



Inteligencia artificial avanzada para la ciencia de datos I (Gpo 102)

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

Mauricio González Soto

Alfredo Esquivel Jaramillo

Dr. Hugo Terashima Marín

Julio Juárez

Frumencio Olivas Alvarez

Antonio Carlos Bento



INTRODUCCIÓN

CONTEXTO

- Titanic, 1912 → más de 1500 víctimas.
- Más que un accidente marítimo: un reflejo de desigualdades sociales.

FACTORES

- Mujeres y niños primero
- Clase social
- Acompañamiento
- Survived

HIPÓTESIS

¿Tenían más probabilidades de sobrevivir los pasajeros que viajaban acompañados frente a quienes lo hacían solos?

LITERATURA

Artículos

- Academicos:
 - Gupta
 - Ekinici
- Kaggle:
 - “Titanic: Machine Learning from Disaster”
 - “Survival Prediction Challenge 2020”
- Se enfocan en mejorar presición, no tanto en interpretación.

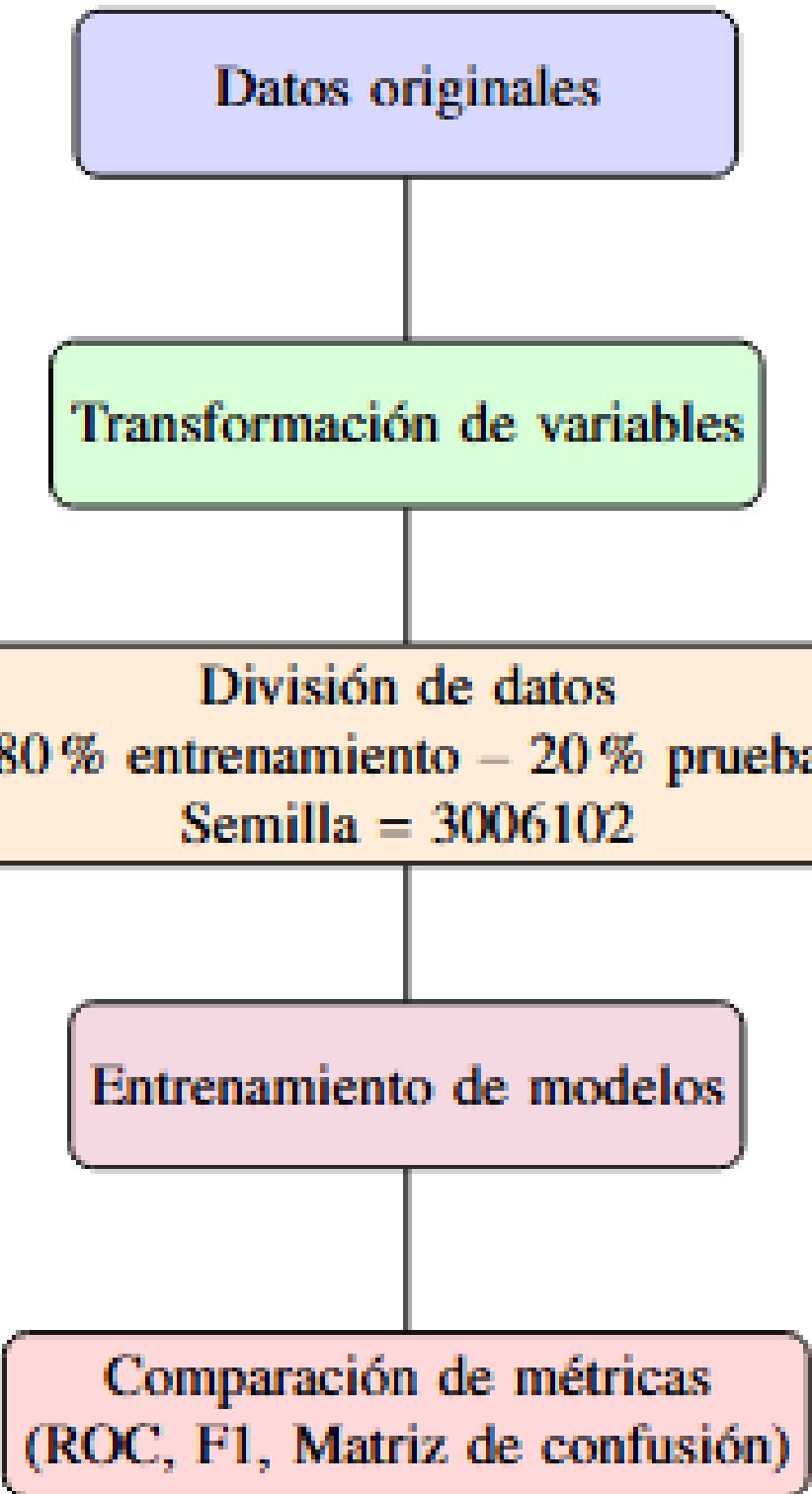
Vocabulario

- Que es la interpretabilidad?
- Paradigmas:
- SHAP(teoría de juegos), LIME(aproximado localmente)
- Que es Fairness?
 - Impossibility theorems.

¿Cuál es el reto?

