

**Momento de Retroalimentación: Reto Análisis del contexto y la normatividad. (Portafolio Análisis)**

Por: Fernando Bustos Monsiváis – A00829931

Escuela de Ingeniería y Ciencias, Instituto Tecnológico y de Estudios Superiores de Monterrey

TC3006C.102: Inteligencia artificial avanzada para la ciencia de datos I

Profesor: Dr. Alfredo Esquivel Jaramillo

Martes 10 de septiembre de 2024

1. **Selecciona uno de los proyectos que hayas desarrollado que involucre el uso de un set de datos reales, puede ser un proyecto del módulo 1: Estadística para la ciencia de datos o en el módulo 2: Aprendizaje máquina. No puede ser el del reto ya que este es un trabajo individual por ser una evidencia.**

Se seleccionó el siguiente proyecto del módulo de Machine Learning: *Momento de Retroalimentación: Módulo 2 Implementación de una técnica de aprendizaje máquina sin el uso de un framework. (Portafolio Implementación).*

En este proyecto se implementó un modelo de regresión lineal simple para predecir la temperatura real utilizando la temperatura aparente como única característica (feature), sin el uso de un framework de aprendizaje de máquina. Dicho proceso se realizó con un dataset de Kaggle que contiene datos de temperaturas en Szeged, Hungría, desde 2006 hasta 2016.

2. **Investiga cual es la normativa asociada al tipo de datos que utilizaste.**

El dataset utilizado en este proyecto, *Weather in Szeged 2006-2016 (weatherHistory.csv)*, está protegido bajo la licencia Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0). Esto implica que:

- **Atribución:** Se debe dar el crédito adecuado a los creadores del dataset.
- **Uso no comercial:** Los datos no pueden utilizarse para fines comerciales.
- **Compartir igual:** Si se modifica o construye sobre el dataset, debe compartirse bajo la misma licencia.

Enlace a la licencia: [CC BY-NC-SA 4.0](https://creativecommons.org/licenses/by-nc-sa/4.0/)

3. **Explica por que medio utilizaste los datos y como te estas asegurando de no violar la normativa establecida. Respalda esto con los documentos oficiales en los que te estás basando y cítalos.**

El proyecto utiliza los datos exclusivamente para fines académicos y de investigación, cumpliendo así con la condición de uso no comercial. Además, en el [repositorio público de GitHub](#) donde se encuentra el proyecto, se ha otorgado el crédito adecuado al propietario del dataset, junto con la licencia que regula su distribución (CC BY-NC-SA 4.0), incluyendo las referencias correspondientes al final del archivo.

Al tratarse de un proyecto académico, no se infringen las condiciones de uso comercial, ya que no se ha monetizado ni utilizado en productos o servicios comerciales. También, si se llegaran a hacer modificaciones al dataset para análisis o futuras investigaciones, estas se compartirían bajo la misma licencia, cumpliendo con el requisito de compartir igual (Creative Commons (CC), 2024).

4. **Una vez analizado el dataset ahora hay que analizar la herramienta, explica cómo tu solución cumple con la normativa establecida en la misma industria y no incurre en ningún sesgo ético.**

El modelo implementado en este proyecto es transparente tanto en el análisis como en la predicción de datos. Se ha utilizado un algoritmo de regresión lineal simple, un método ampliamente aceptado y estadísticamente validado, y se ha documentado detalladamente cada etapa del proceso, desde el entrenamiento hasta la evaluación del modelo.

El proyecto está disponible en un repositorio público de GitHub, lo que permite a cualquier persona verificar que es un trabajo académico sin fines comerciales. Se han proporcionado las referencias correspondientes tanto al dataset original de Kaggle como a la licencia bajo la cual se distribuyen los datos, cumpliendo con las normativas de atribución y la obligación de compartir bajo los mismos términos. Además, los usuarios pueden comparar fácilmente el archivo CSV utilizado en el proyecto con el dataset de Kaggle para confirmar que son idénticos, garantizando la transparencia y el cumplimiento de las condiciones de la licencia (CC BY-NC-SA 4.0).

En cuanto al sesgo ético, se ha asegurado que el dataset no contenga información sensible que pueda identificar a personas, lo cual elimina la posibilidad de violar normativas de privacidad de datos. Además, el objetivo del análisis es puramente académico, sin implicaciones en la toma de decisiones que afecten a individuos.

**5. Analiza bajo qué escenarios se podría incurrir en una falta ética si tu herramienta es mal usada, contempla casos como malicia o negligencia.**

Aunque el proyecto en sí no presenta riesgos éticos inmediatos, se podría incurrir en una falta ética si:

- **Mal uso de los resultados:** Si el modelo desarrollado se utilizara para predecir escenarios meteorológicos que implicaran decisiones críticas (como en emergencias climáticas) sin una debida verificación o ajuste por expertos, se podrían generar decisiones incorrectas que afecten a las personas.
- **Modificación negligente del modelo:** Cambios no controlados en el modelo, como aumentar excesivamente el número de epochs o modificar el learning rate sin supervisión adecuada, podrían generar predicciones inexactas, con implicaciones éticas en el uso de esos resultados.
- **Uso malintencionado de los datos:** Si se manipulara el dataset o los resultados del modelo para presentar conclusiones falsas, se estaría incurriendo en una práctica ética negativa. Esto podría incluir el uso intencionado de predicciones incorrectas para justificar decisiones políticas o comerciales, por ejemplo.

### Referencias

- Creative Commons (CC). (2024). *Atribución/Reconocimiento-NoComercial-CompartirIgual 4.0 Internacional*. Recuperado el lunes 09 de septiembre de 2024, de <https://creativecommons.org/licenses/by-nc-sa/4.0/deed.es>
- Budincsevity, N. (2016). *Weather in Szeged 2006-2016*. Recuperado el 09 de septiembre de 2024, de Kaggle: <https://www.kaggle.com/datasets/budincsevity/szeged-weather?select=weatherHistory.csv>