



Tecnológico de Monterrey
Escuela de Ingeniería y Ciencias

Tecnológico de Monterrey

ESCUELA DE INGENIERÍA Y CIENCIAS

Inteligencia Artificial Avanzada para la Ciencia de Datos

Análisis del contexto y normatividad de una base de datos

05/09/2023

Profesor:

Daniel Otero Fedul

Autor:

Leonardo Ramírez Ramírez - A01351715



1. Introducción

Debido al gran desarrollo de las tecnologías de la información junto con su creciente implementación en tareas cotidianas dentro del quehacer humano, la dependencia de la correcta aplicación de las tecnologías se ha visto cada vez más palpable y necesaria para facilitar ciertas tareas. Con la avenida de esta revolución tecnológica, nació la necesidad de tener registros de información para analizarlos y asegurar estabilidad al tener un nivel de control sobre el futuro con las predicciones que pueden inferirse a través de los datos. Así fue como el Big Data transformó el paradigma contemporáneo actual y con ello surgió la necesidad de asignar normas y reglas a las bases de datos con el fin de asegurar un correcto uso.

Según la compañía u organismo nacional o internacional puede existir ciertas reglas para respetar el uso y privacidad de los datos contenidos en ciertas bases de datos. Por ello, surge la obligación como ingenieros de datos y científicos de datos de buscar la normatividad asociada al dataset en uso y respetar los lineamientos estipulados por los derechos de autor, reglas de uso, etc.

2. Normatividad de una base de datos

Durante el semestre se han utilizado varias bases de datos para diferentes finalidades, pero en esta ocasión se selecciona la base de datos *Summary of Weather* en la cual se aplicó un modelo de regresión de Machine Learning para predecir la temperatura máxima a partir de la temperatura mínima.

2.1. Sobre la base de datos

La base de datos *Summary of Weather* contiene información meteorológica sobre el clima en los años 1940 y 1945 durante la Segunda Guerra Mundial. Según su descripción, la base de datos fue de gran ayuda para determinar el día de los bombardeos y la elección del Día D.

Los reconocimientos de la base de datos son de *United States National Oceanic and Atmospheric Administration (NOAA)*.

La base de datos contiene variables como *Precipitación*, *Temperatura mínima*, *Temperatura máxima*, *Caída de nieve*, *Mal clima* y *Velocidad de ráfagas de viento*. Todas ellas relacionadas a parámetros meteorológicos.

2.2. Licencia de la base de datos

Según lo encontrado en la plataforma de Kaggle, la base de datos está asociada a la licencia **U.S Government Works** la cual se registra a trabajos, informes, estadísticas y registros por parte de organismos o agencias gubernamentales. La mayoría de los documentos registrados con esta licencia



son de acceso público y están disponibles para ser distribuidos, utilizados y copiados libremente sin restricciones de derechos de autor.

La idea detrás de esta licencia en los Estados Unidos es que la información recopilada por parte de organizaciones que trabajan con fondos públicos debe estar disponible al acceso general del público.

Adicionalmente, la **NOAA** anunció en el siguiente link información importante respecto al citar la base de datos, así como información relevante sobre algunas restricciones a tomar en cuenta si se trabaja con ella.

Primeramente se menciona que se debe citar la base de datos si se usa como fuente para algún trabajo de cualquier índole. También se muestran los autores para citar siendo *DOC/NOAA/NWS - National Weather Service, NOAA, U.S. Department of Commerce*. Adicionalmente se mencionan temas importantes tales como:

- **Responsabilidad de uso:** La responsabilidad de uso se refiere a la forma con la que se trabaja la base de datos, incluyendo los objetivos y las metas que se buscan lograr según el trabajo. NOAA y NCEI no pueden ofrecer ninguna garantía sobre la precisión, confiabilidad o integridad de los datos proporcionados. Los usuarios asumen la responsabilidad de determinar la utilidad de estos datos. El usuario es responsable de los resultados de cualquier aplicación de estos datos que vaya más allá de su propósito previsto.
- **Responsabilidad de distribución:** NOAA y NCEI no tienen responsabilidad de posibles daños causados por la omisión o presencia de errores en los datos.

2.3. Sobre el uso ético del proyecto

Teniendo lo anterior en cuenta, el trabajo que se realizó se hizo cumpliendo la normatividad anteriormente especificada pues los datos de uso abierto se usaron para entrenar un modelo de regresión lineal. Los resultados obtenidos fueron simplemente una predicción de la temperatura máxima según la temperatura mínima. Como se trata de una base de datos del siglo pasado durante un acontecimiento ya ocurrido, los datos obtenidos no tienen alguna relevancia actual en cuestión de predicción del clima. En cambio, los resultados ayudan a esclarecer las relaciones lineales de algunas variables según los datos disponibles, además como se especificó en el contexto del problema, en aquel entonces era indispensable predecir el comportamiento del clima para llevar a cabo operaciones militares.

En cuestión del mal uso que se le pueda dar al modelo en situaciones poco éticas o maliciosas realmente no sería un escenario plausible pues como de mencionó, el dataset ya está desactualizado y los resultados poco éticos que de alguna manera se pudieran obtener no serían relevantes en el contexto actual. Además, es importante considerar que el modelo no es lo suficientemente avanzado como para lograr predicciones muy importantes, por el contrario, al ser una actividad de aplicación



de un modelo de ML sin un framework de trabajo o librería, no se le prestó demasiada importancia al generar un modelo avanzado.

Probablemente lo que podría plantear un posible uso poco ético es el manipular los datos incorrectamente para obtener resultados sesgados para justificar falsas conclusiones, donde dichas conclusiones afecten o discriminen a un grupo de personas según el registro histórico del clima en ciertas zonas.

Además, es importante recalcar que los datos se usaron para acciones militares (poco éticas consideradas desde el marco de defensa de los derechos humanos), sin embargo, éstos datos ya no podrían ser útiles si se intentara hacer lo mismo en el contexto actual por lo que es improbable que el modelo desarrollado se utilice para estos fines.

3. Conclusión

La importancia de la normatividad en el manejo de bases de datos y el uso ético de la información contenida en ellas es fundamental en la era digital y en el contexto actual de interconexión global. La normatividad garantiza la protección de los datos personales y la privacidad de los individuos, así como la integridad y confiabilidad de la información. Además, promueve la transparencia en la gestión de datos y evita prácticas perjudiciales o abusivas.

Es importante considerar el respeto de los derechos de autor al ser un componente esencial en esta normatividad. Con esto se garantiza que los autores reciban el crédito y reconocimiento por su trabajo. Esto se ha vuelto cada vez un tema serio en el actual mundo donde la aplicación activa de la tecnología se ha visto envuelta en la vida cotidiana por lo que la normatividad y la ética en el manejo de la información se han vuelto cruciales para proteger los derechos individuales de sus creadores, promover la confianza y trabajo colaborativa así como promover la innovación.

Referencias

[NOAA] NOAA (-). World war ii weather record transmittances. <https://www.ncei.noaa.gov/access/metadata/landing-page/bin/iso?id=gov.noaa.ncdc:C01208>.

Los créditos de las fotografías pertenecen a sus respectivos autores. ©