El rol inesperado del acompañamiento en la supervivencia humana

Regina Reyes Juárez – **A01275790**
Nadia Salgado Álvarez – **A01174509**
Gilberto Ángel Camacho Lara – **A01613895**
Santiago Miguel Lozano Cedillo – **A01198114**
Tecnológico de Monterrey, Escuela de Ingeniería y Ciencias
Equipo: 3
Viernes 12 de septiembre de 2025

***Resumen*—El hundimiento del Titanic en 1912 no solo marcó una tragedia histórica, sino que también puso en evidencia cómo factores sociales podían marcar la diferencia entre la vida y la muerte. Más de un siglo después, este caso sigue siendo un escenario único para aplicar ciencia de datos y reflexionar sobre qué condiciones influyen en la supervivencia en situaciones de emergencia. En este trabajo nos centramos en una pregunta concreta: ¿tenían más probabilidades de sobrevivir los pasajeros que viajaban acompañados frente a quienes lo hacían solos? Para responderla, analizamos el dataset del Titanic con un pipeline de preprocesamiento, ingeniería de variables y modelos de aprendizaje automático como Regresión Logística, Random Forest y XGBoost, evaluados y comparados en su desempeño. Los resultados muestran que, aunque los modelos predicen con alta exactitud (ROC-AUC ¿0.90), viajar acompañado sí parece haber representado una ventaja en las probabilidades de supervivencia, junto con otros factores como el género y la clase social. Este hallazgo no solo ayuda a entender mejor lo ocurrido en 1912, sino que también abre una reflexión actual: los algoritmos modernos pueden ser muy precisos, pero si no consideramos factores humanos como el apoyo social, corremos el riesgo de dejar fuera variables clave en la toma de decisiones.**

I. Introduccón

El hundimiento del Titanic en 1912 no solo simboliza una tragedia marítima, sino también un ejemplo de cómo las dinámicas sociales pueden influir en la supervivencia en situaciones extremas. Más de un siglo después, este episodio sigue capturando la atención de investigadores, no únicamente por su dimensión histórica, sino porque ofrece un escenario único para reflexionar sobre la manera en que factores humanos —como el acompañamiento, la edad o el género— condicionan las decisiones en contextos de crisis. Este caso se ha convertido en un referente en la enseñanza y aplicación de técnicas de ciencia de datos, donde la interpretación de patrones sociales adquiere un valor particular al compararse con dilemas actuales de la inteligencia artificial.

En este estudio proponemos una hipótesis específica: que viajar acompañado aumentaba las probabilidades de sobrevivir frente a quienes lo hacían solos. Más allá de confirmar un fenómeno del pasado, esta pregunta cobra relevancia porque conecta con debates contemporáneos sobre la equidad en sistemas algorítmicos, los cuales, al igual que los protocolos de evacuación de hace un siglo, deben balancear eficiencia con responsabilidad social. Para poner a prueba la hipótesis diseñamos un análisis que combina preprocesamiento de datos, ingeniería de variables y distintos modelos de aprendizaje automático, cuyos resultados permiten no solo evaluar la validez de la hipótesis, sino también abrir una discusión sobre cómo los factores humanos deben considerarse en el desarrollo de modelos predictivos modernos. El resto del documento se organiza de la siguiente manera: la Sección 2 revisa la literatura existente, la Sección 3 describe la metodología, la Sección 4 presenta los resultados y su discusión, y la Sección 5 cierra con conclusiones y líneas de trabajo futuro.

