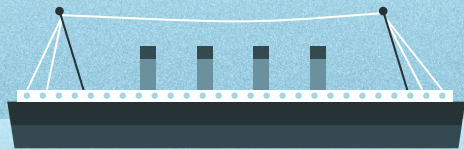


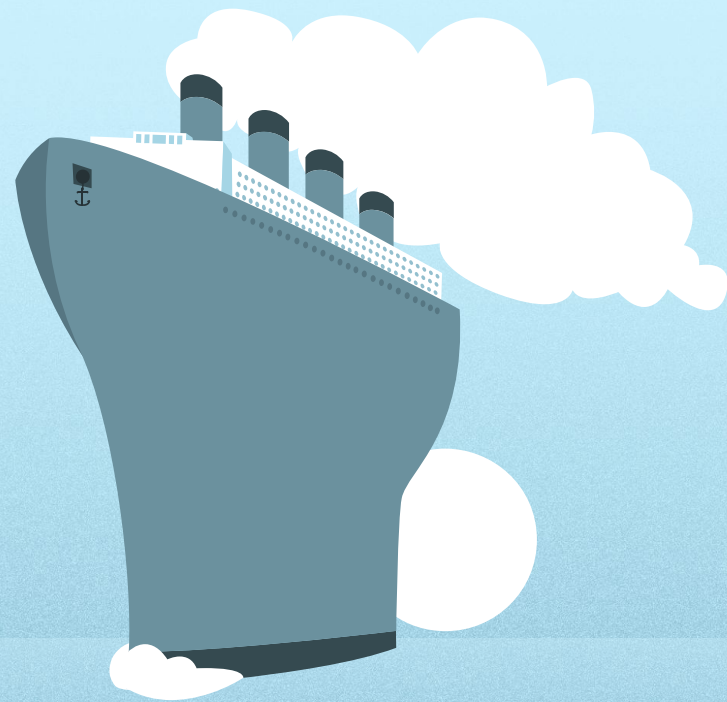
— Presentación de Reto —

# TITANIC

Machine Learning from Disaster



# OVERVIEW



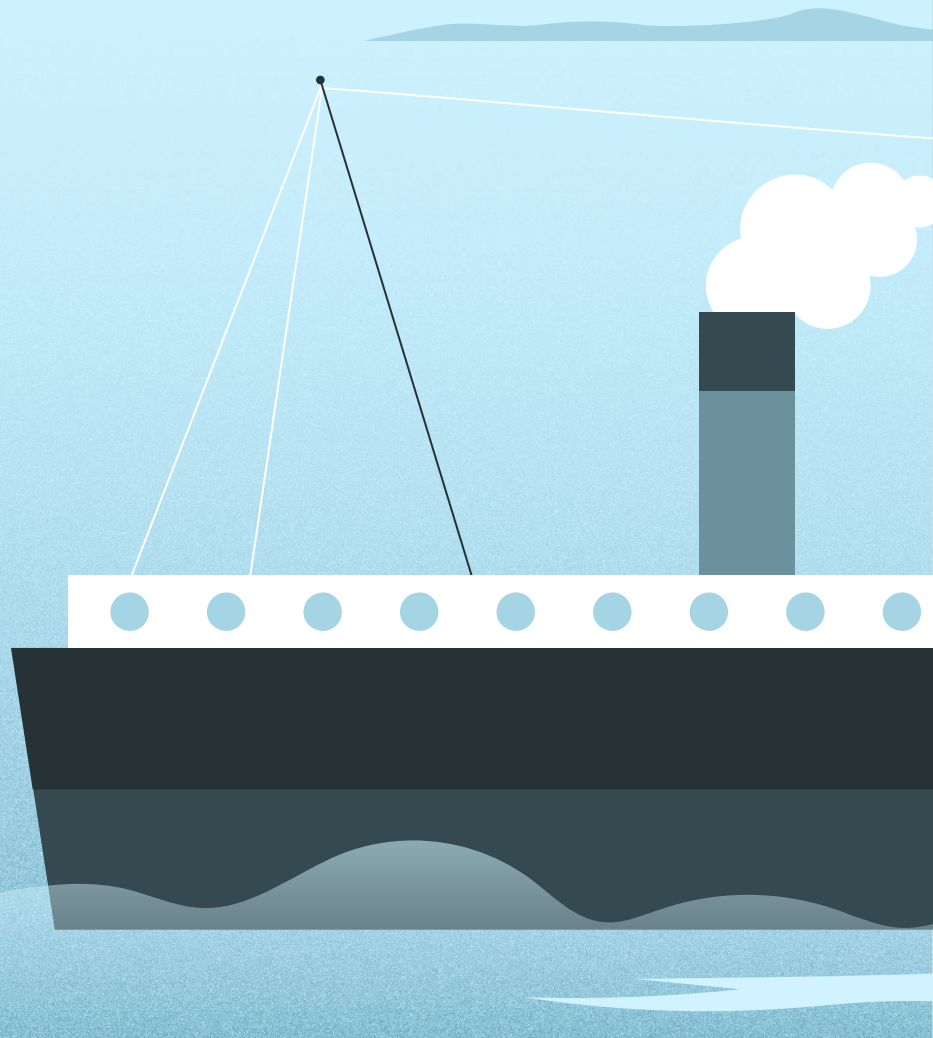