Dado los impossibility theorems, desarrollar un modelo de predicción de supervivencia dado el lema de “mujeres y niños primero”

METODOLOGÍA





VARIABLES ORIGINALES

- PassengerId
- Survived
- Pclass
- Name
- Sex
- Age
- SibSp
- Parch
- Ticket
- Fare
- Cabin
- Embarked

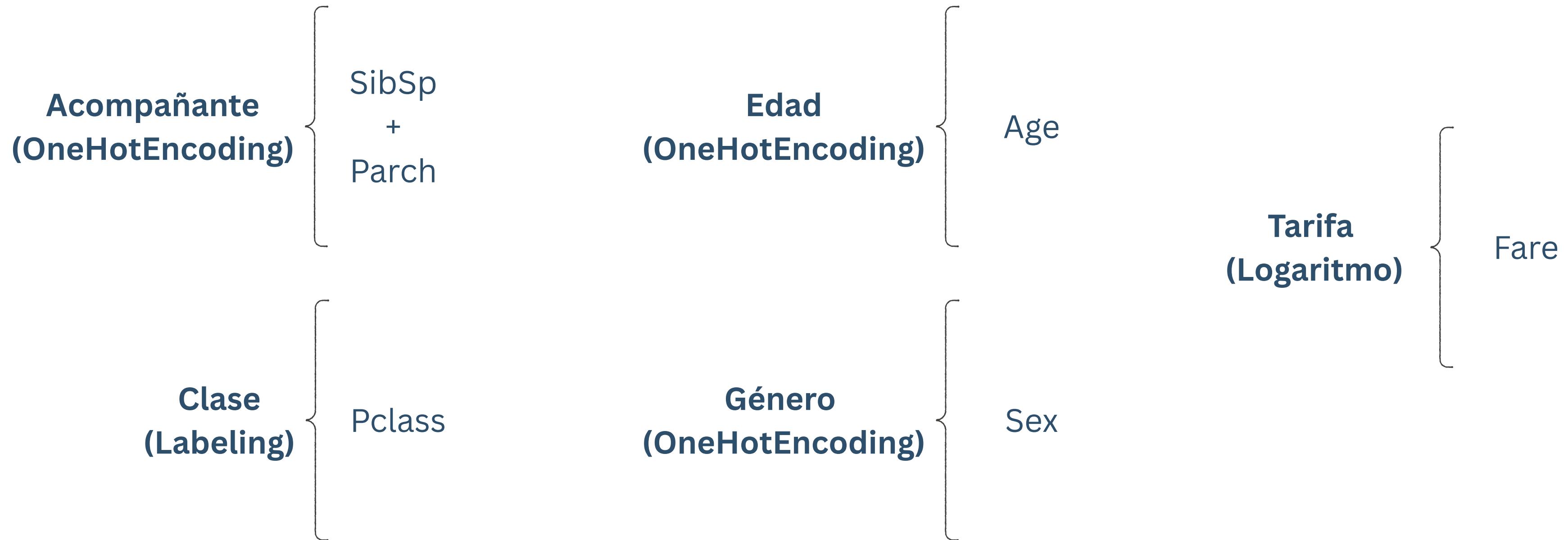


VARIABLES SELECCIONADAS

- PassengerId
- Survived
- Pclass
- Name
- Sex
- Age
- SibSp
- Parch
- Ticket
- Fare
- Cabin
- Embarked

