

## **Inteligencia Artificial Avanzada para la Ciencia de Datos I**

Horario: 09:10–12:50 — Edificio: Aulas III — Salón: 204

# **El Reto del Titanic**

*Paper académico*

### **Equipo 1**

Kevin Horacio Molina Peñuñuri *A01253141*

Santiago Juárez Roaro *A01705439*

David Alejandro Acuña Orozco *A00571187*

Iván Alejandro López Valenzuela *A01284875*

Joaquín Sainz Muleiro *A01783801*

### **Profesorado**

Blanca Rosa Ruiz Hernández

Raúl Valente Ramírez Velarde

Iván Mauricio Amaya Contreras

Frumencio Olivas Álvarez

Antonio Carlos Bento

Alfredo Alan Flores Saldívar

**Fecha: 14 de septiembre de 2025**

ÍNDICE			
<b>I.</b>	<b>Introducción</b>	2	<b>VI. Conclusiones y Trabajo Futuro</b> 8
I-A.	Contexto del Problema . . . . .	2	VI-A. Resumen de Contribuciones . . . . . 8
I-B.	Motivación . . . . .	2	VI-B. Reflexiones Finales . . . . . 8
I-C.	Preguntas de Investigación . . . . .	3	VI-C. Trabajo Futuro . . . . . 8
I-D.	Contribuciones . . . . .	3	VI-D. Llamado a la Acción . . . . . 8
I-E.	Estructura del Paper . . . . .	3	<b>Referencias</b> 9
<b>II.</b>	<b>Revisión de Literatura</b>	3	<b>Apéndice</b> 10
II-A.	Trabajos Previos sobre el Dataset Titanic	3	
II-B.	Machine Learning Interpretable . . . .	3	
II-C.	Fairness en Machine Learning . . . .	3	
II-D.	Ética en IA y Decisiones Algorítmicas	4	
II-E.	Gap en la Literatura . . . . .	4	
<b>III.</b>	<b>Metodología</b>	4	
III-A.	Dataset y Preprocesamiento . . . . .	4	
III-A1.	Descripción del Dataset . .	4	
III-A2.	Análisis de Calidad de los Datos . . . . .	4	
III-A3.	Ingeniería de Features . . .	5	
III-A4.	Estrategia de Imputación .	5	
III-B.	Diseño Experimental . . . . .	5	
III-B1.	Formulación del Problema	5	
III-B2.	Estrategia de Validación . .	5	
III-B3.	Algoritmos Implementados	5	
III-B4.	Análisis de Fairness . . . .	5	
III-C.	Herramientas y Reproducibilidad . . .	5	
<b>IV.</b>	<b>Resultados</b>	5	
IV-A.	Análisis Exploratorio . . . . .	5	
IV-B.	Performance de Modelos . . . . .	6	
IV-C.	Interpretabilidad . . . . .	6	
IV-D.	Análisis de Fairness . . . . .	6	
IV-E.	Validación de Hipótesis . . . . .	7	
<b>V.</b>	<b>Discusión</b>	7	
V-A.	Interpretación de Resultados . . . . .	7	
V-B.	Implicaciones Éticas . . . . .	7	
V-C.	Limitaciones . . . . .	7	
V-D.	Comparación con Literatura . . . . .	7	
V-E.	Aplicaciones Prácticas . . . . .	7	

## ABSTRACT

**Contexto:** El hundimiento del Titanic en 1912 es un caso emblemático para el análisis de desigualdades sociales y sesgos en la toma de decisiones. La abundancia de registros históricos y la existencia de un dataset abierto (Kaggle, [s.f.](#)) lo han convertido en un referente para estudios de aprendizaje automático, donde confluyen factores de clase, género y edad como determinantes de la supervivencia.

**Objetivo:** Este trabajo busca identificar los principales determinantes de supervivencia en el Titanic, examinar cómo se reflejan los sesgos sociales en los datos y evaluar los *trade-offs* entre precisión predictiva y equidad en modelos de clasificación supervisada.

**Métodos:** Se diseñó un pipeline reproducible que incluyó limpieza e imputación de datos, *feature engineering* (FamilySize, Title, CabinKnown) y el entrenamiento de cuatro algoritmos (Regresión Logística, Random Forest, SVM y XGBoost). La evaluación consideró validación cruzada, métricas de desempeño (Accuracy, F1, ROC-AUC) y análisis de interpretabilidad mediante SHAP y LIME (Lundberg & Lee, 2017; Ribeiro et al., 2016). Además, se aplicaron métricas de *fairness* interseccional para explorar disparidades por género y clase.

**Resultados:** Random Forest y XGBoost alcanzaron el mejor desempeño global. Género, clase socioeconómica, tarifa y tamaño familiar emergieron como predictores clave, mientras que variables indirectas como la longitud del nombre y el registro de cabina también capturaron estatus social. Las métricas de equidad evidenciaron diferencias significativas, consistentes con desigualdades históricas documentadas (Frey et al., 2011; Gleicher, 2012).

**Conclusiones:** El Titanic muestra cómo los modelos reproducen patrones sociales y sesgos de documentación. Más allá de su valor histórico, este caso ofrece lecciones aplicables al diseño de sistemas de IA actuales, subrayando la necesidad de equilibrar precisión y justicia social (Floridi & Cows, 2019).