## — La situación —

El 15 de abril de 1912, durante su primer viaje, el RMS Titanic se hundió tras chocar con un iceberg.

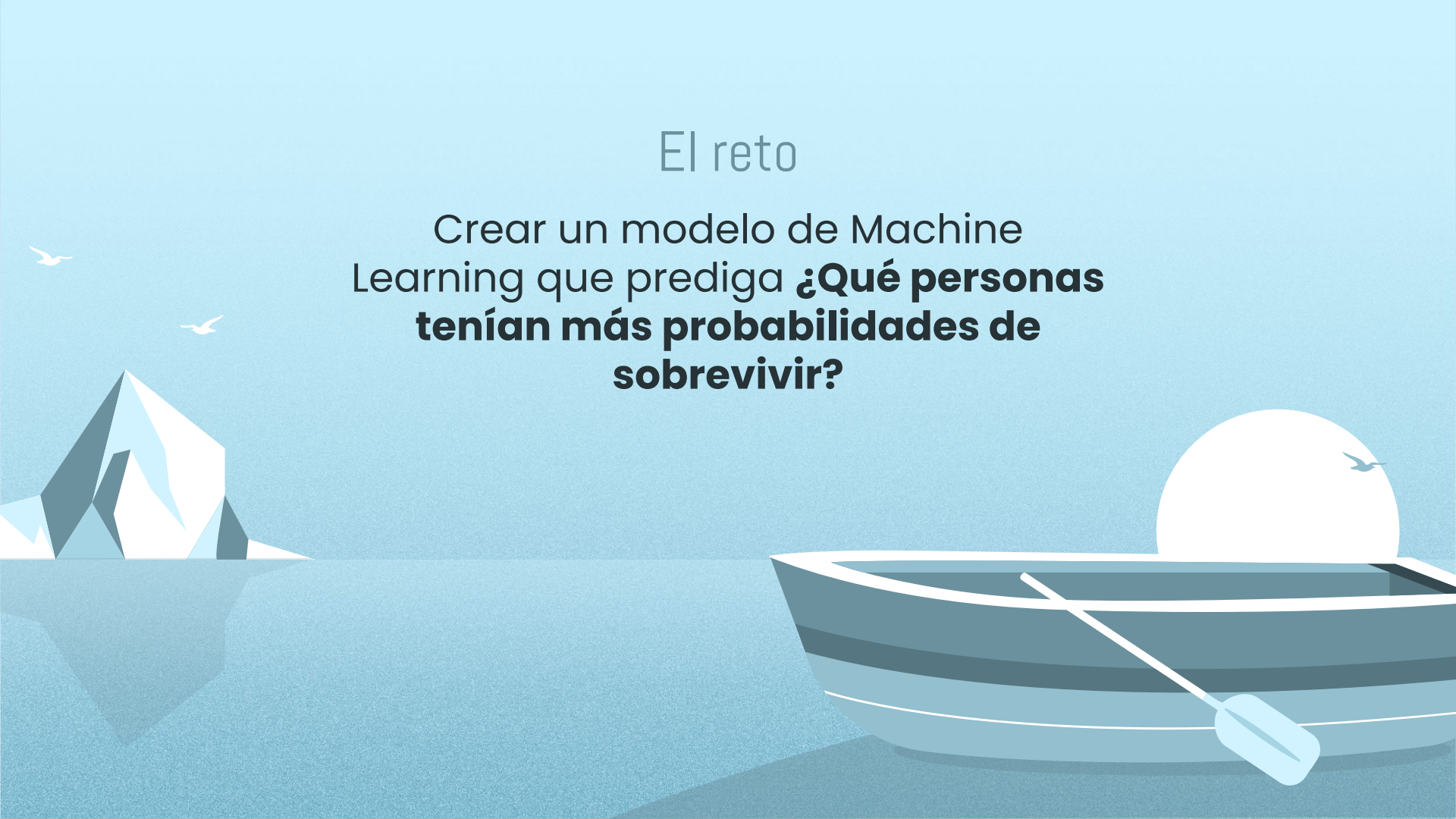
Desgraciadamente, no había suficientes botes salvavidas para todos, por lo que murieron 1,502 de los 2,224 pasajeros.