*Mujeres
y niños primero*

TRANSFORMACIÓN DE VARIABLES

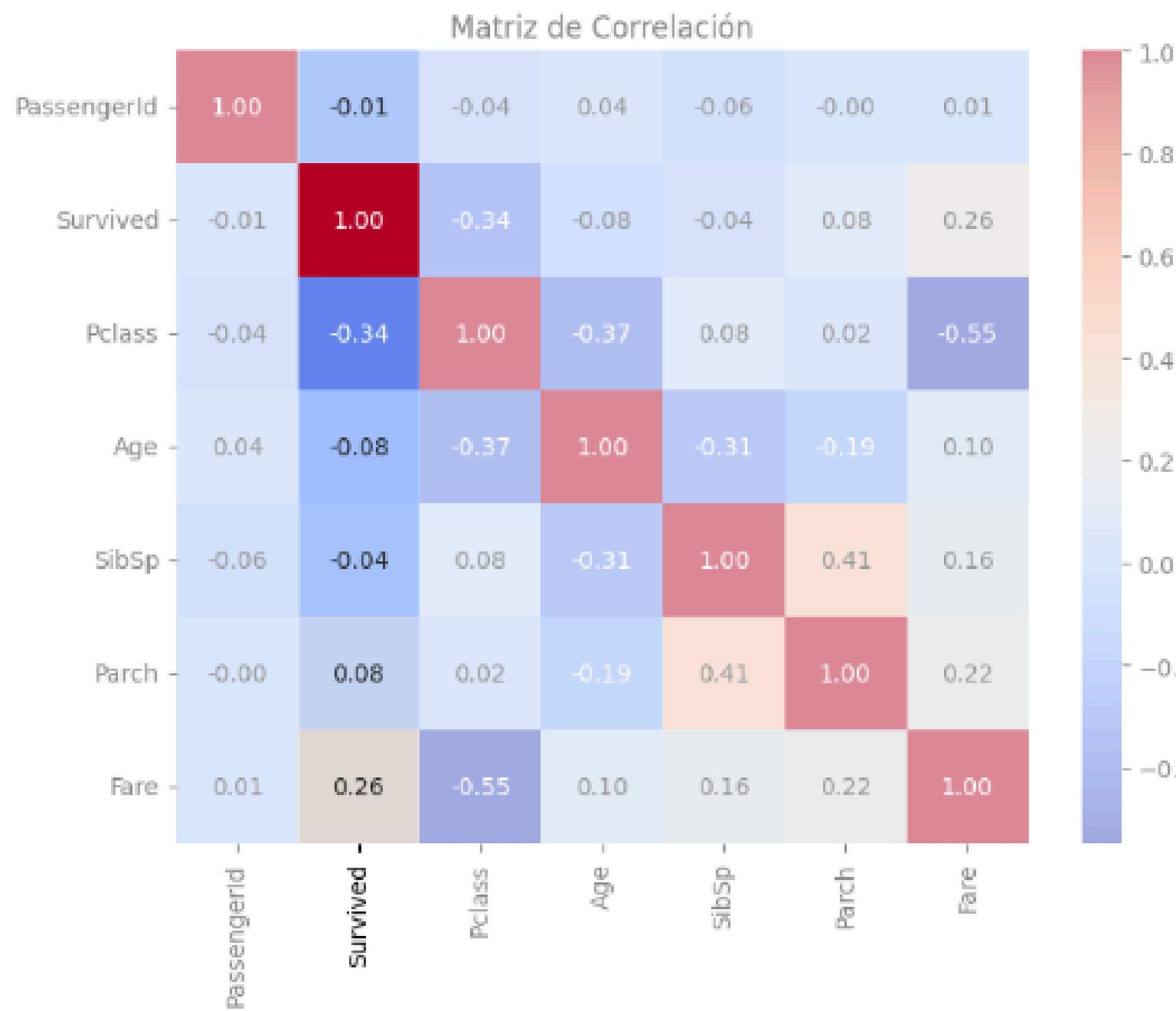


Nota: Se aplicó escaldo a Fare e imputación a age mediante mediana

*Mujeres
y niños primero*

MODELOS

Regresión Logística



Random Forest