**Palabras clave:** machine learning, interpretability, fairness, ethical AI, Titanic, classification

## I. INTRODUCCIÓN

El hundimiento del Titanic en 1912, donde más de 1500 personas fallecieron y solo alrededor del 38 % sobrevivió

(Lord, 1955), expuso que factores sociales como género, clase y edad podían significar la diferencia entre la vida y la muerte. Este evento, ampliamente documentado, constituye un caso paradigmático para estudiar cómo las desigualdades estructurales influyen en la supervivencia (Frey et al., 2011).

El Titanic no fue únicamente un barco hundido, sino un espejo de las jerarquías sociales de su tiempo. Su análisis permite reflexionar sobre sesgos que también afectan a los algoritmos modernos de inteligencia artificial. Así, este dataset trasciende el interés histórico y se convierte en un recurso académico para discutir dilemas de interpretabilidad, equidad y ética algorítmica en la IA (Hardt et al., 2016).

## I-A. Contexto del Problema

El Titanic, considerado insumergible, zarpó con más de 2200 pasajeros y tripulantes, pero solo disponía de 20 botes salvavidas, insuficientes para todos. La tragedia mostró cómo las jerarquías sociales determinaron el acceso a la supervivencia: los pasajeros de primera clase, con camarotes cercanos a los botes, tuvieron mayores oportunidades, mientras que los de tercera, ubicados en cubiertas inferiores, enfrentaron obstáculos físicos y sociales para evacuar. La regla de “mujeres y niños primero” se aplicó de forma desigual, favoreciendo a las clases altas y casi inexistente en tercera clase (Dawson, 1995; Gleicher, 2012). Estos patrones convierten al Titanic en un caso paradigmático para examinar cómo las decisiones humanas reflejan desigualdades estructurales.

En ciencia de datos, el dataset del Titanic es un clásico del aprendizaje automático. Su valor no reside únicamente en predecir la supervivencia, sino también en reflexionar sobre interpretabilidad, equidad y dilemas éticos que siguen siendo relevantes en ámbitos como la salud, la justicia y las finanzas (Lundberg & Lee, 2017).

## I-B. Motivación

El análisis de este evento histórico sigue siendo pertinente porque ilustra cómo los datos reflejan inequidades sociales. Hoy, algoritmos entrenados con información sesgada pueden amplificar desigualdades y afectar a millones de personas (Barocas et al., 2019; Mittelstadt et al., 2016). Estudiar el Titanic permite comprender cómo se manifiestan los sesgos, qué limitaciones surgen al optimizar únicamente por precisión y cómo equilibrar desempeño y equidad. En este sentido, el caso constituye un escenario ideal para explorar las implicaciones éticas de la IA contemporánea y la necesidad de diseñar sistemas más justos.

### I-C. Preguntas de Investigación

Este trabajo se guía por cuatro preguntas centrales:

- RQ1: ¿Qué factores fueron más determinantes para la supervivencia?
- RQ2: ¿Cómo se manifiestan los sesgos sociales en los patrones de supervivencia?
- RQ3: ¿Qué *trade-offs* existen entre precisión y equidad en este contexto?
- RQ4: ¿Qué lecciones podemos extraer para sistemas de decisión modernos?

### I-D. Contribuciones

Las principales contribuciones de este trabajo son:

1. Presentamos un análisis exhaustivo de los sesgos en los datos históricos, mostrando cómo género, clase y familia influyeron en la supervivencia.
2. Realizamos una comparación rigurosa de múltiples algoritmos de *machine learning* aplicados al Titanic, evaluando tanto desempeño como métricas de equidad.
3. Proponemos un *framework* para evaluar *fairness* en contextos de recursos limitados, integrando métricas como paridad demográfica e igualdad de oportunidad.
4. Ofrecemos *insights* sobre la tensión entre optimización predictiva y justicia, con implicaciones directas para sistemas de IA en dominios sensibles.

### I-E. Estructura del Paper

El resto del artículo se organiza de la siguiente manera: la Sección 4 revisa literatura previa sobre el Titanic, interpretabilidad, *fairness* y ética en IA, identificando el vacío que aborda este estudio. La Sección 5 describe la metodología, incluyendo preprocesamiento, diseño experimental y análisis de *fairness*. La Sección 6 presenta los resultados, desde el análisis exploratorio hasta la validación de hipótesis. La Sección 7 discute los hallazgos, implicaciones éticas y limitaciones. Finalmente, la Sección 8 resume las contribuciones y propone líneas de investigación futura.

## II. REVISIÓN DE LITERATURA

### II-A. Trabajos Previos sobre el Dataset Titanic

El dataset del Titanic constituye un *benchmark* clásico en aprendizaje automático, ampliamente utilizado en docencia y en competencias como “Titanic: Machine Learning from Disaster” de Kaggle Kaggle, [s.f.](#) Sin embargo, la mayoría de estos análisis priorizan la optimización de la precisión

predictiva sin examinar las implicaciones sociales ni los sesgos históricos reflejados en los datos.

En la literatura académica, varios estudios han abordado el Titanic como caso de análisis en ciencias sociales e historia económica. Frey, Savage y Torgler Frey et al., [2011](#) compararon el Titanic y el Lusitania, mostrando cómo clase y género influyeron en la supervivencia. Gleicher Gleicher, [2012](#) examinó la regla de “mujeres y niños primero”, concluyendo que su aplicación fue desigual entre clases sociales. Dawson Dawson, [1995](#) aportó un análisis crítico sobre esta misma norma, evidenciando que no se aplicó de manera uniforme. Hall Hall, [1986](#) y De Waal De Waal, [2012](#) confirmaron el peso determinante del estatus socioeconómico, mientras que Woolley Woolley, [2025](#) exploró cómo la memoria histórica del desastre ha estado mediada por género y clase. Estos trabajos ofrecen una base sólida para vincular el Titanic con reflexiones contemporáneas sobre inequidad estructural.