Aunque uno podría pensar que hubo cierto grado de suerte en la supervivencia, parece que algunos grupos de personas tuvieron más probabilidades de sobrevivir que otros.



## El reto

Crear un modelo de Machine Learning que prediga **¿Qué personas tenían más probabilidades de sobrevivir?**





# — Dataset —

Se tienen datos de pasajeros

## CONJUNTOS DE DATOS Y NÚMERO DE REGISTROS

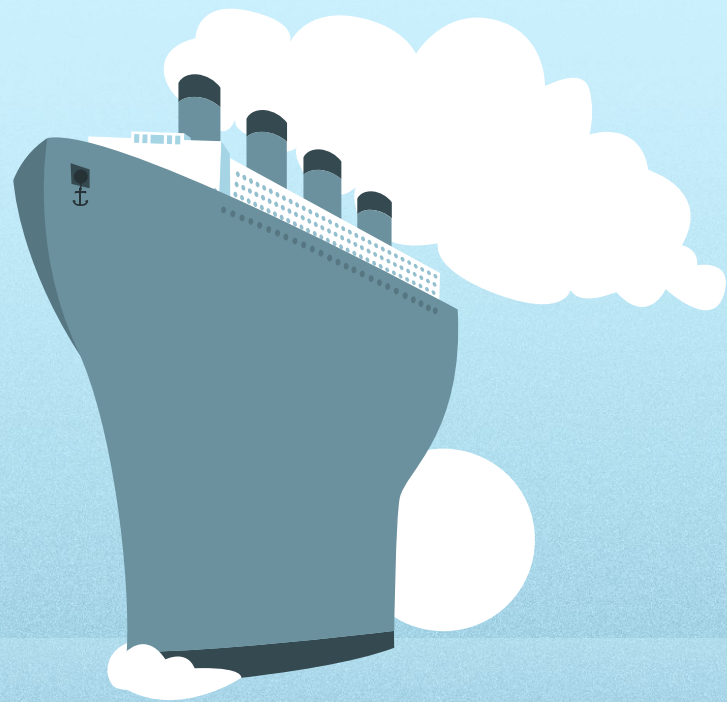
- Entrenamiento (712)
- Validación (179)
- Prueba (418)

## COLUMNAS

(\* Datos Faltantes)

• PassengerId	Categorico	Integer
• Survived	Categorico	Integer
• Pclass	Categorico	Integer
• Name	Categorico	String
• Sex	Categorico	Integer
• Age*	Numérico	Float
• SibSp	Numérico	Integer
• Parch	Numérico	Integer
• Ticket	Categorico	String
• Cabin*	Categorico	String
• Fare	Numérico	Float
• Embarked*	Categorico	String

