



# Tecnológico de Monterrey

**INSTITUTO TECNOLÓGICO DE  
ESTUDIOS SUPERIORES DE MONTERREY**

**Inteligencia artificial avanzada para la ciencia de datos I (Gpo  
101)**

**Portafolio Análisis: Reporte de ética y normatividad**

**05 de Septiembre de 2024**

**Adrian Pineda Sánchez**

**A00834710**

## Introducción a Kaggle

Kaggle es una plataforma digital que brinda a estudiantes, profesionales y expertos en ciencia de datos la oportunidad de acceder a conjuntos de datos, participar en competencias y compartir sus enfoques y soluciones. La plataforma pone a disposición herramientas para el análisis de datos y facilita la colaboración entre sus miembros, lo que contribuye al desarrollo de habilidades en este campo. Además, permite realizar comparaciones de rendimiento en las competencias a través de sistemas de clasificación y métricas definidas para cada reto.

### Normatividad del reto

El uso de datasets en Kaggle debe alinearse con ciertas normativas legales y principios éticos, especialmente en lo que respecta a la protección de los datos y la propiedad intelectual. Algunos de los aspectos más relevantes incluyen:

**Propiedad y Derechos sobre los Datos:** Los términos de uso de Kaggle especifican que quienes suben datasets deben asegurarse de que tienen los derechos necesarios para distribuir esa información. Esto implica que los datos no deben infringir derechos de propiedad intelectual ni violar leyes de privacidad o confidencialidad (Kaggle, n.d.). Además, la plataforma establece que los usuarios deben respetar las leyes de protección de datos vigentes, como el Reglamento General de Protección de Datos (GDPR), que regula el manejo de información personal.

**Permisos para su Publicación y Uso:** Es fundamental que cualquier dataset compartido cuente con los permisos correspondientes para su publicación. Cuando los datos contienen información personal, deben estar anonimizados de acuerdo con normativas como el GDPR, que protege los derechos de privacidad de los individuos (Reglamento General de Protección de Datos, 2018). Esta ley asegura que los datos personales sólo pueden ser tratados si existe un fundamento legal y que los individuos tienen el derecho de solicitar que se elimine su información si ya no es necesaria.

**Cumplimiento del GDPR:** Kaggle está alineado con el GDPR, que establece varios principios clave como la minimización de datos, la integridad y confidencialidad de los datos, y el derecho al olvido, que permite a las personas solicitar la eliminación de su información cuando ya no haya motivos legítimos para conservarla (Comisión Europea, 2018; DPO Centre, 2023).

## **Normatividad Aplicada al Dataset del Titanic**

**Propiedad de los Datos:** El dataset del Titanic en Kaggle es de dominio público. Sin embargo, los usuarios deben respetar las normativas de Kaggle que exigen atribución y respeto a las reglas de uso de la plataforma (Kaggle, n.d.).

**Cumplimiento de Normativas de Privacidad:** Aunque el dataset contiene información histórica que no está sujeta a las leyes de protección de datos como el GDPR, cualquier combinación con datos personales actuales implicaría la obligación de cumplir con estas normativas, que exigen la anonimización y minimización de los datos (Comisión Europea, 2018).

**Ética en el Uso de Datos:** Es importante evitar sesgos en el análisis de datos, sobre todo cuando se utilizan variables como el género o la clase social para predecir la supervivencia. Los científicos de datos deben ser responsables y conscientes de las implicaciones éticas de los modelos que desarrollan (Breidbach & Maglio, 2020).

## **Cumplimiento de Leyes, Normas y Principios Éticos**

En mi primera entrega, realicé un análisis sobre la supervivencia de los pasajeros del Titanic utilizando variables como la clase del boleto, el género y la edad. Al profundizar en este nuevo punto, me enfoqué en cómo los algoritmos toman decisiones automáticas y cómo pueden perpetuar ciertos sesgos, tal como se discute en el artículo *Accountable Algorithms*. Por ejemplo, uno de los principales desafíos que identifiqué es que el uso de variables históricas como el género podría generar un sesgo en los modelos predictivos, lo cual es un problema ético importante cuando se utilizan para tomar decisiones. Este tipo de análisis me permitió reflexionar sobre cómo los algoritmos pueden influir en áreas sensibles, como seguros o préstamos, si no se tiene cuidado en controlar estos sesgos.

También consideré el tratamiento de datos personales, y aunque el dataset del Titanic no involucra información confidencial moderna, es crucial tener en cuenta normativas como el GDPR cuando se trabajan con datos personales.

**Sesgo en Perfiles de Riesgo:** El uso de variables como el género o la clase social en el análisis de supervivencia del Titanic puede perpetuar sesgos históricos. Si los modelos se basan demasiado en estas variables, pueden reforzar desigualdades injustas. Es crucial que los científicos de datos mitiguen estos sesgos para asegurar decisiones justas y equitativas (Breidbach & Maglio, 2020).

**Toma de Decisiones Basadas en Resultados:** Las decisiones automáticas deben ser revisadas para evitar impactos negativos, especialmente en casos modernos como seguros o préstamos. Aunque el dataset del Titanic no tiene repercusiones legales, la lógica del modelo aplicado a datos actuales puede influir en grupos vulnerables si no se manejan los sesgos correctamente (Breidbach & Maglio, 2020).

**Tratamiento de Datos Confidenciales:** Aunque los datos del Titanic no son sensibles, cualquier dataset que contenga información personal debe cumplir con normativas como el GDPR, que exige minimización de datos, procesamiento seguro y derechos de eliminación de datos (Comisión Europea, 2018).

## **Conclusión**

Durante el desarrollo de la solución, se abordó el análisis del dataset del Titanic con un entendimiento académico del contexto histórico en el que se generaron los datos. Aunque los patrones de desigualdad observados en ese momento, como las diferencias en la supervivencia basadas en el género o la clase social, no reflejan los valores actuales, los datos fueron utilizados de la forma más ética posible para el análisis. Se emplearon técnicas que mitigaron cualquier posible sesgo, con el objetivo de evitar la perpetuación de esos patrones históricos en los modelos predictivos.

A pesar de que los datos no tienen relevancia directa en la actualidad, este ejercicio académico permitió explorar el impacto de los sesgos en la toma de decisiones automatizadas, brindando una valiosa oportunidad para reflexionar sobre cómo estos modelos podrían aplicarse en situaciones modernas de manera justa y equitativa (Breidbach & Maglio, 2020). Además, aunque no se maneja información confidencial moderna en este caso, se reafirma el compromiso de cumplir con normativas como el

GDPR cuando se trabaje con datos personales en otros proyectos (Comisión Europea, 2018).

Este enfoque asegura que la solución no solo es técnicamente sólida, sino que también cumple con las leyes, normas y principios éticos requeridos.

## Referencias:

- Comisión Europea. (2018). *Reglamento general de protección de datos (RGPD)*. Recuperado de <https://eur-lex.europa.eu>
- Kaggle. (n.d.). *Kaggle terms of service*. Recuperado de <https://www.kaggle.com/terms>
- Breidbach, C. F., & Maglio, P. (2020). *Accountable algorithms? The ethical implications of data-driven business models*.
- DPO Centre. (2023). Data retention and the GDPR: Best practices for compliance. Recuperado de <https://www.dpocentre.com/gdpr-compliance>
- Reglamento General de Protección de Datos. (2018). *Normas para las empresas de la UE*. Recuperado de [https://commission.europa.eu/law/law-topic/data-protection/reform/rules-business-and-organisations/principles-gdpr\\_en](https://commission.europa.eu/law/law-topic/data-protection/reform/rules-business-and-organisations/principles-gdpr_en)