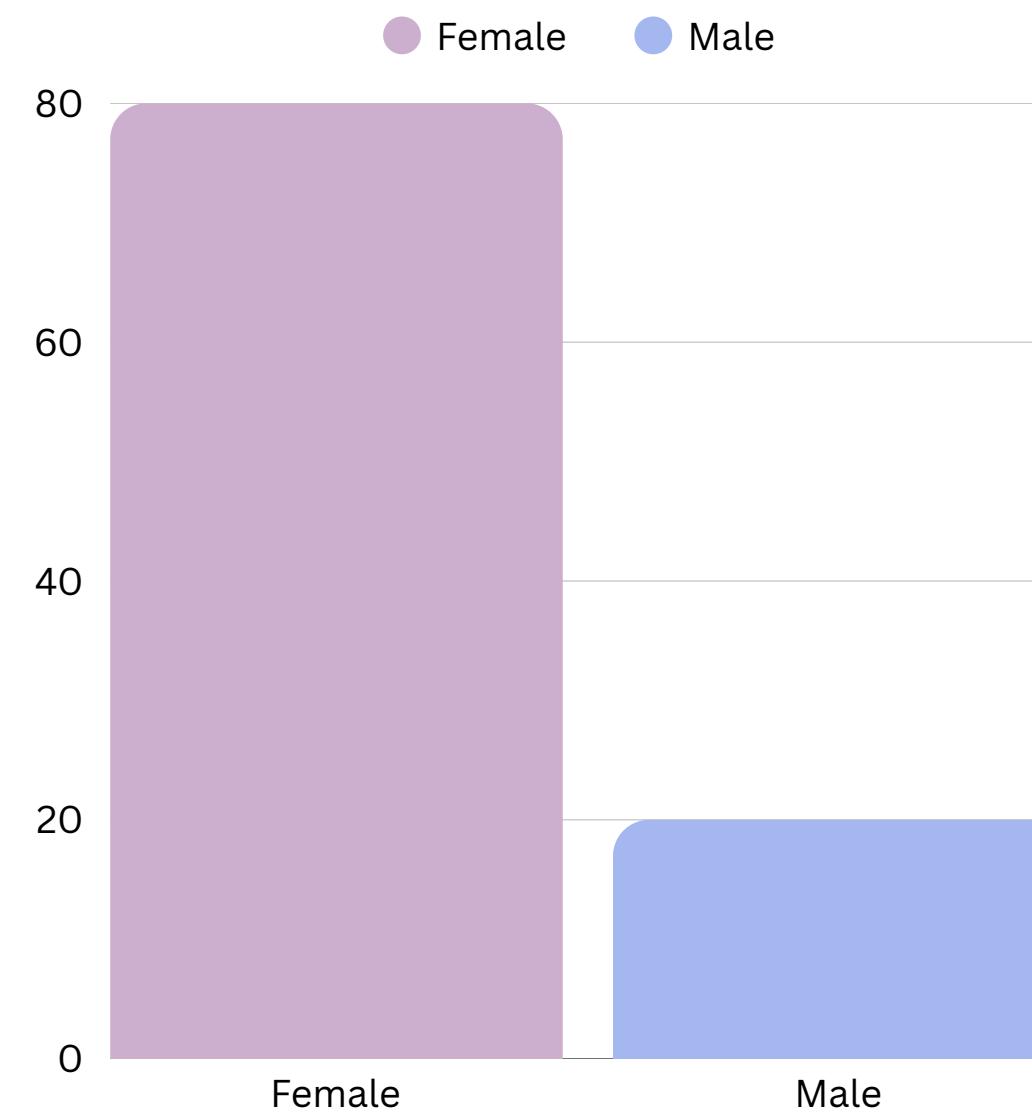


Grafico 1. Porcentaje de supervivencia por género

Figura 2. Matriz de correlación original.

RESULTADOS Y DISCUSIÓN

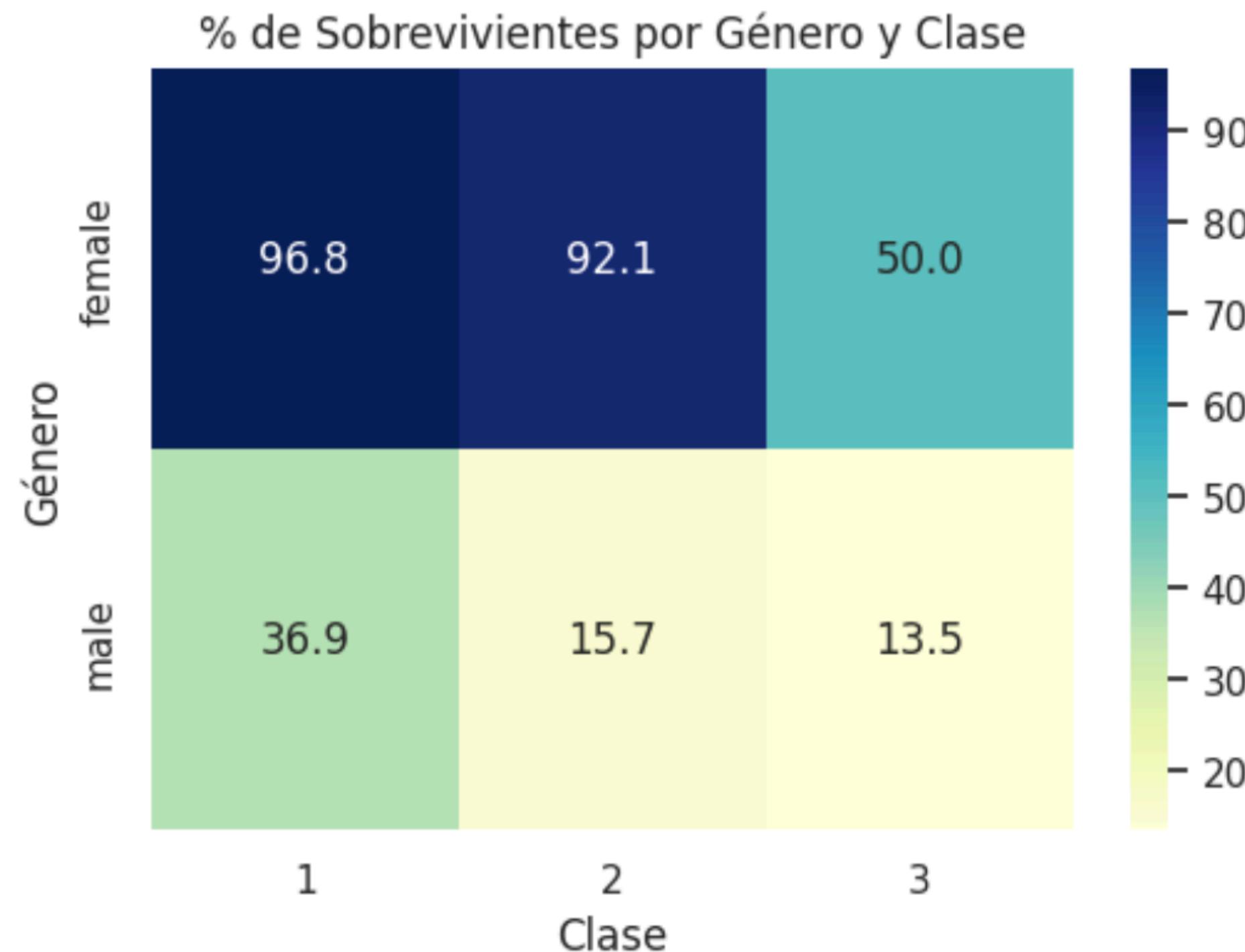


Figura 1. Gráfica que compara supervivencia de género por clase.