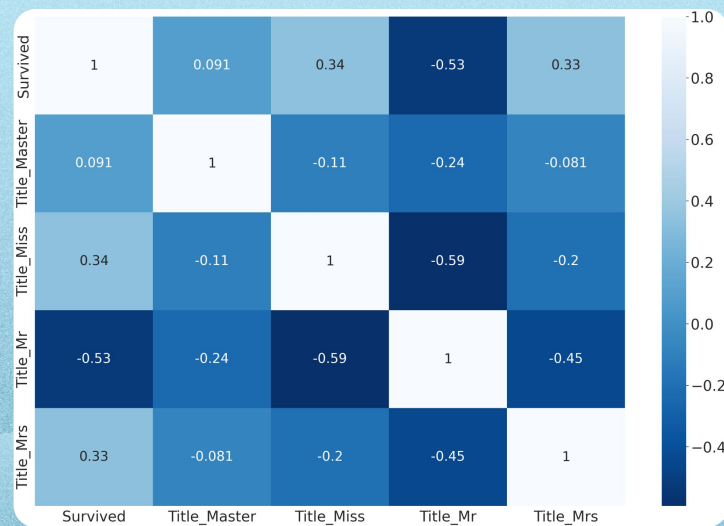
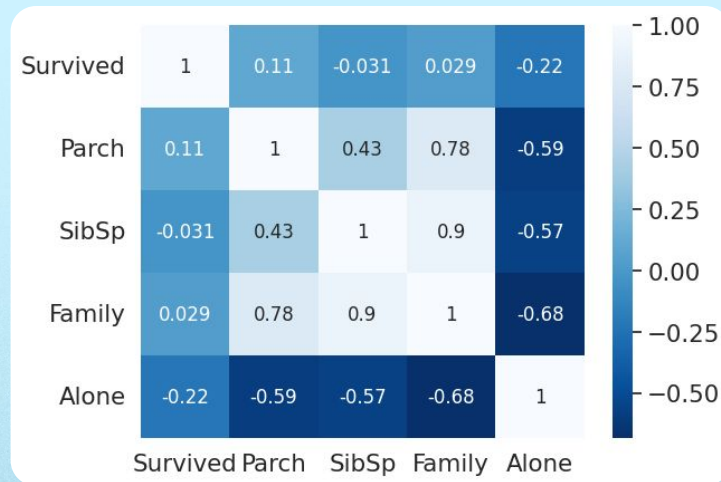
# ANÁLISIS DE LOS DATOS





# — Correlación entre variables —

- Clara relación entre supervivencia y clase
- Quienes sobrevivieron más, pagaron una tarifa mayor
- 81% de los hombres y el 25% de las mujeres fallecieron
- Títulos tienen una relación relevante con la supervivencia (Excepto "Master"), pues correlación con supervivencia es mayor a 0.34



# — Selección de variables —

## Seleccionadas

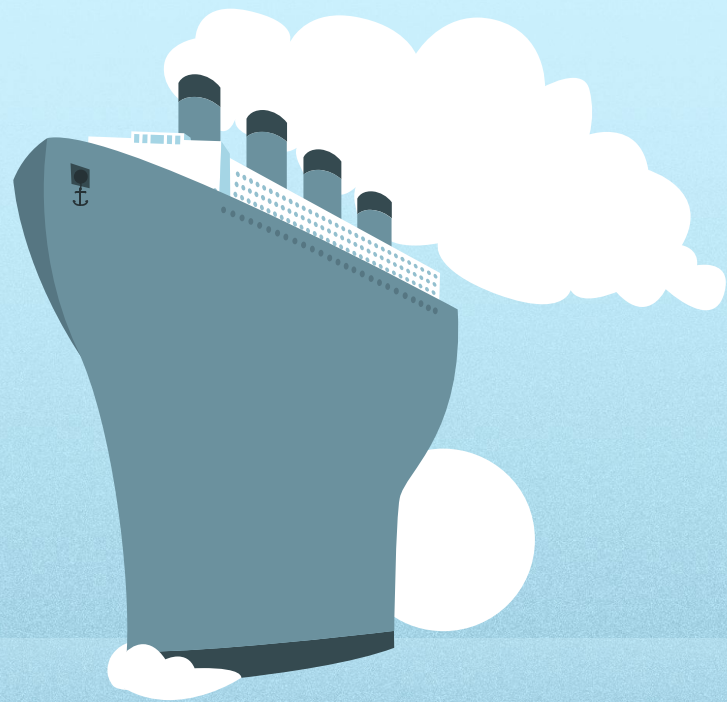
- Sex
- Fare
- Pclass
- Alone
- Title
- Age
- Embarked