II. Revisión de literatura

El Titanic ha sido ampliamente analizado en estudios académicos y competencias de Kaggle. Gupta et al. (2018) mostraron que sexo y edad son predictores clave con modelos como regresión logística y Random Forest, mientras que Ekinci et al. (2018) compararon múltiples algoritmos, destacando la superioridad de los ensambles. En Kaggle, competiciones como “Titanic: Machine Learning from Disaster” y la “Titanic Survival Prediction Challenge 2020” sirvieron como laboratorios para probar feature engineering y modelos avanzados. Sin embargo, la mayoría de trabajos se han enfocado en mejorar precisión, dejando de lado la interpretabilidad y el análisis de sesgos. La interpretabilidad busca explicar por qué los modelos generan ciertas predicciones (CII.IA, 2020), lo cual es clave en algoritmos complejos como XGBoost o Random Forest. Métodos como SHAP descomponen la predicción en contribuciones de cada variable (Awan, 2023), mientras que LIME aproxima el modelo localmente con explicaciones simples (Molnar, n.d.). Estas técnicas han demostrado utilidad en salud, finanzas y justicia penal, facilitando auditorías y comunicación de resultados. No obstante, aún presentan limitaciones como inestabilidad de resultados y dependencia de la interpretación humana. El concepto de fairness se centra en garantizar que las decisiones algorítmicas no discriminen entre grupos sociales. Existen distintas definiciones, como igualdad de oportunidades o paridad demográfica, que a menudo entran en conflicto entre sí, como muestran los impossibility theorems. Para abordar

estos dilemas, se han desarrollado métodos de mitigación de sesgo en tres niveles: preprocesamiento de datos, ajustes en el entrenamiento y correcciones en la salida del modelo (Banco Interamericano de Desarrollo, 2023). Aun así, lograr un equilibrio perfecto entre precisión y equidad sigue siendo un desafío abierto. La ética en IA busca que los modelos sean justos, transparentes y responsables. Marcos como el de la Unión Europea promueven principios de explicabilidad y auditoría. Casos como COMPAS, criticado por sesgos raciales en justicia penal (Díaz, 2018), muestran la necesidad de diseñar algoritmos que no solo sean precisos, sino también respetuosos con los derechos de las personas. Los estudios sobre el Titanic se han centrado en alcanzar la mayor precisión predictiva posible, sin integrar de manera sistemática consideraciones de interpretabilidad, fairness o ética. Esto ha dejado un vacío en la literatura sobre cómo estos enfoques podrían complementar el análisis clásico del dataset. En nuestro caso, el objetivo no es llenar ese vacío, sino desarrollar un modelo de predicción de supervivencia guiado por el lema histórico de "mujeres y niños primero", explorando cómo este principio se refleja en los datos y qué tan bien puede ser capturado por técnicas de Machine Learning.

III. Metodología

La metodología de este proyecto se centra en el análisis de datos del histórico caso del Titanic, abordando de manera estructurada todo el proceso de desarrollo de un modelo de ciencia de datos. Comienza con un análisis exploratorio de las variables disponibles, evaluando su relación con la hipótesis principal del estudio. A partir de esta base, se realiza un preprocesamiento detallado que incluye la imputación de valores faltantes, el escalado de variables y otras técnicas necesarias para preparar los datos de manera adecuada.

Posteriormente, se abordan las técnicas de modelado, incluyendo la selección de algoritmos de aprendizaje automático, las métricas utilizadas para evaluar el rendimiento del modelo y las herramientas empleadas durante todo el proceso. También se consideran aspectos relacionados con la eficiencia, como los costos computacionales asociados a cada enfoque.

Este enfoque metodológico avanza desde una perspectiva exploratoria general hacia un nivel más técnico, detallando las herramientas y recursos computacionales utilizados, con el fin de garantizar la trazabilidad, replicabilidad y solidez del modelo desarrollado

El análisis se basa en el conjunto de datos de pasajeros del Titanic, que consta de 891 registros. Inicialmente, se realizó una selección de características para aislar las variables más relevantes para el análisis de supervivencia. Las columnas `PassengerId`, `Name`, `Ticket` y `Cabin` fueron excluidas del análisis por las siguientes razones:

- **`PassengerId` y `Name`**: Son identificadores únicos y no contienen información predictiva relevante para el modelo de supervivencia.
- **`Ticket`**: Representa un identificador de billete que, al igual que los identificadores de pasajero, no se considera una variable con impacto significativo en la probabilidad de supervivencia.
- **`Cabin`**: Esta columna presentaba un alto porcentaje de valores faltantes (aproximadamente el 77 %), lo que podría introducir ruido o sesgo en el análisis.
- **`Embarked`**: Aunque la ubicación de embarque podría tener cierta correlación con la clase socioeconómica, esta relación ya está mejor capturada por la variable `Pclass`. Por lo tanto, se optó por descartar esta variable para simplificar el modelo y evitar redundancia.

Las variables seleccionadas para el análisis principal fueron **`Age`**, **`Fare`**, **`Pclass`**, **`Sex`**, **`SibSp`** y **`Parch`**.

