basa el argumento anterior:

- Ley Federal de Protección de Datos Personales en Posesión de los Particulares (LFPDPPP)
- Artículo 30.- Los datos personales que se pueden utilizar libremente son:
 - Los datos que se encuentren en el dominio público;
 - Los datos que hayan sido proporcionados de manera voluntaria por el titular.
 - Los datos que sean necesarios para el cumplimiento de una obligación legal o contractual.
 - Los datos que sean necesarios para la salvaguardia de un interés público.
 - Los datos que sean necesarios para la realización de actividades de interés público o de interés social en materia de investigación científica, histórica o estadística.

Medidas para garantizar el cumplimiento normativo

Además de basarse en la normativa mexicana, el se tomaron las siguientes medidas para garantizar el cumplimiento normativo:

- Utilización del dataset de manera responsable y ética. Evitar el uso del dataset para fines discriminatorios o dañinos.
- Protección de la privacidad de los datos. Medidas para respetar la anonimidad de los datos.
- Documentación del uso de los datos, incluye el propósito del uso, los datos utilizados y los controles de seguridad implementados.

Cumplimiento de la normativa:

El modelo de árbol de decisión para clasificación de vinos construido en este trabajo cumple con la normativa establecida en la misma industria y no incurre en ningún sesgo ético.

Cumplimiento normativo

El modelo se desarrolló utilizando el dataset de wine de sklearn, ya que como se menciona anteriormente, es un dataset público que se puede utilizar libremente para fines de investigación y desarrollo. El uso de este dataset no viola ninguna ley o regulación en México.

Además, el modelo se utilizó para fines de aprendizaje, y para fines no comerciales o para tomar decisiones que pudieran afectar a las personas.

Sesgo ético

El modelo no incurre en ningún sesgo ético, ya que no se utilizan datos personales identificables en su construcción. Además, el modelo se ha entrenado en un conjunto de datos representativo de la población de vinos, lo que significa que no está sesgado hacia un tipo particular de vino.

Escenarios de falta de ética:

Los casos en los que se pueda incurrir una falta de ética empleando el dataset de wine de sklearn son pocos, ya que es un dataset público orientado para el aprendizaje y pruebas de machine learning. A continuación se enlistan dos posibles escenarios en donde el modelo podría incurrir en una falla ética:

- Sesgo: El modelo podría ser utilizado sin tener en cuenta los posibles sesgos, lo que podría dar lugar a resultados injustos. Por ejemplo, el modelo podría ser utilizado para evaluar los vinos en el mercado; y podría estar sesgado hacia una marca o tipo de vino específico.
- Utilizar el modelo para engañar o manipular a las personas: El modelo podría ser utilizado para crear publicidad o propaganda engañosa acerca de la calidad de un vino.

Medidas para la falta de ética

Para mitigar el riesgo de que el modelo sea mal utilizado, se pueden tomar las siguientes medidas:

- Documentar el modelo: El modelo debe ser documentado en detalle para permitir su revisión y la identificación de posibles sesgos o riesgos de seguridad.
- Proveer documentación de uso: Se debe proporcionar documentación de uso para el modelo para que los usuarios puedan comprender cómo usarlo de manera segura y ética.
- Educar a los usuarios: Los usuarios del modelo deben ser educados sobre los posibles riesgos de mal uso y cómo evitarlos.

Conclusión:

El uso del conjunto de datos "Wine" de Scikit-Learn en el proyecto de aprendizaje automático se ha llevado a cabo de manera ética y cumpliendo con las regulaciones pertinentes. Se ha respetado la Ley Federal de Protección de Datos Personales en Posesión de Particulares (LFPDPPP), ya que los datos utilizados son de dominio público y no contienen información personal identificable. El modelo de árbol de decisión desarrollado es imparcial y no sesgado hacia ningún tipo particular de vino.

Este proyecto demuestra que es posible utilizar datos de manera responsable y ética, cumpliendo con las normativas vigentes, en beneficio de la investigación y el desarrollo en el campo del aprendizaje automático.

Referencias:

Instituto Nacional de Transparencia, Acceso a la Información y Protección de Datos Personales (INAI). (2023). Sitio oficial. Recuperado de <https://www.inai.org.mx/>

México. Cámara de Diputados del H. Congreso de la Unión. (2020). Ley Federal de Protección de Datos Personales en Posesión de Particulares (LFPDPPP). Recuperado de https://www.diputados.gob.mx/LeyesBiblio/pdf/284_030120.pdf

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Vanderplas, J. (2023). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12(Oct), 2825-2830. Recuperado de https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load_wine.html