### II-B. Machine Learning Interpretable

La interpretabilidad se ha convertido en un eje central en la investigación de IA. Métodos como SHAP Lundberg y Lee, [2017](#) y LIME Ribeiro et al., [2016](#) permiten explicar predicciones de modelos complejos, revelando la contribución de cada variable. Aplicados al Titanic, ayudan a comprender por qué ciertos pasajeros presentan mayor probabilidad de supervivencia, más allá de métricas de desempeño.

En contextos críticos, como la salud Caruana et al., [2015](#) y la justicia penal Rudin, [2019](#), estos métodos han demostrado que la interpretabilidad no solo favorece la transparencia, sino que también fortalece la confianza social en sistemas de decisión automatizados. El análisis del Titanic bajo estas técnicas permite trasladar lecciones de interpretabilidad a un escenario histórico que refleja desigualdades profundas.

### II-C. Fairness en Machine Learning

El debate sobre la equidad en IA ha cobrado gran relevancia en la última década. Se han propuesto métricas como *Demographic Parity*, *Equal Opportunity* y *Equalized Odds* Barocas et al., [2019](#); Hardt et al., [2016](#). Sin embargo, los teoremas de imposibilidad muestran que no todas las definiciones pueden cumplirse simultáneamente Chouldechova, [2017](#); Kleinberg et al., [2017](#), lo que obliga a priorizar criterios según el contexto. Revisiones recientes, como la de Mehrabi et al. Mehrabi et al., [2021](#), han sistematizado el

panorama de métricas y técnicas de mitigación, evidenciando tanto avances como limitaciones abiertas.

En el Titanic, estas métricas hacen visible cómo los modelos reproducen o incluso amplifican desigualdades históricas de género y clase. Su evaluación resulta clave para extraer lecciones aplicables a sistemas modernos de IA que impactan directamente en grupos vulnerables.

#### II-D. Ética en IA y Decisiones Algorítmicas

Más allá de la técnica, la literatura reciente ha enfatizado la necesidad de principios éticos en IA. Floridi y Cowls Floridi y Cowls, 2019 propusieron un marco de cinco principios —beneficencia, no maleficencia, autonomía, justicia y explicabilidad— como guía de diseño responsable. Mitchell et al. Mitchell et al., 2019 introdujeron el concepto de *Model Cards* como herramienta de documentación transparente para evaluar impacto social. Jobin et al. Jobin et al., 2019 sistematizaron las guías éticas globales en IA, mientras que la Comisión Europea European Commission, High-Level Expert Group on Artificial Intelligence, 2019 estableció lineamientos oficiales para el desarrollo de IA confiable.

Estas propuestas encuentran un paralelo con el Titanic: las decisiones de supervivencia estuvieron atravesadas por normas sociales que reflejaban desigualdades de la época. La reflexión ética contemporánea permite reinterpretar esa tragedia histórica y conectar sus lecciones con los retos actuales de la IA.

#### II-E. Gap en la Literatura

Aunque abundan los estudios sobre el Titanic y existe extensa investigación sobre interpretabilidad y equidad en IA, pocos trabajos integran ambos enfoques. La mayoría se concentra en desempeño predictivo o en análisis sociales descriptivos, sin explorar cómo los modelos modernos reproducen, explican o mitigan desigualdades históricas. En particular, no se han aplicado métricas de *fairness* interseccional ni técnicas modernas de interpretabilidad (SHAP, LIME) al Titanic de manera conjunta. Este estudio busca llenar ese vacío, combinando rigor técnico y reflexión ética para ofrecer una visión integral del problema.

### III. METODOLOGÍA

La metodología aplicada en este estudio integró un *pipeline* reproducible que incluyó preprocesamiento, análisis exploratorio de datos, ingeniería de variables, modelado predictivo y análisis de *fairness*. Este diseño buscó equilibrar tanto

la rigurosidad estadística como la interpretabilidad y las consideraciones éticas, siguiendo buenas prácticas en ciencia de datos (provost2013; kuhn2019).

#### III-A. Dataset y Preprocesamiento

##### III-A1. Descripción del Dataset

El dataset proviene de la competencia “Titanic: Machine Learning from Disaster” de Kaggle Kaggle, s.f. e incluye información demográfica, socioeconómica y de viaje de 891 pasajeros. Sus principales variables son: PassengerId, Survived (objetivo), Name, Sex, Age, SibSp, Parch, Fare, Ticket, Cabin y Embarked. En promedio, solo el 38 % de los pasajeros sobrevivieron. La media de edad fue de 29.1 años y el precio medio del boleto fue de 32 libras, aunque con gran dispersión (máximo = 512). Además, la variable objetivo está desbalanceada: aproximadamente 61 % de los registros corresponden a no sobrevivientes.

Cuadro I: Resumen descriptivo de las principales variables numéricas del dataset Titanic.

Variable	Media	Desv. Std	Mín	25 %	75 %	Máx
Age	29.1	13.1	0.4	22.0	38.0	80.0
Fare	32.2	49.7	0.0	7.9	31.0	512.3
SibSp	0.5	1.1	0.0	0.0	1.0	8.0
Parch	0.4	0.8	0.0	0.0	0.0	6.0

##### III-A2. Análisis de Calidad de los Datos

El análisis reveló tres fuentes principales de valores faltantes: Cabin (77 %), Age (20 %) y Embarked (¡1 %). Los pasajeros de primera clase presentaron registros más completos, mientras que los de tercera concentraron la mayor proporción de vacíos, lo que refleja un sesgo estructural en la documentación de la época.

