



**Instituto Tecnológico y de Estudios  
Superiores de Monterrey**  
Campus Monterrey

**Inteligencia artificial avanzada para la ciencia de datos I**  
TC3006C.101

**Análisis del contexto y normatividad**



Arturo Garza Campuzano A00828096

28 ago 2023

## Proyecto seleccionado

El proyecto seleccionado sobre el que se realiza este análisis de contexto y normatividad es **Implementación de una técnica de aprendizaje máquina sin el uso de framework**, el cual se encuentra disponible en el siguiente enlace: [ImplementacionTecnica\\_ML.ipynb](#). Este proyecto es una implementación de un modelo de **regresión logística** sin el uso del framework.

## Normativa asociada al tipo de datos

En el proyecto se utilizó un conjunto de datos de pacientes del sexo femenino, de al menos veintiun años, de ascendencia india prima, con el fin de implementar un modelo clasificador de **regresión logística** sin el uso de framework para predecir si un paciente tiene diabetes o no. Estos datos provienen originalmente del *National Institute of Diabetes and Digestive and Kidney Diseases* (**Akturk, 2020**), ubicado en Estados Unidos. Una normativa asociada al tipo de datos que se usaron es la **Ley de Portabilidad y Responsabilidad del Seguro de Salud (HIPAA)**, la cual fue “creada para proteger a millones de trabajadores y miembros de sus familias en los Estados Unidos que padecen alguna afección médica” (**American Cancer Society, 2019**). En el año 2002, las cláusulas de esta ley se ampliaron para ofrecer a los pacientes un mayor acceso a sus propios expedientes médicos y más control sobre la forma en cómo se usa la información sobre su salud. La ley estipula que la información de salud no se puede compartir sin el permiso escrito del paciente y exige que los proveedores de atención médica y los planes de salud también protejan la privacidad del paciente.

¿Y esta normativa asociada aplica para el conjunto de datos utilizados (*Diabetes Dataset*)? Se sabe que los datos utilizados fueron recibidos el 9 de mayo de 1990 (**Akturk, 2020**). Considerando esta fecha como fecha de publicación, aunque las cláusulas de la HIPAA se promulgaron después de la fecha, se espera que las entidades de atención médica y aquellos que manejen datos de salud los protejan de acuerdo con las regulaciones de la HIPAA. Por lo tanto, esta normativa asociada sí aplica para el conjunto de datos utilizados.

## Uso de los datos y cumplimiento normativo

Los datos se utilizaron por medio de un Jupiter Notebook de Google Colab y, para cumplir con la normativa asociada, se siguieron las siguientes prácticas:

- **Seguridad de datos:** los datos se protegieron restringiendo el acceso al proyecto sólo al personal del Tecnológico de Monterrey. Esto para asegurar que los datos sólo se utilicen para fines educativos.
- **Origen de los datos:** antes de utilizar los datos se investigó acerca del origen de los mismos. Resulta que el dueño original de estos datos era el

*National Institute of Diabetes and Digestive and Kidney Diseases* y el responsable de su donación fue Vincent Sigillito. Los datos se recibieron el 13 de mayo de 1990 y se han utilizado en otros trabajos como en el artículo de Smith et al. (1988), que presenta el uso del algoritmo de aprendizaje ADAP para predecir el inicio de la diabetes mellitus (**Akturk, 2020**). Esta práctica cumple con el artículo 7 de la ley HIPAA donde *“la obtención de datos personales no debe de hacerse a través de medios engañosos o fraudulentos”*.

- **Aviso de privacidad:** en el proyecto se establece y se cumple que el dataset utilizado sólo es usado bajo fines académicos, lo cual sigue con lo establecido en el artículo 13 de la ley HIPAA donde *“el tratamiento de datos personales será el que resulte necesario, adecuado y relevante en relación con las finalidades previstas en el aviso de privacidad”*.

Bajo estas prácticas y la información obtenida del origen de los datos se puede asegurar que la normativa asociada se cumple.

## Sesgo ético de la herramienta

El **sesgo ético de la herramienta** se refiere a la presencia de sesgos injustos, discriminación o resultados éticamente cuestionables en las decisiones o predicciones realizadas por una herramienta de IA. En este caso, la herramienta utilizada fue la **regresión logística**, modelo estadístico que utiliza una función sigmoide para clasificar. Bajo las métricas de desempeño del modelo se puede observar que las decisiones tomadas por el modelo no fueron lo suficientemente justas, por lo que sus resultados son éticamente cuestionables. En otras palabras, el modelo no se puede implementar para predecir si un paciente tiene diabetes o no.

Considerando este alto sesgo en cuanto a las decisiones tomadas por el modelo, ¿se cumple la normativa asociada? Sí, se cumple porque en el mismo proyecto se establece que el resultado del modelo no es óptimo para su implementación, advirtiendo que no serviría para hacer las predicciones solicitadas. También hay que recordar que el modelo fue realizado bajo fines académicos, lo cual implica que los datos son sólo utilizados para propósitos de estudio, no de su mera implementación.

## Escenarios de falta ética y mal uso

Suponiendo que el conjunto de datos proporcionara información más personal, no se especificara bien el origen de estos o no se haya usado previamente para fines legítimos, se abriría la posibilidad de presentarse escenarios de falta de ética y mal uso como:

- **Violación de la privacidad:** si los datos se utilizan de manera que sea posible identificar a pacientes individuales y se divulgan públicamente, se estaría violando la privacidad de los pacientes y sus datos personales.

- Comercialización no ética: si los datos se utilizan con fines comerciales sin el consentimiento adecuado de los pacientes, podría considerarse explotación de su información personal para beneficio económico.
- Falta de transparencia: si se utiliza un modelo de aprendizaje automático sin proporcionar información adecuada a los pacientes sobre cómo se tomarían las decisiones basadas en sus datos, podría erosionarse la confianza del sistema médico.

## Referencias bibliográficas

1. Mehmet Akturk. "Diabetes Dataset." *Kaggle.com*, 2020, [www.kaggle.com/datasets/mathchi/diabetes-data-set](https://www.kaggle.com/datasets/mathchi/diabetes-data-set).
2. American Cancer Society. "Ley de Transferencia Y Responsabilidad de Seguro Médico (HIPAA) de 1996." *Www.cancer.org*, 13 May 2019, [www.cancer.org/es/cancer/asuntos-financieros-y-cobertura/leyes-de-los-seguros-medicos/que-es-hipaa.html#:~:text=La%20Ley%20de%20Transferencia%20y](https://www.cancer.org/es/cancer/asuntos-financieros-y-cobertura/leyes-de-los-seguros-medicos/que-es-hipaa.html#:~:text=La%20Ley%20de%20Transferencia%20y). Accessed 9 Sept. 2023.