

**TECNOLÓGICO DE MONTERREY**

TC3006C: Inteligencia artificial avanzada para la ciencia de datos



**Tecnológico  
de Monterrey**

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

Francisco José Joven Sánchez - A00830564

**Ingeniería en Tecnologías Computacionales**

*Fecha de entrega - 06 / 09 / 2023*

Hasta el momento no he utilizado un dataset que tenga ciertas normas dado que la otra es simple práctica como la del titanic, por lo que tomaré un dataset que, si las tenga y lo usaré como ejemplo hipotético para escribir el reporte y talvez utilizar para una próxima entrega, en esta ocasión usaré el que se utilizó para una competición de Santander correspondiente a la predicción de las transacciones de los clientes.

Este dataset tiene múltiples normas, estas con relación con los participantes, la posibilidad de los grupos, le procedencia de las personas, los premios, etc. Ahora en cuestión a las normas con relación a los datos, primero esta el acceder/utilizar los datos proveídos solo para la competición, investigación académica, de forma educativa o algún otro uso no comercial. En segunda trata de la seguridad, que básicamente es tomar medidas para que ninguna persona que no haya aceptado las reglas de la competición sea capaz de acceder al dataset, además que notificar inmediatamente a Kaggle en caso de llegar a saber que alguien así se haya enterado. Por último, esta el no usar los datos fuera de aquellos dados para la competición para entrenar modelos a menos que este explícitamente indicado en las reglas.

Estos datos son de la competición que como mencione está en Kaggle, dentro de los términos de la página, está establecido en el punto 7 que las reglas de la competición están definidas en la misma y son acordadas al entrar a ellas. En cuanto a estas, los primeros 6 puntos tratan simplemente sobre cumplir con los requisitos para entrar a la competición, en relación con aceptar las reglas, no ser residente de ciertos lugares, ser mayor de edad, el periodo de competición, el como se debe entrar a la competición, sin pagar, por ejemplo, y el participar individualmente o de manera grupal. Después esta el punto 7 en relación con los datos, el cual ya describí en el párrafo anterior. El punto 8 trata sobre el código desarrollado, en términos de uso publico y privado. Del punto 9 hasta el 12 trata sobre determinar ganadores, las obligaciones de estos y los premios, junto con lo que implica recibirlos. Para los demás puntos, concierne en términos un tanto más de legalidad explicita con respecto a la recolección de datos, garantías, indemnizaciones, descalificaciones, modificaciones, extensión de que abarca la competición y finalmente que cualquier reclamo con respecto a las reglas será procesado bajo la ley española. Esos son en resumen todos los puntos de normativas dentro de la competición.

Para hablar del sesgo ético, debemos primero identificar que es, dado que puede ser en cuanto el uso incorrecto de herramientas privadas, la divulgación de datos a agentes externos por cualquier razón o incluso crear un modelo que cree datos verdaderos para la competencia, pero creará datos falsos fuera de esta. Para estos casos simplemente tendría cuidado y tomaría precauciones para que ningún pase, para la primera usaría únicamente código público de ser necesario, en el segundo mantendría los datos en un solo dispositivo bajo contraseña si realmente son datos importantes y para la última cuestión no tendría como objetivo que eso

pase, pero tendría que vigilar casos como no sobreentrenar el modelo hasta cierto punto dado que podría ser la causa.

Las reglas mencionadas dentro de la competición son un tanto estándar para competiciones formales y es necesario tomar precauciones para evitar las consecuencias de cuestiones como la negligencia para alguien inexperto como yo. Ahora que si bien pueden haber actos de malicia no es algo que yo haría dado los problemas que estos podrían implicar.