Las estadísticas descriptivas de las variables numéricas `Age` y `Fare` se presentan en el cuadro **??**.

Cuadro I
Resumen de la Ingeniería de Características

Nueva Característica	Característica Original	Método
Acompañante	`SibSp`, `Parch`	One-Hot Encoding
Edad codificada	`Age`	One-Hot Encoding
Género codificado	`Sex`	One-Hot Encoding
Clase codificada	`Pclass`	Codificación ordinal (labeling)
Tarifa logarítmica	`Fare`	Función matemática (logaritmo)

Tras la selección de variables, se evaluó la calidad de los datos para las columnas restantes. Se identificó que solo la variable `Age` contenía valores faltantes. Para examinar el patrón de estos datos, se realizó un análisis de los valores nulos en relación con la variable objetivo, `Survived`.

La Figura 1 ilustra la distribución de los valores faltantes en la columna `Age` para los grupos de no sobrevivientes y sobrevivientes.

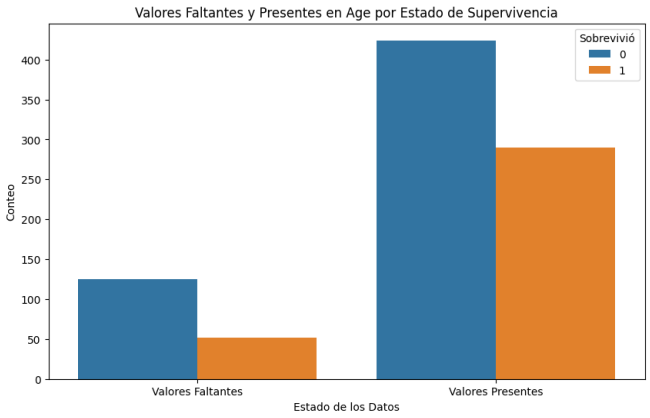


Figura 1. Patrón de valores faltantes en `Age` para Sobrevivientes vs. No Sobrevivientes

El análisis reveló un patrón notable: una mayor proporción de valores faltantes en `Age` se encuentra entre los pasajeros que no sobrevivieron. Este hallazgo es relevante para el preprocesamiento, ya que sugiere que el mecanismo de los datos faltantes no es aleatorio y podría estar relacionado con la variable de supervivencia. Este patrón podría ser considerado al elegir una estrategia de imputación de datos junto con las estadísticas descriptivas del cuadro **??**.

En el dataset se presenta un desbalance en la supervivencia, donde la mayoría de los sobrevivientes fueron mujeres. Esto tiene sentido, ya que el lema era "mujeres y niños primero", por lo que este desbalance no representa un sesgo, sino un reflejo de la situación histórica. En términos porcentuales, aproximadamente el 74 % de las mujeres sobrevivieron, mientras que el porcentaje en hombres fue considerablemente menor.

Como otra modificación, se eliminaron outliers mediante el rango intercuartílico (IQR) para tener rangos más definidos en las columnas `Age`, `SibSp`, `Parch` y `Fare`. Además, se aplicó un método de `StandardScaler` en la columna `Fare`, con el objetivo de que todas las variables numéricas estén escaladas dentro del mismo rango, facilitando posibles transformaciones y garantizando consistencia en el análisis.

Para mejorar la capacidad predictiva del modelo, se aplicaron diversas técnicas de ingeniería de características. El objetivo fue transformar las variables categóricas y continuas seleccionadas en un formato óptimo para el análisis, permaneciendo las ventajas tanto de la ubicación física como del lema de niños/niñas y mujeres primero.

