



Tecnológico de Monterrey

Instituto Tecnológico y de Estudios Superiores de Monterrey

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

TC3004B.104 Inteligencia Artificial Avanzada para la Ciencia de Datos I

Profesores:

Ivan Mauricio Amaya Contreras

Hugo Terashima Marín

Blanca Rosa Ruiz Hernandez

Antonio Carlos Bento

Frumencio Olivas Alvarez

Nombre del estudiante:

Luis Ángel Guzmán Iribe

Matricula:

A01741757

29 de Agosto de 2023

Normativa legal del set de datos

Licencia: CC0: Public Domain

Enlace a Kaggle del dataset utilizado:

<https://www.kaggle.com/datasets/yasserh/student-marks-dataset>

La licencia de Dominio Público CC0 de Creative Commons cede derechos para copiar, modificar, distribuir y ejecutar el trabajo en cuestión sin necesidad de solicitar autorización al autor, incluyendo el uso comercial del trabajo; sin embargo, no se cede ningún derecho de marca registrada o patente, es decir, no es posible afirmar autoría sobre el trabajo original.

Uso del set de datos

El set de datos se utilizó con la finalidad de generar modelos de regresión y polinomial buscando predecir las marcas de los alumnos. No se realiza ninguna modificación sobre el set de datos, no se afirma autoría propia y se comparte junto con el proyecto. Dado que la licencia permite un alto grado de libertad sobre el uso del trabajo, no se infringe ninguno de los puntos establecidos en el acuerdo [1].

Uso éticamente inadecuado del modelo

Si bien las implicaciones legales del uso del conjunto de datos son prácticamente inexistentes además de la autoría original, es importante reconocer las limitaciones en las aplicaciones del modelo predictivo generado a partir de ellas.

Por ejemplo, el modelo predictivo genera un estimado de la calificación de los estudiantes basado única y exclusivamente en la cantidad de horas de estudio que dedican semanalmente, esto no significa que el modelo deba ser utilizada como herramienta para calificar estudiantes basándose únicamente en esta variable, ya que ignora lo méritos y complejidades del mundo real; su única intención es señalar una correlación interesante entre el tiempo dedicado al estudio y los resultados de este esfuerzo.

Tampoco considero pertinente hacer uso de este modelo predictivo para afirmar que el único factor para el éxito de un estudiante, ya que, reitero, ignora otros factores y complejidades del mundo real. De este modo resultaría poco ético afirmar que aunque se haya encontrado una fuerte relación entre las horas semanales dedicadas al estudio, estas sean el único factor para el mismo, y se divulgue este pedazo de información como conocimiento absoluto; es irresponsable hacer afirmaciones tan contundentes tomando

como base un modelo predictivo sujeto a errores tanto matemáticos como humanos en el momento de entrenamiento, recolección de datos, análisis, etc...

Aterrizando esta reflexión en un escenario más conciso, sería irresponsable que, por ejemplo, un profesor en una escuela utilice este modelo como métrica de evaluación para sus estudiantes, basándose en la asunción que hace el modelo de que solamente el tiempo dedicado al estudio semanalmente para determinar la calificación final de sus estudiantes. Esta acción constituye un malentendido de la intención del modelo, al utilizarlo como herramienta de evaluación e incluso con fines punitivos en el caso de alumnos que determine con notas no aprobatorias, independientemente de su desempeño real.

Normativa legal de herramientas

- Visual Studio Code
- Python
- Pandas

En las licencias de los 3 productos se habla acerca de modificaciones y distribución del código fuente del software, uso malicioso del mismo, o su uso comercial sin acreditación (en el caso de pandas) [2]. Algunas acciones que pudieran quebrantar sus términos de servicio son la distribución de versiones alteradas del software como Python [3], o en el caso de VScode, usando el producto de tal manera que rompa el Acuerdo de Licencia de Usuario Final (EULA) de Microsoft, esto incluye el uso del producto para finalidades ilegales [4].

Referencias

Licencias del set de datos:

- [1] Creative Commons. (s.f.). CC0 1.0 Universal (CC0 1.0) Public Domain Dedication. [Online]. Available: <https://creativecommons.org/publicdomain/zero/1.0/>

Licencias de herramientas:

- [2] NumFOCUS (2023). PANDAS Software License. [Online]. Available: <https://github.com/pandas-dev/pandas/blob/main/LICENSE>
- [3] Python Software Foundation (2023). History and License. [Online]. Available: <https://docs.python.org/3/license.html>
- [4] Microsoft Corporation. (2023). Terminos de licencia de uso de software. [Online]. Available: <https://code.visualstudio.com/license?lang=es>