## NO seleccionadas

- Cabin
- Ticket
- Parch
- SibSp
- Name



# PREPARACIÓN DE LOS DATOS

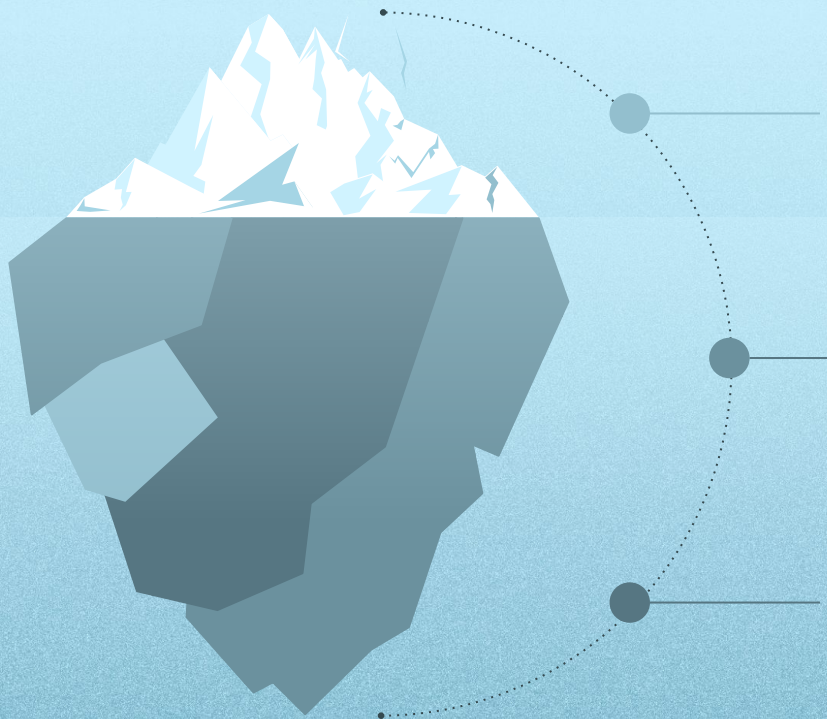


## — Transformación de datos —

- Se obtuvieron variables dummies de título
- Se creó la variable “Alone”, utilizando los datos de Parch y Sibsp, y se transformó a binaria
- Variable sexo se transformó a binaria
- Se separaron edades por intervalos
- Variable Embarked se cambió a numérica







## Manejo de valores faltantes

Se eliminaron regiones donde Embarked = NaN y se rellenaron edades faltantes con base al sexo, título y clase.

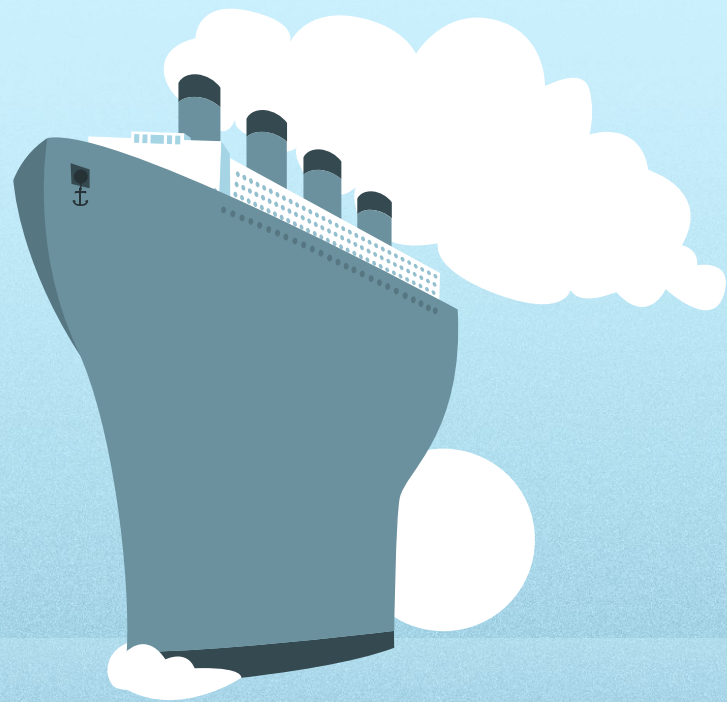
## Escalamiento

Se realizó un escalamiento a todas las variables con StandardScaler().