A continuación, se detalla la lógica de cada transformación realizada:

- **Acompañante**: Se creó una variable binaria a partir de `SibSp` y `Parch` para identificar de forma simple si los pasajeros viajaban con familiares o solos, donde 0 indicaba que no aparecía ningún acompañante en `Parch` ni en `SibSp`, y 1 indicaba que al menos había uno en estas dos variables.
- **Edades**: Se categorizó la variable continua `Age` para niños menores a 12 años y otros por encima de esta edad, creando grupos que permitieran simplificar la edad en categorías, para transformarla posteriormente en una variable binaria donde 1 representaba niño/niña y 0 otros; esto era relevante para la hipótesis.
- **Género**: La variable `Sex` se transformó en binaria para cuantificar el género, donde 1 correspondía a femenino y 0 a masculino, considerando su impacto en el modelo.
- **Clase**: La variable `Pclass` se codificó de forma ordinal para reflejar la ubicación física en la embarcación, donde 3 era primera clase, 2 segunda y 1 tercera, relevante a la hora de la evacuación.
- **Tarifa**: La variable `Fare`, al tener un rango tan amplio en cuanto a las tarifas, se decidió transformar aplicando una función matemática para reducir su varianza mediante un logaritmo.

A continuación, el cuadro II resume las características originales y el método de transformación aplicado.

Cuadro II
RESUMEN DE LA INGENIERÍA DE CARACTERÍSTICAS

Nueva Característica	Característica Original	Método
Acompañante	`SibSp`, `Parch`	One-Hot Encoding
Edad codificada	`Age`	One-Hot Encoding
Género codificado	`Sex`	One-Hot Encoding
Clase codificada	`Pclass`	Codificación Ordinal (labeling)
Tarifa logarítmica	`Fare`	Función matemática (logaritmo)

En cuanto a la imputación de datos, solo fue necesario en la variable `Age`, ya que era la única que presentaba valores faltantes. Se decidió realizar la imputación mediante la mediana, dado que en la figura 2 se observa que su mínimo es 0 y su máximo 80, con una media de 30. Considerando el contexto histórico, donde existía desigualdad económica, se optó por mantener esa media mediante la imputación con la mediana, ya que los casos de personas mayores o con mejor economía eran poco frecuentes.

Además, después de la imputación se observa un aumento de los datos en el rango de edad alrededor de los 20 y 40 años, reflejando la intención de conservar este rango. Asimismo, en la figura 2 se puede ver la comparación entre la distribución de edad original y la imputada.

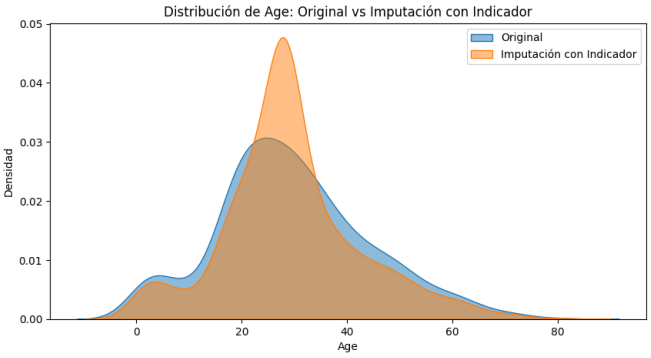


Figura 2. Distribución de Age: Original vs Imputación con Indicador.

La intención de realizar las transformaciones a las variables está alineada con nuestra hipótesis: si los pasajeros viajaban acompañados, aumentaban sus probabilidades de sobrevivir en el Titanic. Además, el lema también está relacionado con el problema, por lo que decidimos considerarlo al momento de la transformación.

Como métricas de evaluación utilizamos el ROC y el F1 score. El ROC es considerado más confiable que la accuracy, ya que esta última es más sensible a clases desbalanceadas, mientras que el ROC refleja mejor el desempeño real del modelo. Por su parte, el F1 score permite evaluar el balance

entre sensibilidad y precisión, lo que es especialmente útil cuando las clases están desbalanceadas.

Para complementar las predicciones, también utilizamos una matriz de confusión, que nos permitió identificar los verdaderos positivos, falsos positivos, verdaderos negativos y falsos negativos, evaluando así de manera integral el desempeño del modelo.

El objetivo es preparar, entrenar y evaluar los modelos bajo condiciones controladas y comparables, asegurando que las transformaciones de las variables y la división de los datos permitan medir de manera justa su desempeño. En la figura 3 se observa el workflow que va desde los datos originales hasta la comparación de modelos mediante métricas.