RESULTADOS Y DISCUSIÓN

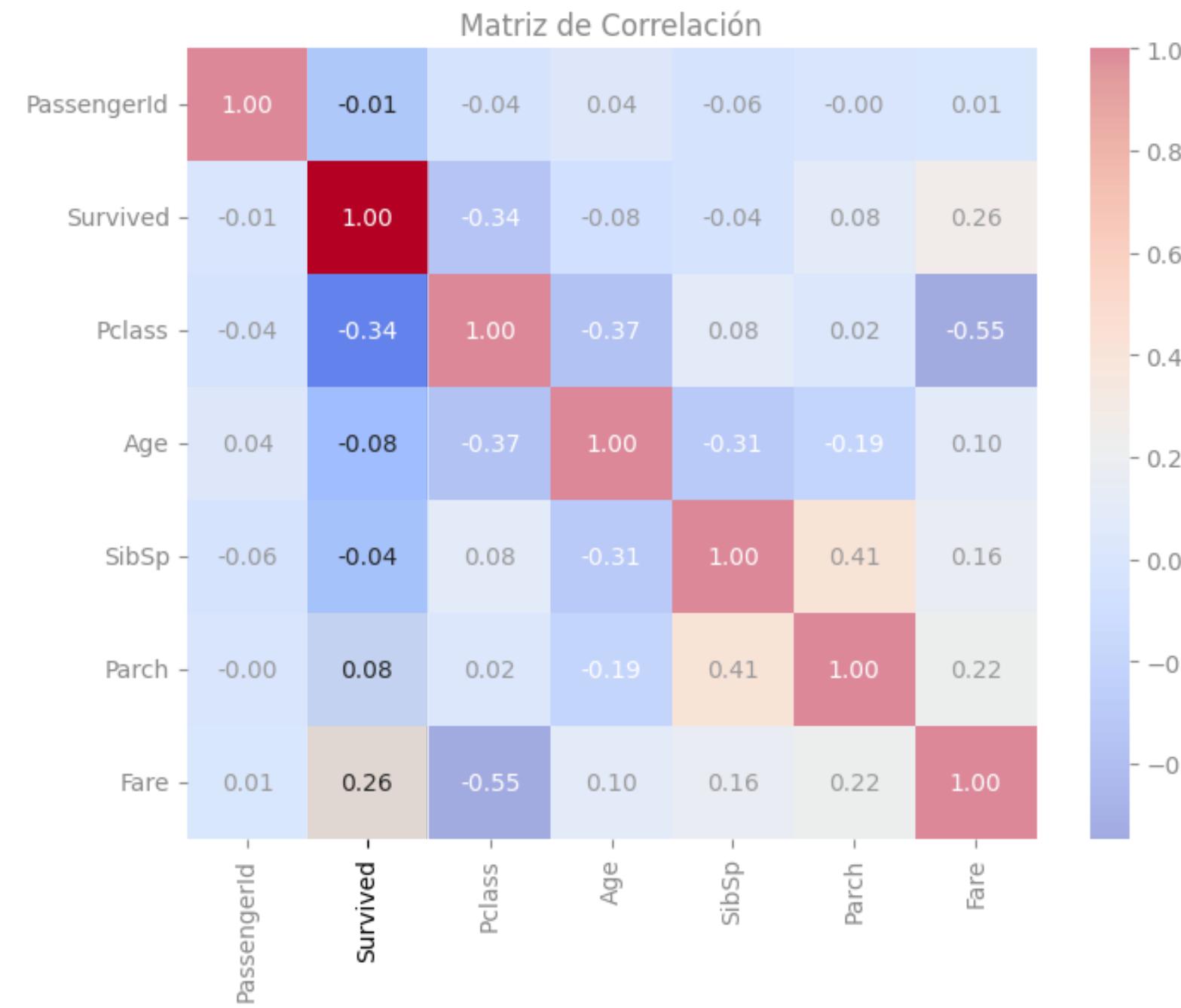


Figura 2. Matriz de correlación original.