## Eliminación de columnas

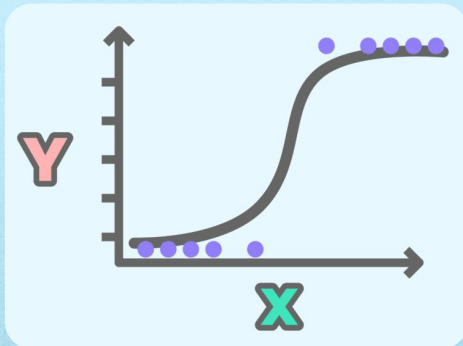
Se eliminaron columnas no utilizadas (Cabina, Ticket, SibSp, Parch y Name)

# ENTRENAMIENTO DEL MODELO





## — Prueba de modelos —

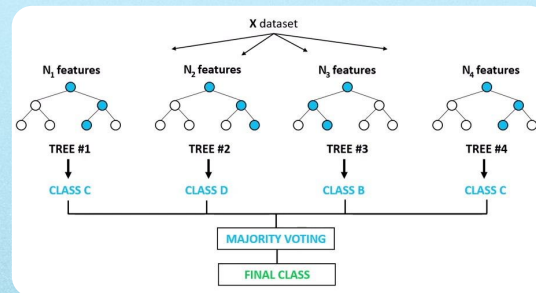


Regresión Logística

Accuracy = 0.83

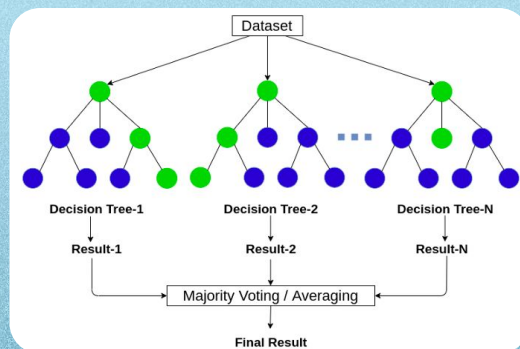
Random Forest  
Regressor

Accuracy = 0.78

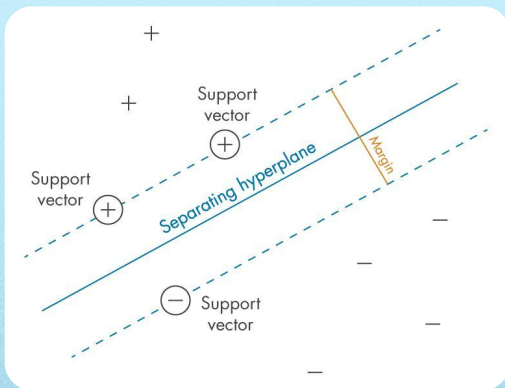


Random Forest  
Classifier

Accuracy = 0.79



## — Prueba de modelos —

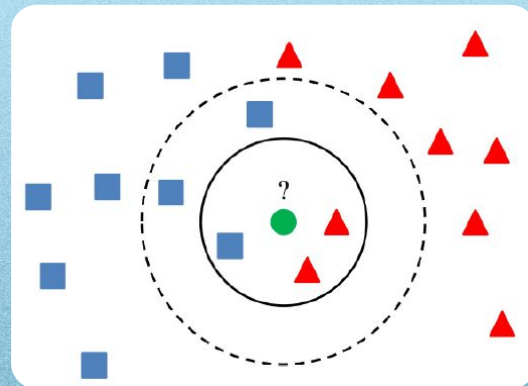


Support Vector  
Machine

Accuracy = 0.82

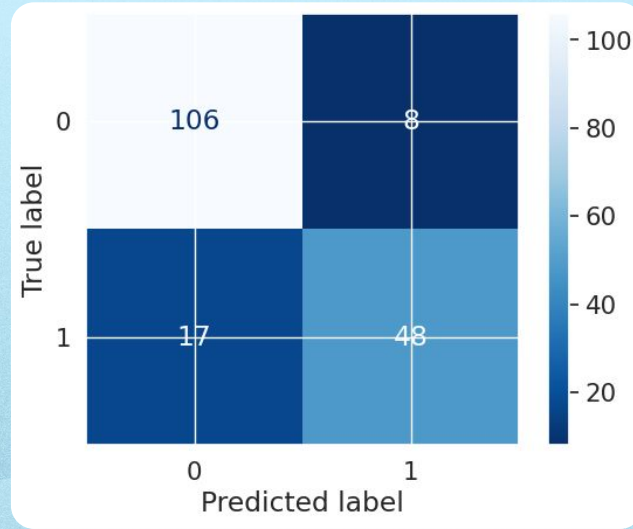
K-nearest  
Neighbors (KNN)