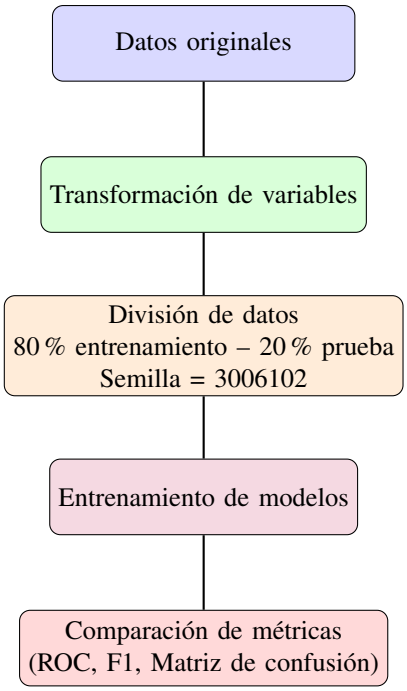


Figura 3. Diagrama del proceso de entrenamiento y evaluación de modelos.

Los algoritmos implementados fueron, como línea base, una regresión logística, dado que no existía una alta correlación entre las variables como se observa en la figura 5 y al tratarse de un problema de clasificación, era necesario considerar mejores alternativas. Asimismo, teniendo en cuenta el desbalance por género, se optó por probar un modelo de Random Forest, aplicando técnicas de búsqueda de hiperparámetros como GridSearchCV y RandomizedSearchCV.

- **Métricas seleccionadas**: Se utilizaron métricas como sensibilidad (recall), soporte (support) y precisión (precision) para evaluar el desempeño del modelo en los distintos grupos protegidos.
- **Grupos protegidos**: Los grupos protegidos considerados, por contexto histórico y social, son niños/niñas y mujeres, quienes tenían prioridad en la supervivencia.
- **Trade-offs considerados**: En este contexto, cualquier desbalance reflejado por las métricas no se considera necesariamente negativo, ya que representa fielmente el comportamiento histórico observado y el lema *"mujeres y niños primero"*.

Como principal herramienta, se utilizó Google Colab, donde se emplearon librerías de machine learning, manejo de datos y visualizaciones para el análisis y toma de decisiones. Los datasets fueron obtenidos de Kaggle en formato CSV.

En cuanto al costo computacional, considerando que el dataset tiene aproximadamente 900 filas, no se requiere de un procesamiento de alto costo. Al utilizar la pipeline 3 diseñada, es posible obtener resultados precisos y eficientes sin necesidad de recursos computacionales intensivos.

IV. Métodos

Para abordar el problema de predicción de supervivencia en el Titanic, se seleccionaron dos métodos principales de aprendizaje supervisado: **regresión logística** y **Random Forest**. La elección de estos métodos se fundamenta en las características del conjunto de datos y los objetivos del análisis:

- **Regresión logística:** Se utilizó como modelo lineal de referencia, debido a la ausencia de correlaciones lineales fuertes entre las variables, tal como se evidencia en la matriz de correlación (Figure 5). Este método permite interpretar fácilmente el efecto de cada variable en la probabilidad de supervivencia, facilitando la validación de la hipótesis sobre el impacto de viajar acompañado.
- **Random Forest:** Este modelo de ensamble se empleó para capturar interacciones no lineales y patrones complejos entre las variables, que la regresión logística podría no detectar. Además, permite evaluar la importancia relativa de cada característica, incluyendo la variable **acompañamiento**, y proporciona un rendimiento más robusto en términos de precisión y curva ROC.

Ambos métodos se aplicaron sobre un conjunto de datos preprocesado que incluyó:

1. Ingeniería de características (*feature engineering*) para generar variables binarias y categóricas relevantes como **acompañamiento**, **edad codificada** y **género codificado**.
2. Imputación de valores faltantes, escalado de variables y eliminación de outliers, asegurando la calidad de los datos.
3. División de los datos en conjuntos de entrenamiento y prueba con una proporción 80/20 y una semilla fija para garantizar reproducibilidad.

La combinación de estos métodos y técnicas de preprocesamiento permite evaluar de manera confiable la hipótesis planteada y comparar el impacto de cada variable sobre la supervivencia de los pasajeros.

V. Resultados y Discusión

Al analizar la supervivencia por clase y género en la Figure 4, se observa que la probabilidad de sobrevivir fue significativamente mayor en las mujeres, especialmente en las clases 1 y 2. Sin embargo, en la clase 3 este porcentaje disminuye de forma notable, reflejando que las condiciones

socioeconómicas también jugaron un papel decisivo en la probabilidad de supervivencia. Estos hallazgos refuerzan la idea de que la evacuación no se regía únicamente por el lema social de *"mujeres y niños primero"*, sino que este se encontraba condicionado por la posición social y el acceso físico a los recursos de evacuación.