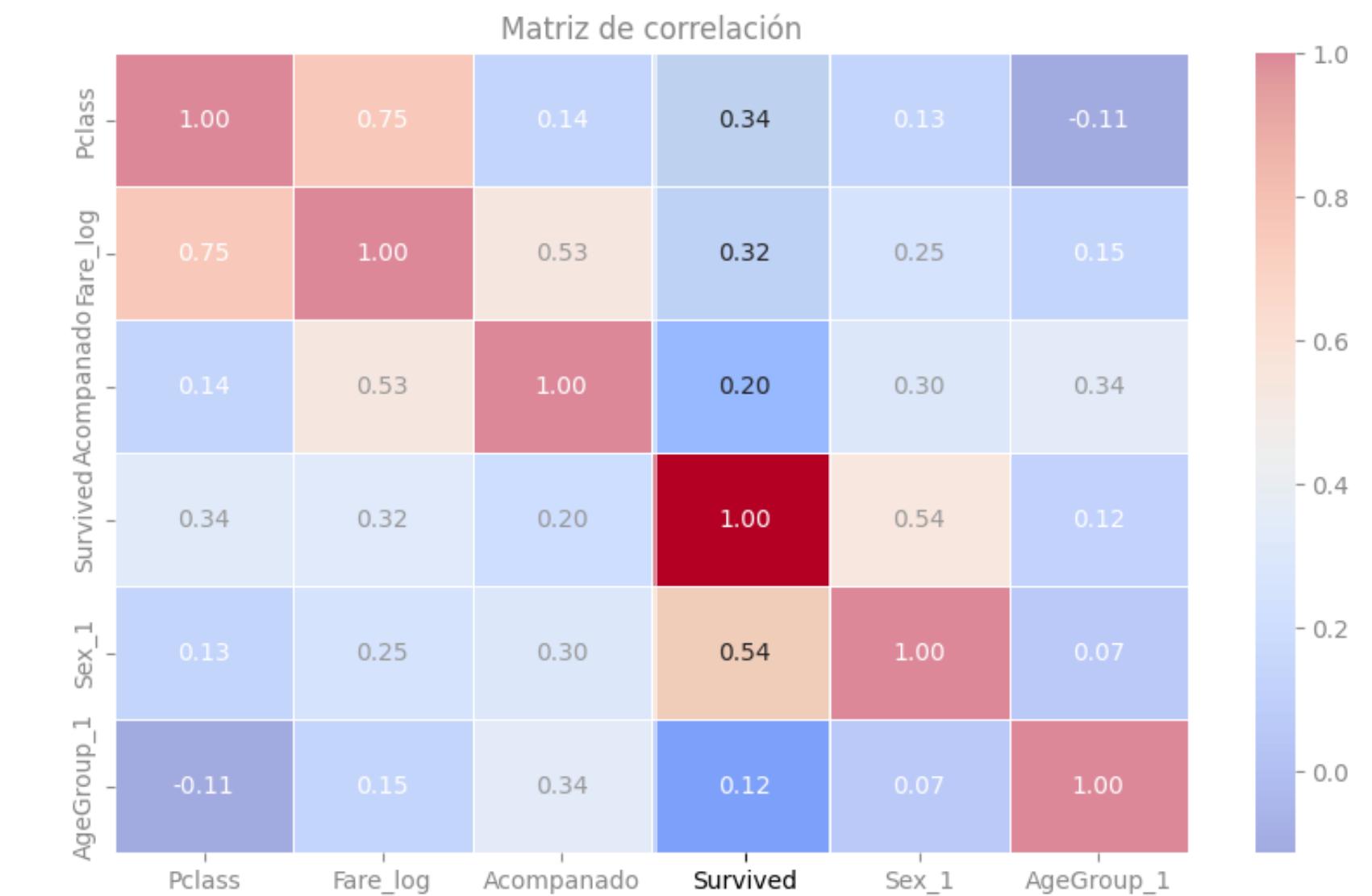


Figura 3. Matriz de correlación con el lema.

RESULTADOS Y DISCUSIÓN

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

Figura 4. Comparación de desempeño entre modelo de regresión logística con y sin acompañamiento.

RESULTADOS Y DISCUSIÓN

== Métricas del Random Forest por defecto ==

Accuracy: 0.8212290502793296

ROC-AUC: 0.891699604743083

Matriz de confusión:

```
[[91 19]
 [13 56]]
```

Reporte de clasificación:

	precision	recall	f1-score	support
0	0.88	0.83	0.85	110
1	0.75	0.81	0.78	69
accuracy			0.82	179
macro avg	0.81	0.82	0.81	179
weighted avg	0.83	0.82	0.82	179

Figura 5. Resultados Random Forest.