Accuracy = 0.86



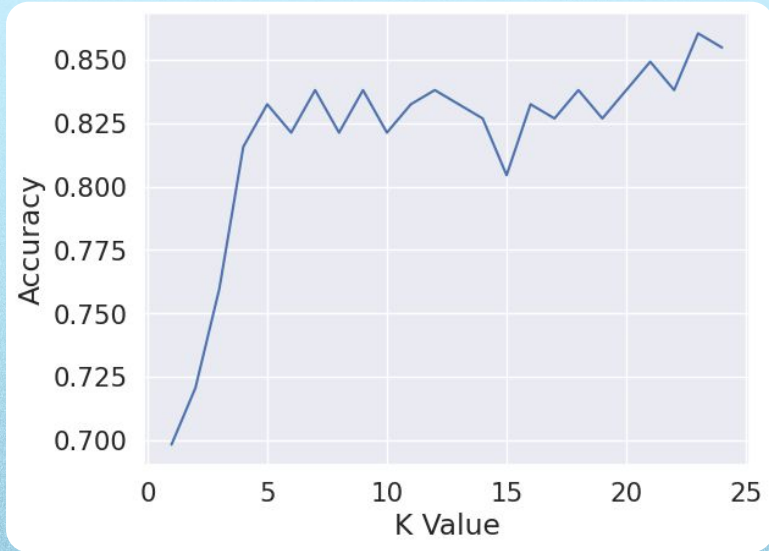


## — Elección del modelo —



Tras comparar modelos, se eligió aquel con mayor accuracy para el conjunto de validación

**K-nearest Neighbours | Accuracy = 0.86**



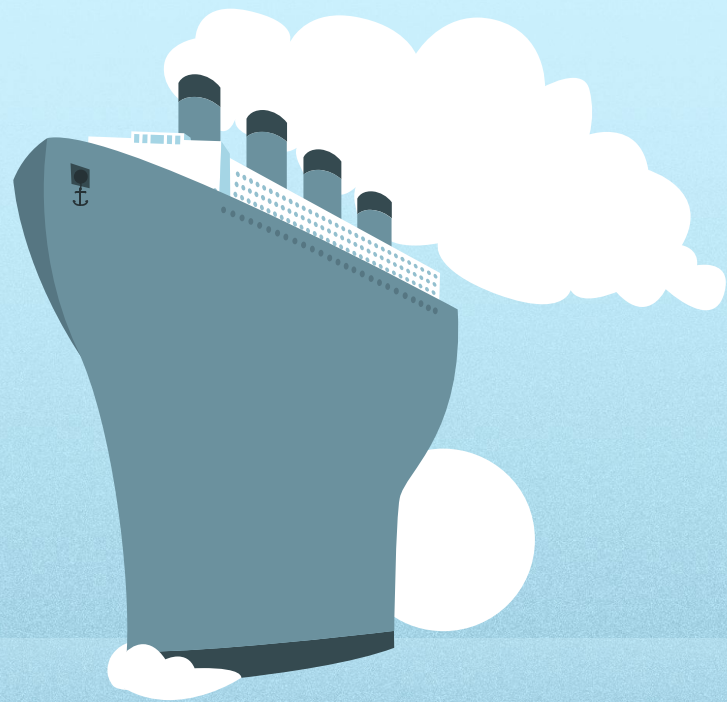
## Ajuste de hiperparámetros

Se probaron diferentes hiperparámetros como `n_neighbors`, `p` y `leaf_size` con `GridSearchCV`

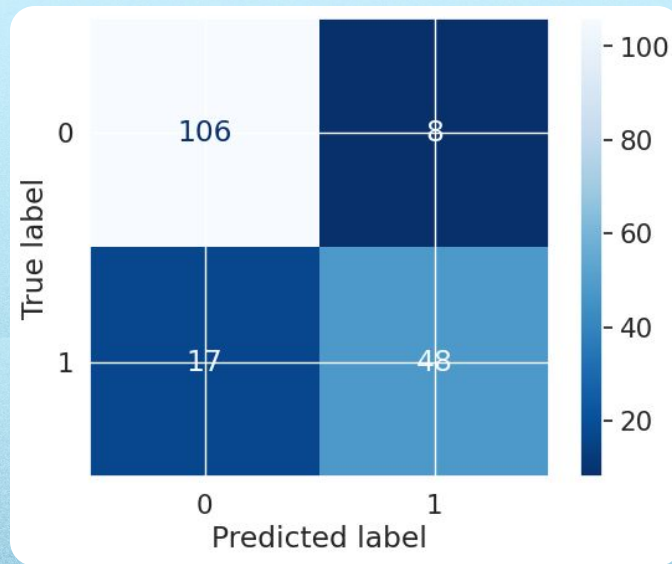
`n_neighbors = 23` tuvo el mejor desempeño