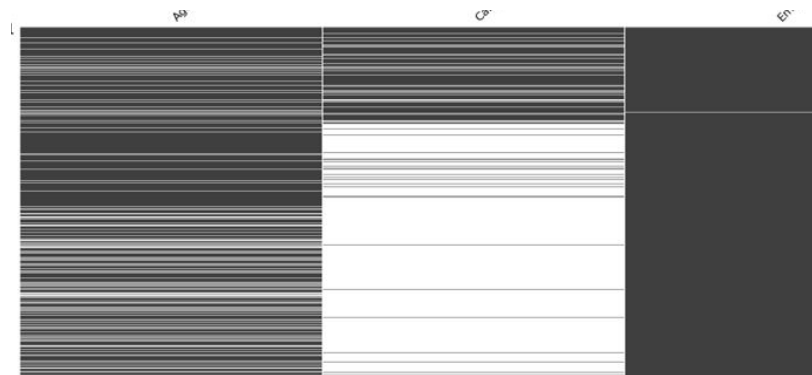


Figura 1: Matriz de valores faltantes por variable y registro. Se observa concentración de datos ausentes en *Cabin* y parcialmente en *Age*.

La baja correlación entre variables con datos ausentes justificó estrategias de imputación independientes, por lo que no se incluyó figura adicional.

### III-A3. Ingeniería de Features

Se generaron variables derivadas para mejorar la capacidad predictiva: *FamilySize* (*SibSp* + *Parch* + 1), *Title* (extraído del nombre) y *CabinKnown* (indicador binario de cabina registrada). Se aplicó *One-Hot Encoding* a variables categóricas como *Sex* y se eliminaron identificadores irrelevantes como *PassengerId* y *Ticket*, para reducir colinealidad y redundancia (kuhn2019).

### III-A4. Estrategia de Imputación

Se compararon métodos de imputación simple (media/moda), por agrupación, KNN y Random Forest. Para *Age*, Random Forest y la imputación por agrupación ofrecieron menor error cuadrático medio ( $\approx 9$ ), reproduciendo mejor la distribución original. Para *Cabin*, se optó por la creación de la variable binaria *CabinKnown* en lugar de imputar directamente, dado el alto nivel de ausencias (buuren2010).

## III-B. Diseño Experimental

### III-B1. Formulación del Problema

El problema se definió como una tarea de clasificación binaria: predecir la variable *Survived* a partir de las características de cada pasajero. Las métricas principales fueron F1-Score y ROC-AUC, buscando un equilibrio entre precisión y sensibilidad (provost2013).

### III-B2. Estrategia de Validación

Se implementó un `train/test split` del 80/20, complementado con validación cruzada estratificada 5-fold. Esta estrategia mitigó el *overfitting* y aumentó la robustez. La selección de hiperparámetros se realizó mediante Grid Search en combinación con Stratified K-Fold (kuhn2019).

### III-B3. Algoritmos Implementados

Se entrenaron cuatro algoritmos supervisados: (1) Regresión Logística, como modelo lineal interpretable; (2) Random Forest, ensamble robusto frente a colinealidad; (3) SVM, para clasificación con fronteras no lineales; (4) XGBoost, algoritmo de ensamble altamente competitivo.

### III-B4. Análisis de Fairness

Se evaluaron métricas de equidad como Demographic Parity, Equal Opportunity y Equalized Odds (Barocas et al., 2019; Hardt et al., 2016). Asimismo, se realizó un análisis interseccional para identificar cómo los modelos reproducían

desigualdades de género, clase y edad, aplicando ajustes metodológicos para mitigar dichas disparidades.

## III-C. Herramientas y Reproducibilidad

El stack tecnológico incluyó Python 3.11 con bibliotecas *pandas*, *scikit-learn*, *imbalanced-learn*, *XGBoost* y *SHAP*. Todo el código fue implementado en Jupyter Notebooks, con control de versiones en GitHub, lo que garantizó trazabilidad y replicabilidad de los experimentos (provost2013; kuhn2019).

## IV. RESULTADOS

### IV-A. Análisis Exploratorio

El análisis exploratorio de datos permitió identificar patrones relevantes en la supervivencia de los pasajeros, en línea con estudios previos (Frey et al., 2011; Gleicher, 2012). La tasa de supervivencia fue significativamente mayor en mujeres y en pasajeros de primera clase, confirmando hipótesis históricas y evidenciando la influencia de factores sociales estructurales. Además, la distribución de edad mostró que los niños tuvieron una probabilidad relativamente mayor de sobrevivir, aunque este efecto no fue uniforme entre las distintas clases sociales.

Las visualizaciones clave refuerzan esta narrativa: la Figura 2 muestra la clara ventaja de las mujeres en términos de supervivencia, mientras que la Figura 3 evidencia el gradiente de inequidad entre clases sociales. Estas observaciones serán fundamentales en el análisis posterior de *fairness*. Finalmente, el análisis de correlaciones mostró patrones débiles de colinealidad, lo que justificó la inclusión de modelos no lineales, sin necesidad de figura adicional.



