# Normatividad Reto

Rogelio Lizárraga Escobar

September 10, 2024

## 1 Introducción de Kaggle

Kaggle es una plataforma digital que ofrece un espacio colaborativo para estudiantes, profesionales y científicos de datos, donde pueden compartir y trabajar en retos de análisis y modelado de datos. Además, Kaggle facilita el acceso a una gran variedad de conjuntos de datos, tanto de fuentes públicas como privadas, lo que permite a los usuarios poner en práctica sus habilidades en ciencia de datos, *machine learning*, y análisis estadístico. Su sistema de competiciones y evaluación comparativa (*leaderboards*) permite que las soluciones presentadas sean evaluadas objetivamente en términos de rendimiento, ofreciendo un entorno de aprendizaje constante. Para los estudiantes y profesionales, Kaggle es una herramienta educativa inigualable por su acceso a grandes bases de datos y por la posibilidad de desarrollar proyectos reales de alto impacto.

## 2 Normatividad del reto o socio formador (Kaggle)

La normatividad que rige el uso de bases de datos y conjuntos de datos en Kaggle es estricta en cuanto al manejo, propiedad y publicación de los mismos. De acuerdo con los términos de servicio de Kaggle, los usuarios deben respetar los derechos de propiedad de los datos que suben o utilizan en sus análisis. Los datos deben estar alineados con normativas globales como el Reglamento General de Protección de Datos (*GDPR*) para la Unión Europea y otras leyes de privacidad, como la Ley Federal de Protección de Datos Personales en Posesión de los Particulares en México [1].

El *GDPR* establece que cualquier procesamiento de datos personales debe cumplir con ciertos principios clave, incluyendo la legalidad, transparencia y

equidad en la recolección y manejo de la información. Para los sets de datos compartidos en Kaggle, como el caso del reto Titanic, es esencial que las personas o entidades que los proporcionan aseguren que los datos han sido anonimizados o que se ha obtenido el debido consentimiento para su uso y análisis. Además, cualquier dato compartido en la plataforma debe contar con los permisos legales correspondientes para su publicación y distribución. Los usuarios también deben estar conscientes de la política de Kaggle respecto al uso de los datos fuera de la plataforma, ya que las restricciones sobre cómo se puede utilizar un conjunto de datos varían dependiendo del propietario de los datos y del tipo de licencia asociada [2].

Kaggle establece claramente en sus términos de servicio que cualquier violación a los derechos de propiedad intelectual o la distribución indebida de los conjuntos de datos puede resultar en la suspensión de la cuenta o en acciones legales por parte de los propietarios de los datos [3]. Esto asegura que el uso de los datos esté alineado con las normativas legales vigentes en cada jurisdicción.

# 3 Cumplimiento de leyes, normas y principios éticos en la solución del reto

En el contexto del análisis algorítmico aplicado al reto del Titanic, es importante considerar los elementos éticos y legales involucrados en la toma de decisiones automatizadas. Como se discutió en la primera entrega del análisis, uno de los riesgos más grandes en la toma de decisiones algorítmica es el sesgo, que puede surgir tanto del diseño del modelo como de los datos utilizados para entrenarlo. En el caso de Kaggle, aunque el conjunto de datos del Titanic no contiene información sensible que pueda ser directamente rastreada a individuos vivos, el sesgo algorítmico sigue siendo un riesgo importante que debe abordarse.

Uno de los principios clave en cualquier análisis de datos es la *equidad*. Los modelos algorítmicos deben evitar favorecer o perjudicar injustamente a ciertos grupos de personas, en este caso, clases sociales, géneros, o edades. En el reto del Titanic, los datos de pasajeros incluyen variables como el sexo y la clase social ($1^{\underline{a}}$, $2^{\underline{a}}$, o $3^{\underline{a}}$ clase), y los modelos podrían introducir sesgos si no se manejan cuidadosamente estas variables. Por ejemplo, un modelo podría reflejar los prejuicios históricos presentes en los datos (como la supervivencia desigual de hombres y mujeres o de personas de diferentes clases sociales), perpetuando estos sesgos en los resultados de predicción.

Otro aspecto ético crucial es la *transparencia*. Al desarrollar modelos en

Kaggle o cualquier otra plataforma, es esencial que los resultados del modelo sean explicables y comprensibles. Los usuarios finales deben poder entender cómo y por qué un modelo toma ciertas decisiones, especialmente cuando estas decisiones tienen un impacto significativo, como lo es en el caso de seguros, créditos bancarios, o, en este caso, las predicciones de supervivencia basadas en datos históricos. La *explicabilidad* del modelo ayuda a garantizar que no se tomen decisiones arbitrarias o injustificadas [4].

Finalmente, el *tratamiento de datos confidenciales* es otro tema que no puede ser ignorado. Aunque el conjunto de datos del Titanic ha sido anonimizado y no contiene información confidencial según los estándares actuales, cualquier proyecto de análisis de datos debe adherirse a las mejores prácticas de privacidad de datos. Esto incluye garantizar que los datos sean almacenados de manera segura, que no se compartan indebidamente y que se cumplan todas las leyes de protección de datos aplicables, como el GDPR mencionado anteriormente. En el caso de un análisis que utilice datos sensibles, sería necesario asegurar que se hayan implementado todas las medidas necesarias para proteger la identidad y la privacidad de los individuos.

# 4 Conclusión

El análisis y la solución propuesta para el reto del Titanic en Kaggle deben cumplir con una serie de normativas éticas y legales, desde el tratamiento justo y equitativo de los datos hasta la protección de la privacidad de los individuos involucrados. A través del cumplimiento de estas normativas, no solo garantizamos la legalidad del análisis, sino también su legitimidad y confianza en los resultados obtenidos. Los científicos de datos y los desarrolladores de modelos deben estar siempre atentos a estos aspectos para evitar la introducción de sesgos, la discriminación y la vulneración de derechos de privacidad en el uso de datos.

El análisis ético y legal es esencial no solo para cumplir con la normativa, sino también para garantizar que el uso de la ciencia de datos contribuya al bienestar social en lugar de perpetuar desigualdades o vulnerar derechos fundamentales.

# 5 Referencias

# References


[1] Reglamento General de Protección de Datos (GDPR), *Parlamento Europeo y Consejo de la Unión Europea*, Recuperado de `https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX%3A32016R0679`.

[2] Kaggle, *Términos de Servicio*, Recuperado de `https://www.kaggle.com/terms`.

[3] Cámara de Diputados del H. Congreso de la Unión, *Ley Federal de Protección de Datos Personales en Posesión de los Particulares*, Diario Oficial de la Federación, Recuperado de `https://www.dof.gob.mx/nota_detalle.php?codigo=5150631&fecha=05/07/2010`.

[4] IEEE Standards Association (2021), *IEEE 7000-2021: IEEE standard model process for addressing ethical concerns during system design*, Recuperado de `https://standards.ieee.org/standard/7000-2021.html`.

[5] Hardt, M., Price, E., & Srebro, N. (2016), *Equality of opportunity in supervised learning*, Advances in Neural Information Processing Systems (NeurIPS), Recuperado de `https://arxiv.org/abs/1610.02413`.