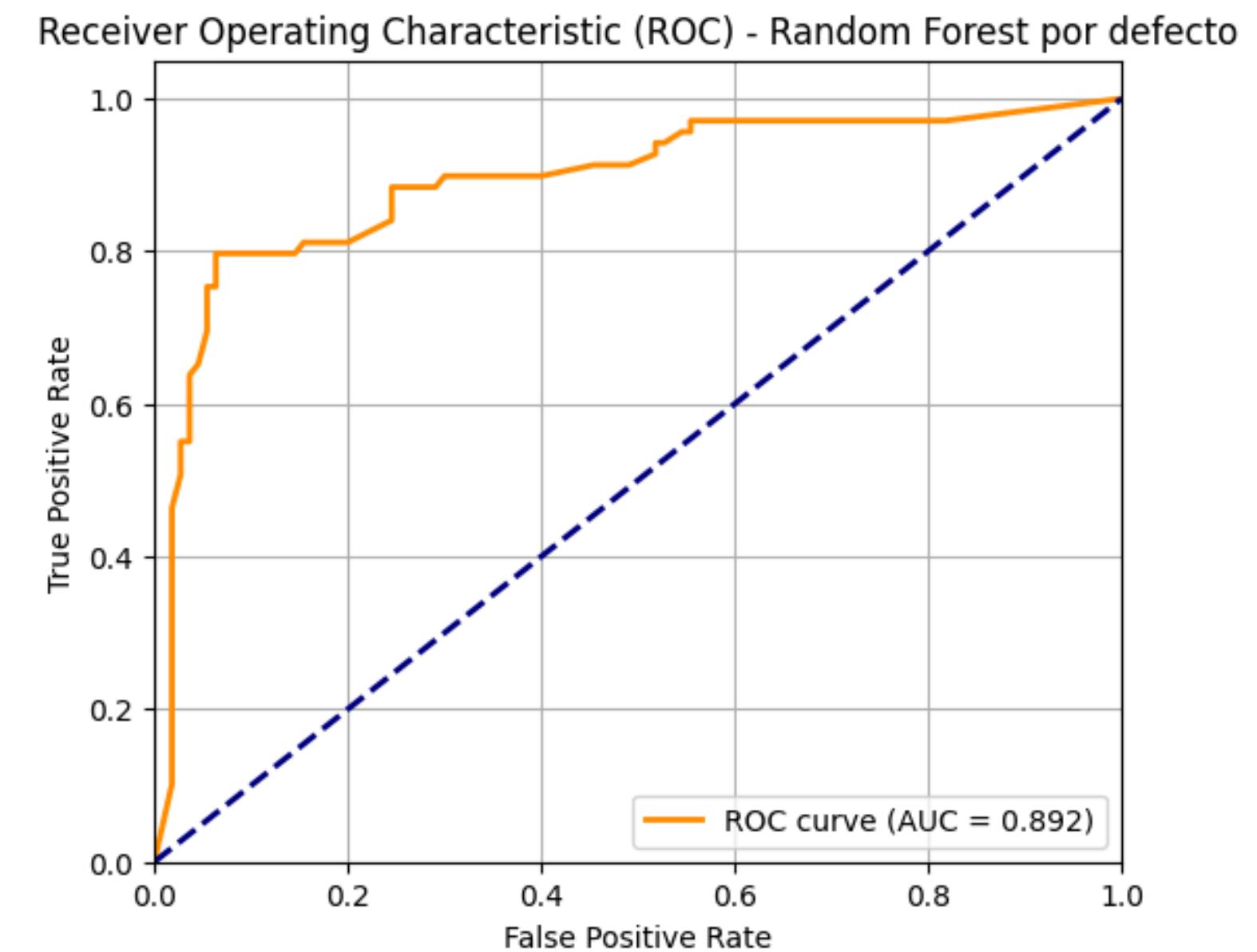


Figura 6. Gráfica ROC modelo Random Forest.

RESULTADOS Y DISCUSIÓN

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

Figura 7. Resultados Random Forest con Grid Search.