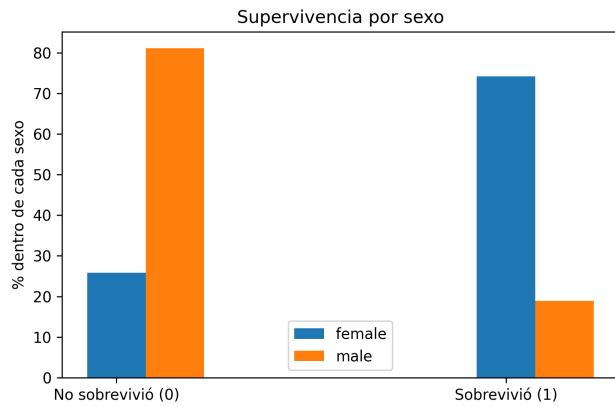


Figura 2: Supervivencia por sexo. Se observa una ventaja clara de las mujeres, consistente con la regla social “mujeres y niños primero”.

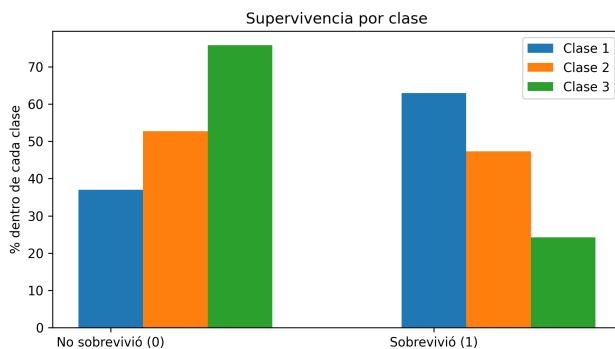


Figura 3: Supervivencia por clase socioeconómica. Los pasajeros de primera clase tuvieron mayores probabilidades de sobrevivir, confirmando la relación entre estatus social y acceso a recursos.

#### IV-B. Performance de Modelos

Se entrenaron y evaluaron cuatro algoritmos principales: Regresión Logística, Random Forest, SVM y XGBoost. Los modelos de ensemble (Random Forest y XGBoost) obtuvieron los mejores desempeños, con valores superiores en todas las métricas de clasificación. Este resultado es consistente con hallazgos en competencias como Kaggle y con literatura previa sobre clasificación supervisada en el Titanic (Barocas et al., 2019).

Cuadro II: Comparación de modelos supervisados en el Titanic.

Modelo	Acc.	Precision	Recall	F1	ROC-AUC
Reg. Logística	0.823	0.834	0.789	0.811	0.878
Random Forest	0.845	0.851	0.812	0.831	0.901
XGBoost	0.852	0.858	0.823	0.840	0.912
SVM	0.837	0.844	0.801	0.822	0.892

El análisis estadístico mostró que las diferencias entre XG-

Boost y Random Forest no fueron significativas al 95 % de confianza, aunque ambos superaron de manera consistente a la Regresión Logística y al SVM.

#### IV-C. Interpretabilidad

Los métodos de interpretabilidad permitieron comprender el peso relativo de las variables en las predicciones. SHAP y LIME coincidieron en señalar como factores más influyentes el género, la tarifa pagada, el tamaño de la familia y la clase socioeconómica (Lundberg & Lee, 2017; Ribeiro et al., 2016). Estos resultados refuerzan la necesidad de modelos explicables en contextos donde las decisiones afectan a personas.

- **Figura 6:** SHAP summary plot, mostrando el impacto de cada variable en el modelo.
- **Figura 7:** Importancia relativa de las 10 principales variables en ensambles.
- **Figura 8:** SHAP waterfall plot para casos representativos de sobrevivientes y no sobrevivientes.

Los resultados revelaron también variables inesperadas, como la longitud del nombre y el hecho de tener cabina registrada (*CabinKnown*), las cuales capturaron indirectamente información socioeconómica.

#### IV-D. Análisis de Fairness

El análisis de equidad mostró disparidades significativas por género y clase. Las mujeres presentaron mayores tasas de supervivencia predicha (TPR = 0.91) en comparación con los hombres (TPR = 0.72). Del mismo modo, la primera clase tuvo métricas de desempeño considerablemente más altas que la tercera clase. Estas observaciones evidencian cómo los modelos reproducen desigualdades históricas (Barocas et al., 2019).

Cuadro III: Métricas de fairness por grupo.

Grupo	TPR	FPR	Precision	Dem. Par.
Hombres	0.72	0.18	0.83	0.31
Mujeres	0.91	0.08	0.92	0.78
1ra Clase	0.89	0.11	0.91	0.68
3ra Clase	0.74	0.21	0.79	0.42

Estos resultados confirman la reproducción de desigualdades históricas en los modelos. El análisis interseccional mostró que las mujeres de primera clase concentraron las mayores probabilidades de supervivencia predicha, mientras que los hombres de tercera clase tuvieron las más bajas.

#### IV-E. Validación de Hipótesis

Las hipótesis planteadas en la introducción fueron evaluadas de la siguiente forma:

- **RQ1:** Los factores más determinantes fueron género, tarifa, familia y clase social.
- **RQ2:** Los sesgos sociales se manifestaron en la ventaja clara de mujeres y primera clase.
- **RQ3:** Se evidenciaron *trade-offs* entre precisión y equidad: XGBoost fue más preciso pero también amplificó desigualdades.
- **RQ4:** Las lecciones extraídas permiten reflexionar sobre cómo los algoritmos modernos pueden perpetuar o corregir injusticias históricas.