# EVALUACIÓN DEL MODELO



## — Conjunto de validación —

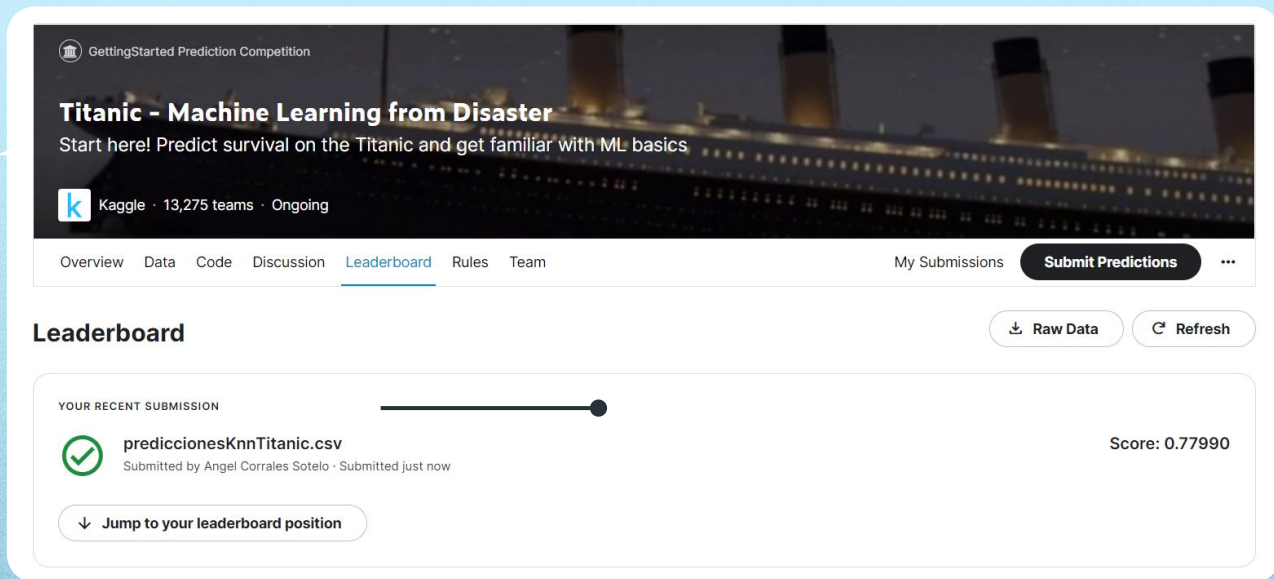


**Optimal K = 23**

**Accuracy = 0.86**



# — Conjunto de prueba —



The screenshot shows the Kaggle competition page for "Titanic - Machine Learning from Disaster". The header includes the competition title, a description "Start here! Predict survival on the Titanic and get familiar with ML basics", and the Kaggle logo with "13,275 teams · Ongoing". The navigation bar has links for Overview, Data, Code, Discussion, Leaderboard (active), Rules, and Team. On the right, there are links for "My Submissions" and a "Submit Predictions" button. Below the navigation bar, the "Leaderboard" section is visible, featuring a "Raw Data" button and a "Refresh" button. The "YOUR RECENT SUBMISSION" section shows a submission named "prediccionesKnnTitanic.csv" with a score of 0.77990, submitted by "Angel Corrales Sotelo" just now. A "Jump to your leaderboard position" button is also present.

GettingStarted Prediction Competition

## Titanic - Machine Learning from Disaster

Start here! Predict survival on the Titanic and get familiar with ML basics

Kaggle · 13,275 teams · Ongoing

Overview Data Code Discussion Leaderboard Rules Team

My Submissions **Submit Predictions** ...

### Leaderboard

Raw Data Refresh

YOUR RECENT SUBMISSION

✓ prediccionesKnnTitanic.csv Score: 0.77990  
Submitted by Angel Corrales Sotelo · Submitted just now

Jump to your leaderboard position

**Score: 0.77990**

# Gracias

## Equipo 5

Ángel Corrales

Aralí Mata

Daniel Cázares

Izrael Rascón

Yolanda Elizondo

