



Tecnológico de Monterrey

Instituto Tecnológico y de Estudios Superiores de Monterrey

**Momento de Retroalimentación. Análisis del contexto y la
normatividad**

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

Profesores:

Ivan Mauricio Amaya Contreras

Blanca Rosa Ruiz Hernandez

Antonio Carlos Bento

Frumencio Olivas Alvarez

Hugo Terashima Marín

Alumno:

Alberto H Orozco Ramos – A00831719

11 de Septiembre de 2023

Análisis del contexto y la normatividad

Para esta actividad, seleccioné el conjunto de datos correspondiente al módulo "Aprendizaje Máquina" del conjunto de actividades, el cual contiene información, principalmente recaudaciones, relacionada con películas de Marvel. Es así que, a lo largo de este reporte, se busca demostrar si los datos utilizados para dicha entrega cumplen con normatividad de datos.

Aplicación de la Norma ISO 8000 en la Calidad de Datos

La implementación de la norma ISO 8000 desempeñó un papel fundamental en el proceso de garantizar la calidad de los datos utilizados en este proyecto. Esta norma, que se enfoca en la calidad de los datos, establece directrices específicas para la gestión y el control de los datos utilizados en diversas aplicaciones. Al aplicar esta norma a el conjunto de datos de películas de Marvel, se han obtenido varios beneficios concretos.

En primer lugar, la norma ISO 8000 nos alentó a realizar una revisión concreta de la documentación disponible en Kaggle antes de iniciar cualquier análisis o entrenamiento del modelo. Esto aseguró la comprensión a fondo sobre la calidad de los datos, su estructura y su idoneidad para nuestros propósitos. Identificar las características esenciales de los datos, como la consistencia y la integridad, se convirtió en una prioridad gracias a esta norma.

Además, la norma ISO 8000 nos impulsó a verificar la veracidad de los datos que estábamos utilizando. No solo confiamos en la fuente original (Kaggle), sino que también consultamos estadísticas y opiniones de sitios reconocidos como IMDb y Rotten Tomatoes para corroborar la precisión de los datos, especialmente en lo que respecta a la reacción del público y la crítica a las películas.

En resumen, la aplicación de la norma ISO 8000 en nuestro proyecto de Aprendizaje Automático garantizó que estuviéramos cumpliendo con los estándares de calidad de datos definidos por esta norma internacional. Nos ayudó a abordar la calidad, la veracidad y la integridad de los datos de manera metódica y rigurosa. Esta norma proporcionó un marco sólido para garantizar que los datos utilizados en nuestro análisis y modelado fueran confiables y adecuados para la toma de decisiones.

A continuación, se detalla la implementación del modelo paso a paso para verificar si se cumple con las normativas:

- Previo a cualquier implementación de los datos dentro del modelo, se aseguró que los datos se encontraran limpios y sin campos faltantes, esto debido a que no se es posible entrenar un algoritmo de Machine Learning teniendo un dataset incompleto o con datos que puedan provocar irregularidades no deseadas en el comportamiento del algoritmo.
- Se analizaron todas y cada una de las variables dentro del set de datos con el fin de entender a qué se refiere cada encabezado, sus valores y cómo influyen el uno del otro sobre una misma entrega de Marvel.
- Para asegurarse de obtener un conjunto de datos confiable, es decir, que los datos que se utilizaron sean verídicos, no solo bastó con obtenerlos de una página de confianza

como lo es Kaggle, sino que además se consultaron estadísticas de páginas reconocidas por distintos medios como IMDb o Rotten Tomatoes, que proporcionan confiables acerca de películas y estrenos en cines, ya sean recaudaciones u opiniones del público, dónde se consultaron películas como IronMan, Avengers: Infinity War, Spider-Man: No Way Home y Guardians of the Galaxy Vol. 1 (IMDb, 2023); en dónde pude corroborar que los datos incluidos dentro del dataset eran verídicos, más que nada la opinión y rating tanto del público como de la crítica en general (Rotten Tomatoes, 2023).

- El código desarrollado no modifica de ninguna manera el conjunto de datos original ni sobrescribe ningún dato en el archivo CSV. El archivo se carga en el entorno de Google Colab, se lee y se almacena en variables utilizando Python. Estas variables se utilizan como una "copia" de respaldo de los datos originales para evitar alterarlos.
- Durante y después de todos los procesos realizados para generar predicciones, que incluyen dividir el conjunto de datos en dos y generar las predicciones, no se altera el archivo original en absoluto. Todo el proceso se basa en la lectura de datos y no implica escritura ni reemplazo.

Es importante señalar que, aunque este caso involucra un algoritmo de Aprendizaje Automático simple que realiza lectura de datos de un archivo CSV de recaudaciones de películas, esta práctica podría tener consecuencias graves si se aplica a datos sensibles o personales, o si se utiliza de forma maliciosa. Algunos ejemplos de problemas potenciales incluyen:

- **Violación de Privacidad:** Si el algoritmo tratara datos sensibles o personales y no se diseñará para protegerlos mediante técnicas de ocultamiento, codificación o cifrado, esto podría considerarse una violación de la privacidad de los usuarios.
- **Falta de Transparencia:** A pesar de que se trate de un algoritmo que solo realiza lectura de archivos, en ciertos casos, la falta de transparencia en el proceso de lectura, transformación, decodificación o tratamiento de datos, incluyendo la presentación de resultados, podría ser considerada un incumplimiento ético y una violación de las normas, incluyendo los principios éticos y la norma ISO 8000 (datos.gob.es, 2022).

En conclusión, incluso algoritmos aparentemente simples como este pueden tener un impacto significativo si no se aplican siguiendo las normativas adecuadas. Cumplir con la normatividad es fundamental para garantizar la privacidad, la ética y la integridad de los datos en cualquier proyecto de Aprendizaje Automático.

Bibliografía

- Gordon Keith. (2013). "J.9 ISO 8000 - Data Quality." En "Principles of Data Management - Facilitating Information Sharing" (2da edición). BCS The Chartered Institute for IT.
- datos.gob.es. (2022, 9 de junio). "Normas técnicas para alcanzar la calidad del dato." datos.gob.es. [En línea] Disponible en: <https://datos.gob.es/es/blog/normas-tecnicas-para-alcanzar-la-calidad-del-dato#:~:text=La%20ISO%208000%2D61%20establece,%2DDo%2DCheck%2DAct>. Consultado el 29 de Agosto de 2023.
- IMDb: Ratings, reviews, and where to watch the best movies & TV shows. (s. f.). IMDb. https://www.imdb.com/?ref_=nv_home
- Rotten Tomatoes: movies | TV shows | Movie trailers | reviews. (s. f.). Rotten Tomatoes. <https://www.rottentomatoes.com/>