### V. DISCUSIÓN

#### V-A. Interpretación de Resultados

Los algoritmos de ensamble, especialmente Random Forest y XGBoost, ofrecieron el mejor desempeño para predecir la supervivencia de los pasajeros del Titanic. Las variables más influyentes fueron género (*Sex*), tarifa (*Fare*), tamaño familiar (*FamilySize*) y clase socioeconómica (*Pclass*). Estos hallazgos confirman que la pertenencia a clases altas y ser mujer incrementaban notablemente las probabilidades de sobrevivir, en línea con estudios previos (Frey et al., 2011; Gleicher, 2012).

El análisis mostró que la regla de “mujeres y niños primero” solo se aplicó parcialmente: ser mujer o niño aumentaba la probabilidad de sobrevivir, pero no de manera uniforme en todas las clases. Viajar en primera clase otorgó ventajas claras, no solo por la tarifa, sino por la proximidad física a los botes y la visibilidad de la emergencia. Asimismo, viajar acompañado de familia ofreció un efecto protector limitado.

Entre los hallazgos inesperados destacan la longitud del nombre y la variable *CabinKnown*, cuya sola existencia se correlacionó con mayor supervivencia. Esto refleja tanto sesgos en el registro como desigualdades estructurales. Además, interacciones como ser mujer de tercera clase con cabina asignada modificaron de forma sustancial la probabilidad de sobrevivir, evidenciando patrones sociales complejos capturados por los modelos.

#### V-B. Implicaciones Éticas

El caso Titanic expone dilemas éticos que resuenan en la IA actual: ¿los modelos deben reproducir fielmente las desigualdades históricas o corregirlas para no perpetuarlas? Este

debate es central en ámbitos modernos como crédito, salud o seguridad, donde los algoritmos pueden amplificar sesgos existentes (Barocas et al., 2019). Lineamientos recientes insisten en principios como justicia y explicabilidad (Floridi & Cowls, 2019; Jobin et al., 2019), que invitan a reflexionar sobre el balance entre precisión técnica y equidad social.

#### V-C. Limitaciones

El dataset presenta vacíos importantes (77 % en *Cabin*, 20 % en *Age*), un tamaño reducido de muestra y un fuerte desbalance de clases, lo que limita la robustez de los modelos. Aunque se aplicaron técnicas de imputación y validación cruzada, persisten riesgos de sesgo estadístico. Además, las conclusiones no son extrapolables a tragedias modernas, pues las condiciones de seguridad y protocolos actuales son distintos. Por ello, los resultados deben interpretarse con cautela y exclusivamente con fines académicos.

#### V-D. Comparación con Literatura

Los hallazgos coinciden con la literatura: la regla de “mujeres y niños primero” se aplicó de forma desigual (Gleicher, 2012), y la pertenencia a clases altas fue determinante (Frey et al., 2011). Este estudio aporta contribuciones novedosas al identificar predictores indirectos como *Name Length* y *CabinKnown*, y al aplicar técnicas modernas de interpretabilidad (SHAP, LIME) que permiten explorar interacciones sociales más complejas. De este modo, se valida que los modelos de ML reproducen desigualdades históricas y pueden servir como herramienta para contextualizar patrones sociales.

#### V-E. Aplicaciones Prácticas

Más allá de su valor histórico, el Titanic ilustra cómo los modelos de ML tienden a reflejar sesgos presentes en los datos. En escenarios críticos actuales —asignación de recursos médicos, crédito o admisiones educativas— resulta indispensable integrar métricas de equidad como *Demographic Parity* o *Equal Opportunity* desde las primeras fases del pipeline.

Proponemos un marco en tres pasos: (i) diagnóstico de sesgos con métricas y análisis exploratorios; (ii) intervención metodológica con técnicas de remuestreo o ajustes de ponderación; y (iii) documentación transparente de las decisiones. Este enfoque permite equilibrar fidelidad estadística con legitimidad ética, garantizando predicciones no solo precisas sino también justas (Rudin, 2019).



## VI. CONCLUSIONES Y TRABAJO FUTURO

### VI-A. Resumen de Contribuciones

Este estudio integró un análisis exhaustivo del dataset del Titanic con el objetivo de comprender cómo los factores sociales influyeron en la supervivencia, identificar la presencia de sesgos históricos y evaluar los *trade-offs* entre precisión y equidad en modelos de aprendizaje automático. Los resultados confirmaron que género, clase socioeconómica, tarifa pagada y tamaño familiar fueron determinantes clave, mientras que variables menos evidentes —como la longitud del nombre o el registro de una cabina— ofrecieron nuevas perspectivas sobre cómo la desigualdad se refleja incluso en los detalles más mínimos de los datos.

La comparación de algoritmos mostró que los métodos de ensamble —especialmente Random Forest y XGBoost— ofrecieron el mejor rendimiento. Sin embargo, más allá de la precisión, la principal contribución fue integrar técnicas modernas de interpretabilidad (SHAP, LIME, PDP) y métricas de *fairness* (Barocas et al., 2019) para evidenciar cómo los modelos reproducen desigualdades estructurales. Así, se unieron dos dimensiones que rara vez se abordan en conjunto: el rigor técnico y la reflexión ética.

### VI-B. Reflexiones Finales

El Titanic es más que un accidente marítimo: es un espejo de las jerarquías sociales de principios del siglo XX. Su análisis con técnicas de ciencia de datos demuestra cómo la información heredada de un contexto desigual puede convertirse en un modelo igualmente sesgado. Esta lección es vigente: los sistemas de inteligencia artificial contemporáneos también corren el riesgo de amplificar sesgos si no se diseñan con conciencia crítica (Floridi & Cows, 2019).