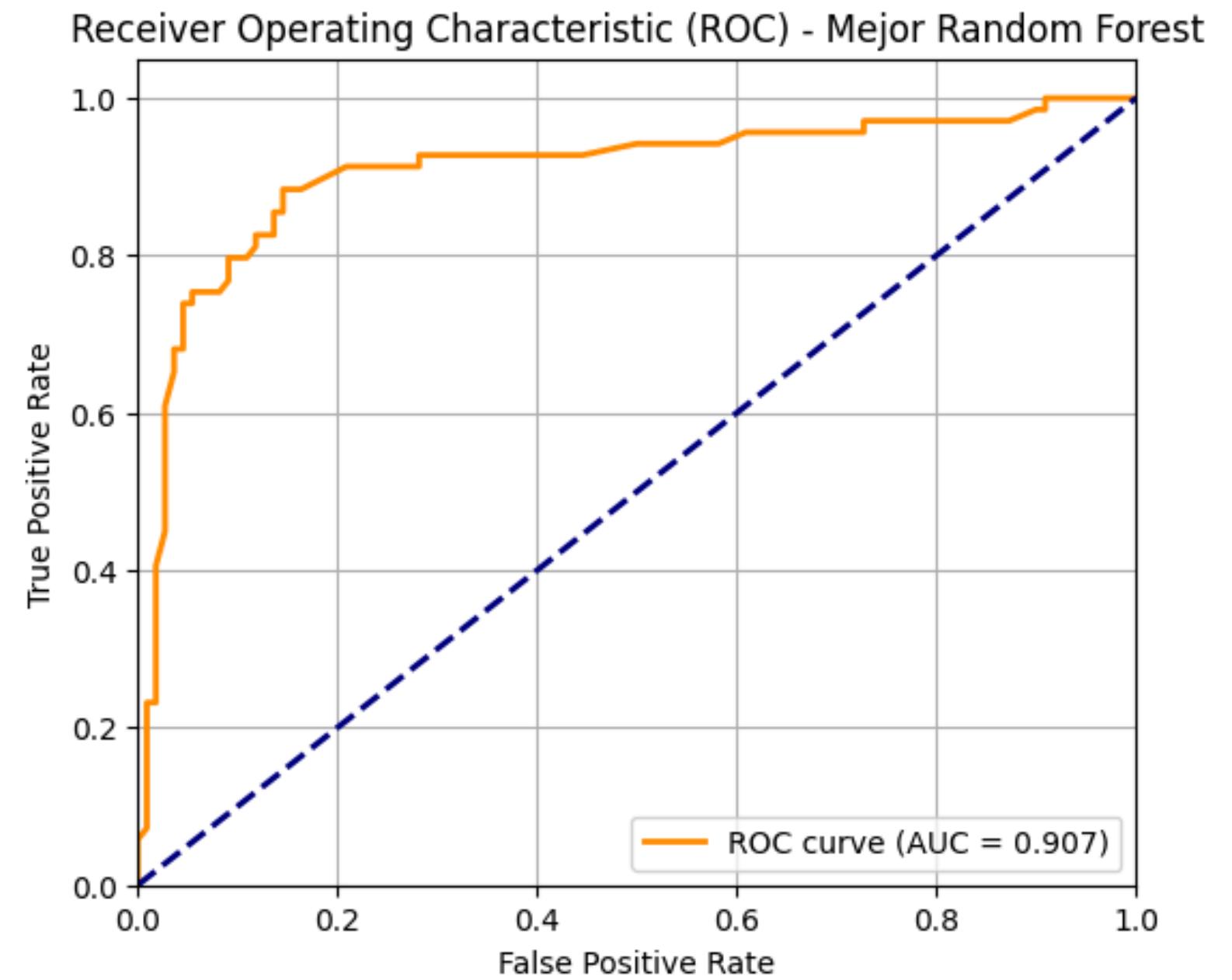
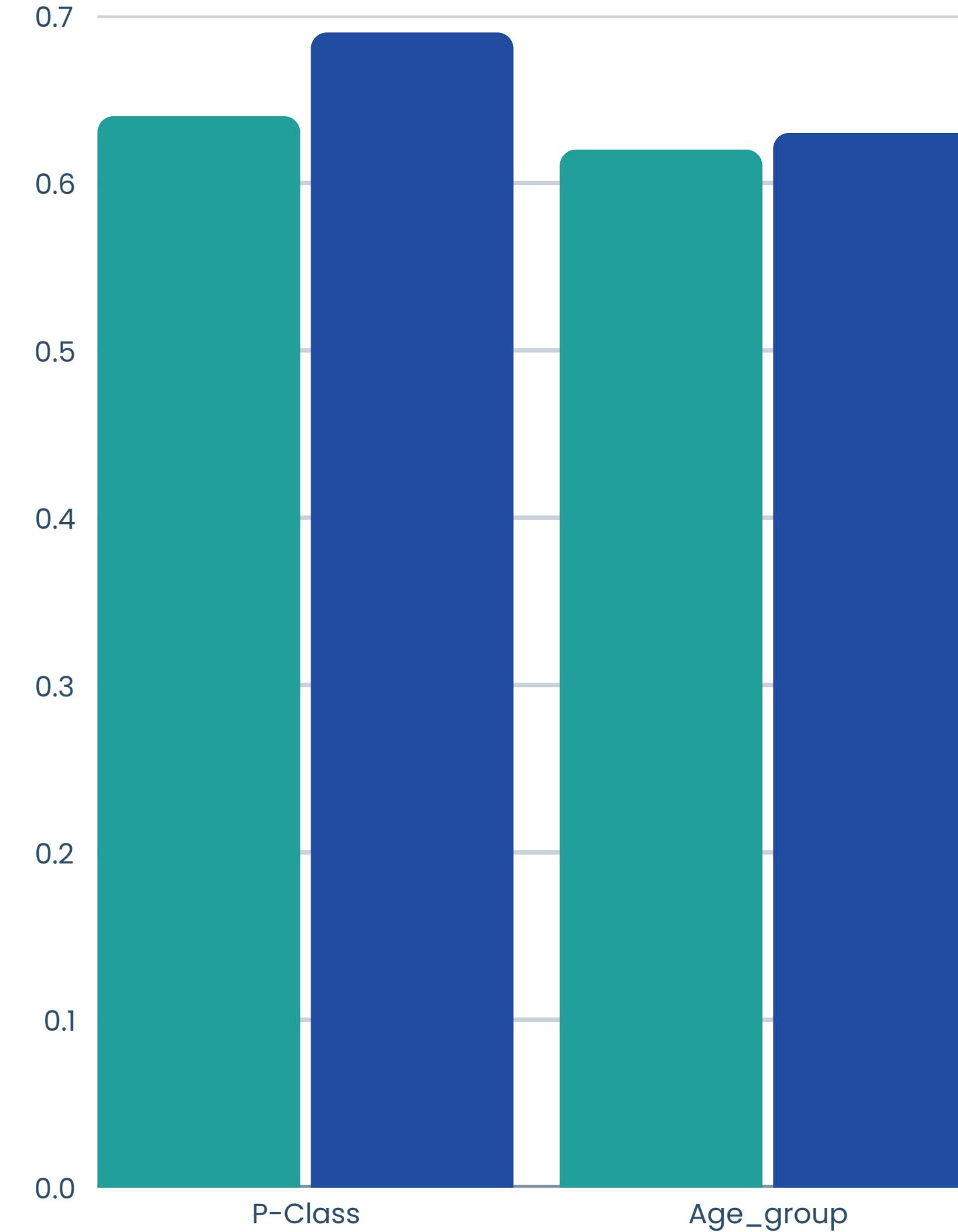


Figura 8. Gráfica ROC modelo Random Forest con Grid Search.

RESULTADOS Y DISCUSIÓN

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

Figura 9. Resultados de accuracy en modelo Ranfom Forest con Grid Search por variable .



ACOMPAÑAMIENTO TIENE UN IMPACTO.

NO UNIFORME

No es uniforme en todas las categorías.

NO
DETERMINANTE

Aunque relevante, no constituye un predictor determinante

MODELO

Random Forest pudo reflejar de mejor manera las relaciones entre variables.

TRABAJO FUTURO

Aplicación en contextos actuales.

Replicar el modelo
en escenarios contemporáneos.

Analizar impacto por tipo de
acompañamiento.

Explorar diferencia de sobrevivir por tipo de
familiar.

Optimización de modelos.

Experimentar con arquitecturas más
complejas.

GRACIAS