

Trabajo Práctico Final 2025

Sistemas de Soporte para la Toma de Decisiones

Modelo de Predicción de Supervivencia en el Titanic





Introducción

Los modelos estadísticos y de aprendizaje automático permiten predecir comportamientos futuros en base a datos históricos. En este trabajo se aplicarán dichos modelos sobre el conocido Titanic Dataset disponible en Kaggle (<https://www.kaggle.com/competitions/titanic/data>), que contiene información sobre los pasajeros del RMS Titanic y si lograron sobrevivir o no al hundimiento ocurrido en 1912.

Este dataset incluye características demográficas y socioeconómicas tales como edad, sexo, clase de boleto, puerto de embarque, entre otros. La columna objetivo indica si el pasajero sobrevivió (1) o no sobrevivió (0).

Cada registro corresponde a un pasajero único. El objetivo final de este trabajo será desarrollar un modelo que permita predecir la probabilidad de supervivencia de un pasajero en base a sus características y posteriormente construir una aplicación que permita ingresar los datos de un nuevo pasajero y obtener una predicción automática.

Descripción del dataset

Columna	Descripción
PassengerId	Identificador único del pasajero
Survived	Indicador de supervivencia (0 = No sobrevivió, 1 = Sobrevivió)
Pclass	Clase del pasaje (1^a, 2^a, 3^a)
Name	Nombre del pasajero
Sex	Sexo
Age	Edad en años
SibSp	Cantidad de hermanos/esposos a bordo
Parch	Cantidad de padres/hijos a bordo
Ticket	Código del boleto
Fare	Monto pagado por el pasaje
Cabin	Código de cabina
Embarked	Puerto de embarque (C = Cherbourg, Q = Queenstown, S = Southampton)



Objetivo del trabajo

Desarrollar un modelo predictivo que, a partir de los datos del pasajero, indique si hubiera sobrevivido al hundimiento del Titanic. El modelo final deberá poder ser utilizado por un usuario final mediante una aplicación diseñada para tal fin.

Actividades

- 1. Análisis Exploratorio del Dataset (Tener en cuenta los principios vistos en el tema visualización de datos)**
 - Analizar la calidad de los datos, detectar valores faltantes y posibles outliers.
 - Realizar visualizaciones para comprender la distribución de las variables.
 - Realizar gráficos comparativos (por ejemplo, supervivencia según clase, sexo y edad).
 - Analizar si algunas características influyen más que otras en la supervivencia.
 - Determinar si el dataset está balanceado respecto a la variable objetivo.
- 2. Preprocesamiento de los Datos (Utilizar las técnicas de preprocesamiento adecuadas vistas en la materia)**
 - Completar o eliminar valores faltantes según criterio justificado.
 - Codificar variables categóricas.
 - Normalizar o estandarizar valores numéricos cuando corresponda.
 - Analizar si es apropiado aplicar técnicas de reducción de dimensión (PCA por ejemplo).
- 3. Modelado Predictivo (Poner en práctica técnicas de ML y de evaluación de resultados adecuadas, re-entrenar, modificar parámetros, etc.)**
 - Entrenar y comparar al menos los siguientes modelos:
 - i. Árbol de decisión
 - ii. Un método de ensamble (Random Forest o Gradient Boosting)
 - iii. Una red neuronal artificial
 - iv. Otro modelo adicional a elección del alumno.
 - Para cada modelo:
 - i. Seleccionar métricas adecuadas (ej. accuracy, precision, recall, F1-score, ROC-AUC).
 - ii. Realizar validación y graficar matrices de confusión y/o curvas ROC.
 - iii. Comparar el desempeño de los modelos y justificar la selección del mejor.
- 4. Aplicación final**
 - Desarrollar una aplicación web que permita ingresar datos de un pasajero hipotético y obtener la predicción de supervivencia de forma clara y visualmente adecuada para un usuario final.
- 5. Entrega de informe y presentación**
 - Elaborar un informe escrito
 - Preparar una presentación para exponer el proyecto en clase