A partir de esta evidencia, en el tratamiento de los datos se decidió otorgar mayor peso a las variables **género** y **edad**, priorizando a mujeres y niños por su relevancia tanto histórica como predictiva. De esta manera, se buscó capturar en el modelo no solo las decisiones técnicas, sino también los factores culturales y sociales que influían en las dinámicas de supervivencia.

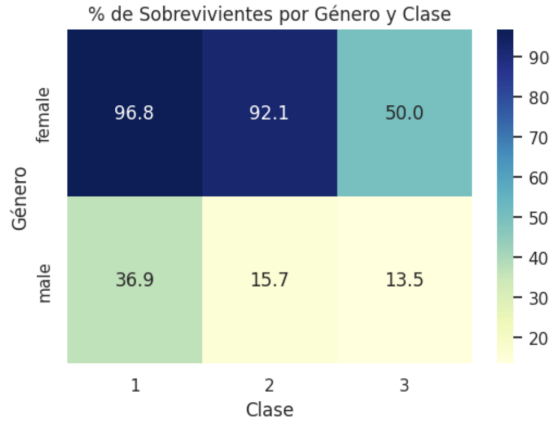


Figura 4. Gráfica que compara supervivencia de género por clase.

Posteriormente, en la matriz de correlación mostrada en la Figure 5, se identificó la ausencia de correlaciones lineales fuertes entre las variables. Este hallazgo justificó la elección de un modelo de regresión logística, ya que permite evaluar de forma independiente la contribución de cada variable en la predicción de supervivencia, sin depender de relaciones estrictamente lineales. Con ello, fue posible comprobar cómo factores como **acompañamiento**, **clase socioeconómica**, **género** y **edad** interactúan para explicar la hipótesis de que viajar acompañado incrementa las probabilidades de sobrevivir.

En conjunto, este análisis no solo valida parcialmente la hipótesis, sino que también revela que la supervivencia fue el resultado de una interacción compleja entre factores sociales, culturales y estructurales, lo que hace indispensable considerarlos de manera conjunta dentro de los modelos predictivos.

Si bien en los resultados de la regresión logística en la Table III se observa que la *accuracy* mejora únicamente en una proporción pequeña al añadir la variable de **acompañamiento**, este hallazgo adquiere relevancia bajo distintos enfoques. Desde una perspectiva estadística, incluso variaciones marginales en la precisión son significativas cuando se trabaja con un conjunto de datos de tamaño limitado y con variables que no presentan correlaciones lineales fuertes. Dichas mejoras refuerzan la estabilidad del modelo y confirman que el acompañamiento no es un efecto espurio, sino un factor con impacto real en la predicción.

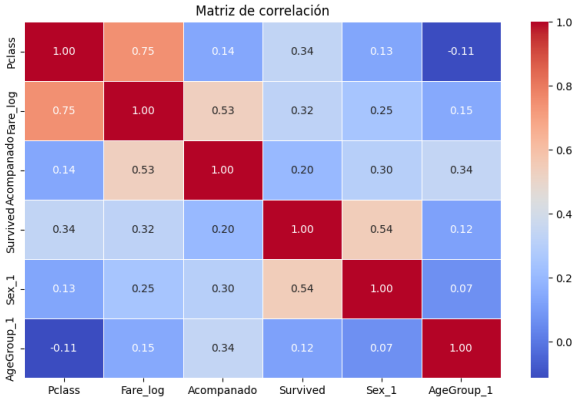


Figura 5. Matriz de correlación.

Desde una perspectiva social e histórica, la presencia de acompañantes pudo haber desempeñado un papel determinante en las decisiones de evacuación. Viajar en familia o en grupo favorecía la cooperación, la priorización de mujeres y niños, y aumentaba la probabilidad de acceder colectivamente a los recursos de rescate. Aunque este tipo de comportamiento no siempre se refleja en un salto considerable de *accuracy*, sí aporta evidencia cualitativa que complementa la interpretación numérica del modelo.

