

Instituto Tecnológico y de Estudios Superiores de Monterrey

Campus, Querétaro



**Tecnológico
de Monterrey**

*Inteligencia Artificial Avanzada para
la Ciencia de Datos - (Gpo 103)*

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

Estudiante:

Samantha Guanipa Ugas

A01703936

Fecha de entrega:

05 de septiembre de 2023

I. Proyecto desarrollado que incluye el uso de un dataset con datos reales

Para el módulo 2 de aprendizaje máquina desarrollé un clasificador binario que es capaz de predecir si una persona padece Alzheimer o no, utilizando un modelo de regresión logística para su entrenamiento. El dataset utilizado fue obtenido desde la plataforma “Kaggle”, y fue obtenido con el nombre de “Handwriting Data to Detect Alzheimer’s Disease”. Este dataset cuenta datos de la escritura de 174 participantes, y cuenta con 451 atributos.

II. Normativa que se debe cumplir al tipo de datos utilizados

Al momento de utilizar datos personales que existen en un dataset hay una serie de normativas que deben cumplirse en México. El dataset utilizado cuenta con algunos datos que pueden llegar a ser sensibles y pueden incurrir en faltas a la persona de la cuál se poseen sus datos personales.

Por ejemplo, en el artículo 16 de la constitución política de los Estados Unidos Mexicanos se menciona que “Toda persona tiene derecho a la protección de sus datos personales, al acceso, rectificación y cancelación de los mismos, así como a manifestar su oposición, en los términos que fije la ley, la cual establecerá los supuestos de excepción a los principios que rijan el tratamiento de datos, por razones de seguridad nacional, disposiciones de orden público, seguridad y salud públicas o para proteger los derechos de terceros.”(Constitución Política de los Estados Unidos Mexicanos, 2019).

Esto es muy importante de tener en cuenta al momento de utilizar el dataset para el análisis de los datos, ya que se deben proteger los mismos para no incurrir en una falla respecto a este artículo constitucional.

Por otra parte, el Instituto Nacional de Transparencia, Acceso a la Información y Protección de Datos Personales (INAI) explica que la Ley Federal de Protección de Datos Personales en Posesión de los Particulares (LFPDPPP) es aplicable para este proyecto, debido a que “la LFPDPPP es aplicable a todas las personas o empresas privadas que utilicen tus datos personales, como podría ser una compañía de teléfono celular, una empresa donde se realizan compras en línea, o bien, una aplicación donde se pueden ver películas y series.”(INAI, 2021). Por lo tanto, se debe tomar en cuenta los principios de esta ley en el tratamiento de los datos personales recabados para el análisis.

III. Medio por el cuál se utilizaron los datos (normativa para la base de datos)

Como se mencionó anteriormente, este dataset fue obtenido desde “Kaggle” que es una plataforma que contiene una gran cantidad de bases de datos que han sido donadas a la misma, de manera consciente por parte de los donadores. Sin embargo, para constatar que no se está violando la normativa establecida, entré a la plataforma que le donó esta base de datos a Kaggle, y se observa que esta base de datos fue creada por Francesco Fontanella, y cuenta con una licencia “Commons Attribution

4.0 International (CC BY 4.0)” que permite compartir y adaptar el dataset teniendo en cuenta el nombre del autor y se menciona que “esta licencia permite que otros distribuyan, mezclen, modifiquen y desarrollen tu obra, incluso con fines comerciales, siempre que se atribuya el mérito de la creación original al autor. Debe dar el crédito apropiado , proporcionar un enlace a la licencia e indicar si se realizaron cambios. Puede hacerlo de cualquier manera razonable, pero no de ninguna manera que sugiera que el licenciante lo respalda a usted o su uso.” (Creative Commons, 2023).

IV. Cumplimiento de la normativa

Teniendo en cuenta los artículos mencionados anteriormente en este reporte, se analizará las modificaciones realizadas al dataset y el tratamiento de los datos del mismo. Primeramente, al dataset no se le hizo alguna modificación, ni se le agregaron datos adicionales a lo largo del proyecto. Lo único que se hizo fue analizar los datos contenidos en el mismo, removiendo en una copia ubicada en un dataframe los datos que no se utilizarían, como el ID del paciente, con el fin de cumplir con el reglamento de la LFPDPPP, donde no se debe mostrar datos personales del paciente cumpliendo con el deber de confidencialidad que establece que “el responsable debe guardar secreto respecto de los datos personales que son tratados, es decir, que los datos personales del titular no se difundan o compartan con terceros, salvo que se cuente con el consentimiento para ello, o bien, por alguna obligación normativa que exija su difusión” (INAI, 2021). En este caso, solo se conservaron los datos del estudio, como el tiempo que le tomó escribir a cada paciente, la presión con la que agarraba el lápiz, entre otros datos que no se consideran vulnerables o que puedan ser confidenciales.

V. Escenarios en los que se puede incurrir en una falta ética

Considero que se puede incurrir en una falta ética al momento de publicar la base de datos y el lugar donde se recolectaron los mismos, y que alguien pueda ir al lugar mencionado, por ejemplo, un hospital, y preguntar a qué persona corresponde cada ID de cada paciente, eso podría ser un acto de malicia para investigar a los pacientes más a fondo y afectarlos de manera monetaria o física. Sería un acto negligente de mi parte el redactar de qué lugar físico se obtuvieron los datos y dejar el ID de cada paciente, por ejemplo.

Por otra parte, al ser este un proyecto meramente académico, se podría caer en negligencia si se le solicitara a un adulto mayor que escriba en un papel con el fin de hacerle un diagnóstico basado en el modelo de clasificación binaria realizado, sin la presencia de un médico, ya que no se cuenta con un certificado médico que valide el mismo para su uso en la medicina.

VI. Conclusión

Como se pudo observar a lo largo de este reporte, en México existen diversas normativas que regulan el uso de los datos, protegiendo a la persona que proporciona sus datos personales y a la persona que crea la base de datos con el fin de que otros

científicos o académicos puedan realizar estudios o experimentos por su cuenta. Es por esto que a lo largo del proyecto realizado para la detección de Alzheimer en pacientes se protegieron los datos vulnerables, y a su vez no se les hizo una modificación que pudiera afectar al creador de la base de datos, ni se robó el crédito de la misma.

Referencias

Suprema Corte de Justicia de la Nación. (2020). Constitución Política de los Estados Unidos Mexicanos (CPEUM).

<https://www.scjn.gob.mx/sites/default/files/cpeum/documento/2020-06/CPEUM-016.pdf>

Instituto Nacional de Transparencia, Acceso a la Información y Protección de Datos Personales (INAI). (s.f.). Marco de competencias del INAI.

https://micrositios.inai.org.mx/marcocompetencias/?page_id=370#_ftnref1

Creative Commons. (2023). Attribution 4.0 International (CC BY 4.0).

<https://creativecommons.org/licenses/by/4.0/>