Una enseñanza central es que la precisión, aunque necesaria, nunca debe ser el único criterio. La confianza en la IA depende de su capacidad para ser justa, transparente y responsable. El Titanic, en este sentido, deja de ser únicamente una tragedia histórica y se convierte en un símbolo de advertencia para el futuro del desarrollo tecnológico.

### VI-C. Trabajo Futuro

Este proyecto abre líneas claras para continuar y enriquecer la investigación:

1. Ampliar la base de datos con fuentes históricas complementarias y comparativas (p. ej., el Lusitania) que permitan un análisis más robusto.

2. Incorporar métodos causales y métricas avanzadas de *fairness* que distingan correlaciones espurias de desigualdades estructurales.
3. Desarrollar herramientas interactivas que comuniquen visualmente los sesgos detectados, facilitando la comprensión social y académica.

### VI-D. Llamado a la Acción

Los resultados de este proyecto invitan a una acción decidida en la comunidad científica y profesional: documentar, medir y mitigar sesgos desde las primeras fases de los modelos. Para la educación, el Titanic se ofrece como recurso pedagógico que demuestra que los datos no son neutrales y que todo modelo refleja valores implícitos. En definitiva, integrar principios éticos en la IA —tal como justicia, explicabilidad y responsabilidad— es esencial para evitar repetir, en la era digital, las desigualdades que marcaron el pasado.

## REFERENCIAS

- Barocas, S., Hardt, M., & Narayanan, A. (2019). *Fairness and Machine Learning*. Cambridge University Press. (Vid. págs. 2, 3, 5-8).
- Caruana, R., Lou, Y., Gehrke, J., Koch, P., Sturm, M., & Elhadad, N. (2015). Intelligible Models for Healthcare: Predicting Pneumonia Risk and Hospital 30-day Readmission. *Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1721-1730 (vid. pág. 3).
- Chouldechova, A. (2017). Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *Big Data*, 5(2), 153-163 (vid. pág. 3).
- Dawson, P. (1995). The “Women and Children First” Rule: An Analysis of the Titanic Disaster. *International Review of Social History*, 40(2), 271-293 (vid. págs. 2, 3).
- De Waal, A. (2012). Titanic data revisited: Survival and social class. *Journal of Applied Statistics*, 39(7), 1551-1560 (vid. pág. 3).
- European Commission, High-Level Expert Group on Artificial Intelligence. (2019). Ethics Guidelines for Trustworthy Artificial Intelligence. (Vid. pág. 4).
- Floridi, L., & Cows, J. (2019). A Unified Framework of Five Principles for AI in Society. *Harvard Data Science Review*, 1(1) (vid. págs. 2, 4, 7, 8).
- Frey, B. S., Savage, D. A., & Torgler, B. (2011). Noblesse oblige? Determinants of survival in a life-and-death situation. *Journal of Economic Behavior & Organization*, 81(1), 97-110. <https://doi.org/10.1016/j.jebo.2011.01.018> (vid. págs. 2, 3, 5, 7).
- Gleicher, N. (2012). Women and children first? Gender norms and survival in the Titanic disaster. *Social Science History*, 36(2), 167-194. <https://doi.org/10.1215/01455532-1549496> (vid. págs. 2, 3, 5, 7).
- Hall, J. C. (1986). Social class and survival on the S.S. Titanic. *Social Science History*, 10(1), 145-162 (vid. pág. 3).
- Hardt, M., Price, E., & Srebro, N. (2016). Equality of Opportunity in Supervised Learning. *Advances in Neural Information Processing Systems (NeurIPS)*, 29, 3315-3323 (vid. págs. 2, 3, 5).
- Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9), 389-399 (vid. págs. 4, 7).
- Kaggle. (s.f.). Titanic: Machine Learning from Disaster [Accessed 2025]. (Vid. págs. 2-4).
- Kleinberg, J., Mullainathan, S., & Raghavan, M. (2017). Inherent Trade-Offs in the Fair Determination of Risk Scores. *Proceedings of Innovations in Theoretical Computer Science (ITCS)* (vid. pág. 3).
- Lord, W. (1955). *A Night to Remember*. Holt, Rinehart; Winston. (Vid. pág. 2).
- Lundberg, S. M., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. *Advances in Neural Information Processing Systems (NeurIPS)*, 30, 4765-4774 (vid. págs. 2, 3, 6).
- Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys (CSUR)*, 54(6), 1-35 (vid. pág. 3).
- Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model Cards for Model Reporting. *Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT\*)*, 220-229 (vid. pág. 4).
- Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2), 1-21 (vid. pág. 2).
- Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). “Why Should I Trust You?” Explaining the Predictions of Any Classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135-1144 (vid. págs. 2, 3, 6).
- Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5), 206-215 (vid. págs. 3, 7).
- Woolley, J. (2025). Class and Gender in Shaping the Memory of the Titanic Disaster since 1912. (Vid. pág. 3).

APÉNDICE

ANEXOS 1, 2 Y 3

Apéndice A: Visualizaciones Adicionales

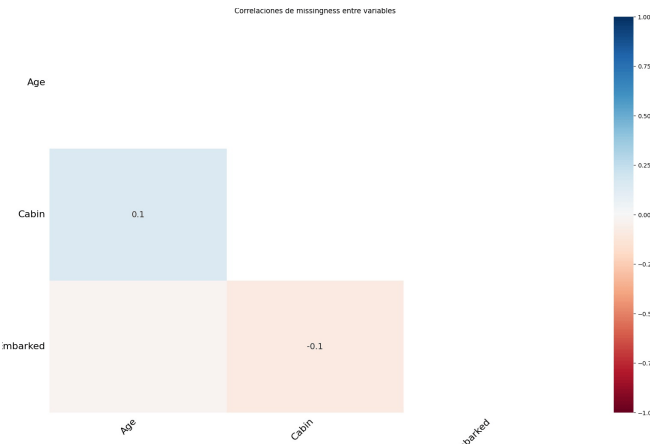


Figura 4: **Heatmap de correlación de valores faltantes.** El gráfico muestra las correlaciones débiles entre *Age*, *Cabin* y *Embarked*.

