



**Tecnológico  
de Monterrey**

**Momento de Retroalimentación: Análisis de contexto y la  
normatividad**

Profesor:

Alfredo Esquivel Jaramillo

Alan Alberto Mota Yescas

A01753924

Martes 10 de septiembre del 2024

Grupo 102

Instituto Tecnológico y de Estudios Superiores de Monterrey

**Inteligencia artificial avanzada para la ciencia de datos**

---

## Normativa Asociada al Tipo de Datos Utilizados

El dataset utilizado en este proyecto, el Pima Indians Diabetes Database, contiene datos clínicos sensibles, como niveles de glucosa, presión arterial, y antecedentes médicos de pacientes. Estos datos caen bajo las categorías de datos personales y datos de salud, que son regulados en muchos países por normativas específicas para la protección de la privacidad y seguridad.

1. Reglamento General de Protección de Datos (GDPR): En la Unión Europea, la GDPR regula el uso de datos personales, particularmente aquellos que son sensibles, como los datos médicos. Establece que se deben obtener los consentimientos adecuados y que se deben aplicar medidas estrictas para proteger los datos personales de los individuos.
2. Ley de Portabilidad y Responsabilidad del Seguro de Salud (HIPAA): En los Estados Unidos, HIPAA regula la protección de información médica. El uso de datos de salud sin anonimización o consentimiento explícito viola esta normativa.

## Medios Utilizados para Garantizar el Cumplimiento de la Normativa

Medios de Uso de los Datos: Los datos fueron obtenidos de Kaggle, una plataforma que proporciona datasets para uso académico y de investigación. En este contexto, Kaggle asegura que los datos han sido anonimizados para evitar la identificación de pacientes individuales. Según la fuente oficial de los datos, estos están disponibles únicamente con fines de investigación, y no contienen información que permita identificar a individuos específicos.

Medidas para No Violentar la Normativa:

1. Anonimización de los Datos: Todos los identificadores personales han sido removidos del dataset, por lo que se garantiza la anonimización completa de los sujetos. No se maneja información directamente identificable (como nombres o direcciones) que pudiera requerir un consentimiento explícito bajo la normativa GDPR o HIPAA.
2. Uso para Fines Académicos: Los datos se utilizan estrictamente para fines de investigación y desarrollo de modelos de machine learning, alineados con el propósito original de la recopilación de los mismos.

Referencias: GDPR (Artículos 6 y 9) sobre el procesamiento de datos sensibles y el consentimiento. HIPAA, que regula la privacidad de los datos médicos.

## Cumplimiento de la Normativa en la Solución

La solución implementada sigue los siguientes puntos para asegurar el cumplimiento de la normativa y evitar sesgos éticos:

1. No Introducción de Datos Personales: El modelo no requiere ni maneja datos identificables de individuos. Además, se garantiza la anonimización de los datos mediante el preprocesamiento inicial.
2. Justificación Clínica y Sin Sesgos: Las características utilizadas están basadas en evidencia médica relevante y probada para la detección de diabetes, minimizando cualquier posible introducción de sesgos no justificados.
3. Evaluación Imparcial del Modelo: Se utilizaron métricas claras y equitativas (como AUC, precisión, y F1-score) para evaluar el rendimiento del modelo sin favorecer ninguna clase en particular.

## Escenarios Potenciales de Uso Ético Incorrecto

A pesar de que la solución cumple con las normativas y no introduce sesgos de manera deliberada, existen ciertos escenarios en los que se podría incurrir en faltas éticas si la herramienta es mal utilizada:

1. Malicia: Si alguien usa el modelo para realizar decisiones clínicas sin el debido consentimiento o sin un médico profesional presente, podría inducir errores diagnósticos que afecten negativamente a la salud de los pacientes.
2. Negligencia en el Uso del Modelo: Si el modelo es mal interpretado por personal no capacitado o se usa sin supervisión médica adecuada, existe el riesgo de proporcionar diagnósticos erróneos.
3. Discriminación Algorítmica: Si se amplía su uso a poblaciones diferentes sin reentrenamiento adecuado, el modelo podría volverse sesgado hacia ciertos grupos poblacionales.