En síntesis, aunque la variable de acompañamiento no produzca una mejora drástica en términos de métricas globales, sí constituye un predictor esencial que, al combinarse con clase, género y edad, refuerza de manera convincente la hipótesis central: no viajar solo incrementaba las probabilidades de supervivencia. Incorporar esta dimensión permite capturar con mayor realismo la complejidad social y humana del desastre del Titanic.

Cuadro III
COMPARACIÓN DE DESEMPEÑO ENTRE MODELO DE REGRESIÓN LOGÍSTICA CON Y SIN ACOMPAÑAMIENTO

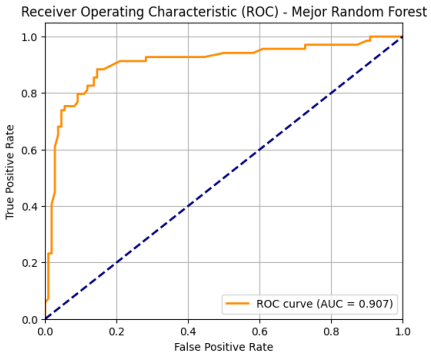
Métrica	**Con acompañamiento**	**Sin acompañamiento**
Accuracy	0.837	0.832
Precisión (0)	0.86	0.86
Precisión (1)	0.80	0.79
Recall (0)	0.88	0.87
Recall (1)	0.77	0.77
F1-score (0)	0.87	0.86
F1-score (1)	0.79	0.78

Con el objetivo de evidenciar con mayor claridad el impacto de la variable de **acompañamiento**, se empleó el modelo de *Random Forest*, ya que este permite capturar interacciones complejas entre variables, incluir múltiples factores simultáneamente y no depender de relaciones lineales, las cuales se mostraron ausentes en la matriz de correlación (Figure 5). Esta elección metodológica proporciona un análisis más profundo y robusto, capaz de reflejar la influencia real de los acompañantes en la probabilidad de supervivencia.

Cuadro IV
RESULTADOS MODELO RANFOM FOREST

Métrica	Valor
Accuracy	0.865
Precisión (0)	0.92
Precisión (1)	0.79
Recall (0)	0.85
Recall (1)	0.88
F1-score (0)	0.89
F1-score (1)	0.94

Como se observa en la Table IV, la *accuracy* mejora respecto a la regresión logística y la curva ROC mostrada en la Figure 6 comienza a presentar la forma característica de un buen clasificador. Esto evidencia que el modelo, al ser entrenado con la misma lógica de preprocesamiento, logra capturar patrones no lineales que aumentan la precisión, especialmente en la combinación de variables como clase, edad y acompañamiento.



g

Figura 6. Gráfica ROC modelo Random Forest.

No obstante, surge la interrogante de si, más allá de la mejora en métricas globales, la hipótesis planteada es correcta: ¿realmente la variable de acompañamiento tiene un impacto sustancial en la supervivencia? Para abordar esta cuestión, se comparó la *accuracy* de los modelos generados con cada variable, tanto incluyendo la variable de acompañamiento como excluyéndola, obteniéndose los resultados presentados en la Table V.

Cuadro V
RESULTADOS DE ACURRACY EN MODELO RANFOM FOREST POR VARIABLE.

Variable	Con acompañamiento	Sin acompañamiento
Pclass	0.698	0.648
Farelog	0.642	0.670
Sex	0.826	0.826
AgeGroup	0.636	0.625

Es relevante destacar que algunas variables muestran un efecto innegable. Por ejemplo, en el caso de la edad, de acuerdo con el lema *"mujeres y niños primero"*, se evidencia que ser niño incrementa la probabilidad de supervivencia, especialmente cuando se combina con la condición de encontrarse acompañado o solo. De manera similar, la variable **clase** revela que estar acompañado puede convertirse en un factor determinante para sobrevivir, mientras que la variable **género** muestra un impacto menos pronunciado.

VI. CONCLUSIONES

El análisis realizado confirma que la variable **acompañado** tiene un impacto positivo en la predicción de la supervivencia de los pasajeros del Titanic, aunque su efecto no es uniforme en todas las categorías. Los modelos evidencian que este factor resulta particularmente relevante en combinación con variables como **clase socioeconómica** y **edad**, mientras que en el caso de **género** su influencia es menos significativa.