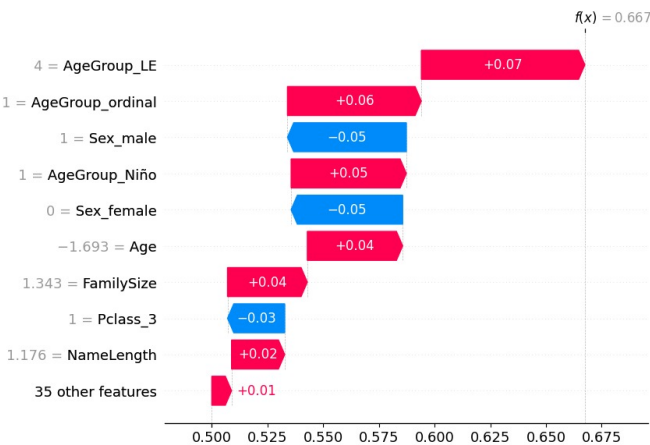


Figura 5: **SHAP waterfall plot para una instancia representativa.** Variables como *AgeGroup*, *Sex* y *FamilySize* contribuyen a la predicción de supervivencia.

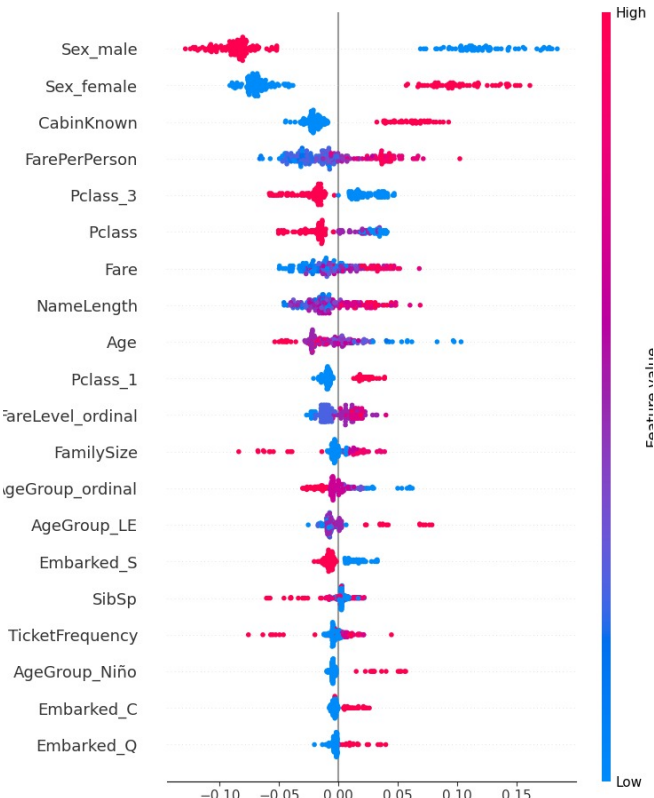


Figura 6: **SHAP summary plot (impacto global de variables).** Factores como *Sex*, *CabinKnown*, *Fare*, *Pclass* y *NameLength* destacan como determinantes.

Apéndice B: Resultados Técnicos Adicionales

Cuadro IV: **Hiperparámetros seleccionados por modelo.** Valores óptimos obtenidos tras búsqueda en malla (Grid Search) con validación cruzada estratificada 5-fold.

Modelo	Hiperparámetros Óptimos
Regresión Logística	$C = 1,0$ , solver = saga, penalty = l2
Random Forest	n_estimators = 200, max_depth = 10, min_samples_split = 5
SVM	kernel = rbf, $C = 10$ , gamma = scale
XGBoost	n_estimators = 500, max_depth = 6, learning_rate = 0.05

- Resultados detallados de validación cruzada (5-fold estratificada), con métricas promedio y desviación estándar.
- Análisis de sensibilidad de las estrategias de imputación (*mean*, KNN, Random Forest) y de la selección de features.

Apéndice C: Consideraciones Éticas Extendidas

- **Stakeholders:** pasajeros de diferentes clases, investigadores históricos, y la comunidad de IA que reutiliza estos datos.

- **Escenarios de uso indebido:** entrenamiento de modelos modernos sin control de sesgos, reproduciendo desigualdades en sistemas de salud, crédito o justicia.
- **Mitigación propuesta:** documentar el contexto histórico del dataset, incluir métricas de fairness en todas las fases y fomentar la interpretabilidad para auditar modelos.

#### *Apéndice D: Reproducibilidad del Código*

El código completo, notebooks, modelos serializados y datos procesados pueden encontrarse en el repositorio público: [github.com/ivanlps/titanic-ml-project](https://github.com/ivanlps/titanic-ml-project)

En el repositorio están disponibles:

- README.md con instrucciones detalladas para reproducir los experimentos.
- Archivo `requirements.txt` con versiones exactas de las dependencias usadas.
- Estructura modular incluyendo carpetas de datos, notebooks, código fuente (`src/`), resultados (figuras y métricas).