En términos de desempeño, el modelo de *Random Forest* superó a la *regresión logística*, mostrando un mayor poder predictivo y una curva ROC con características propias de un buen clasificador. Sin embargo, el incremento de precisión al incluir la variable *acompañado* fue moderado, lo cual sugiere que, aunque relevante, no constituye un predictor determinante por sí solo.

Estos hallazgos validan parcialmente la hipótesis inicial: viajar acompañado aumenta las probabilidades de sobrevivir, pero su efecto depende del contexto de otras variables, especialmente la edad y la clase.

VII. TRABAJO FUTURO

Como líneas de trabajo a desarrollar, se plantean las siguientes:

1. **Aplicación en contextos actuales:** Replicar el modelo en escenarios contemporáneos de evacuación para evaluar si el lema "mujeres y niños primero" sigue vigente o si los criterios de priorización han cambiado.
2. **Inclusión de nuevas categorías de género:** Explorar la integración de identidades no binarias u otras categorías en el modelado, con el fin de reflexionar sobre cómo podrían afectar a las políticas de evacuación modernas.
3. **Optimización de modelos:** Experimentar con arquitecturas más complejas (p. ej., redes neuronales o boosting) que puedan capturar interacciones no lineales entre variables con mayor precisión.

VIII. AGRADECIMIENTOS

Queremos expresar nuestro más sincero agradecimiento a nuestros profesores y mentores Mauricio González Soto, Julio Antonio Juárez Jiménez, Hugo Terashima Marín y Alfredo Esquivel Jaramillo quienes con su orientación, dedicación y apoyo nos guiaron a lo largo de este proyecto. Reconocemos también la importancia de los recursos utilizados, en particular los archivos .csv proporcionados, que constituyeron la base de nuestro análisis. Asimismo, extendemos nuestro agradecimiento a las librerías de Python empleadas para la construcción y predicción de modelos, así como a los recursos disponibles en internet, los cuales facilitaron el desarrollo de este trabajo. Finalmente, valoramos la retroalimentación recibida a lo largo del proceso, que fue fundamental para mejorar continuamente y alcanzar los objetivos planteados.

IX. Referencias

Gupta, K., Sharma, P., Bouza Herrera, C. N. (2018). Surviving the Titanic tragedy: A sociological study using machine learning models. Suma de Negocios, 9(20), 86-92. https://doi.org/10.14349/sumneg/2018.V9.N20.A2

Ekinci, E., İlhan Omurca, S., Acun, N. (2018, April). A comparative study on machine learning techniques using Titanic dataset [Conference presentation]. 7th International Conference on Advanced Technologies, Antalya, Turkey. https://www.researchgate.net/publication/324909545$_{A}Comparative_{S}tudy_{o}n_{M}achine_{L}earning_{T}echniques_{U}sing_{T}itanic_{D}ataset$

Maldonado, A. (2020, 21 de mayo). La interpretabilidad en Machine Learning. CII.IA. https://www.ciiia.mx/noticiasciiia/la-interpretabilidad-en-machine-learning

Awan, A. A. (2023, 28 de junio). Una introducción a los valores SHAP y a la interpretabilidad del machine learning. DataCamp. https://www.datacamp.com/es/tutorial/introduction-to-shap-values-machine-learning-interpretability

Molnar, C. (n.d.). LIME. In Interpretable Machine Learning: A Guide for Making Black Box Models Explainable. Retrieved September 12, 2025, from https://christophm.github.io/interpretable-ml-book/lime.html

Saturn Cloud. (2023, August 15). Fairness-aware Machine Learning. Saturn Cloud. https://saturncloud.io/glossary/fairnessaware-machine-learning

Banco Interamericano de Desarrollo. (2023, agosto 15). Desarrollo del modelo y validación. En Uso responsable de la IA para las políticas públicas: Manual de ciencia de datos. https://el-bid.github.io/Manual-IA-Responsable/desarrollo-del-modelo-y-validaci

Díaz, M. (2018). La inteligencia artificial en el ámbito educativo: Retos y perspectivas. Revista Latinoamericana de Tecnología Educativa, 17(1), 45–60. https://www.redalyc.org/journal/6739/673971913008/html

Referencias