Titanic Classification.

By Pascalia Maiga.

Business Understanding.

The RMS Titanic, stood as one of the largest ships of its era. In 1912, the Titanic tragically struck an iceberg in the North Atlantic Ocean, leading to catastrophic damage and, ultimately, a devastating loss of lives.

This project aims to analyze the likelihood of survival among passengers and to identify the factors that significantly impacted survival rates, as well as those that held no substantial influence.

Business Problem

The goal is to predict the passenger survival on the Titanic.

KEY FACTORS:

Socio-economic status (ticket class)

Number of siblings/ spouses aboard

Number of parents/Children aboard

Gender

Age

Key Objectives.

1. Develop a Predictive Model

2. Identify Key Survival Factors

3. Evaluate Model Performance and Generalization

4. Provide Interpretability of Model Predictions

5. Create a Reproducible Data Science Workflow

6. Analyze Limitations and Ethical Implications

7. Provide a Data-Driven Narrative on Survival Patterns

Data Understanding.

I downloaded my data from kaggle *https://www.kaggle.com/datasets/brendan45774/test-fileat* . My data had a 418 rows and 12 columns which I used in my project

Data Preparation.

Cleaning involved removing missing values in the dataset and in other cases filling using back filling and forward filling techniques.

I conducted some EDA that yielded some domain knowledge I could use to inform future steps and modelling. Some of the plots I came up with are shown in the next slides:

Modelling.

In this section, I built classification models using

1. Baseline Model / KNNBasic

However, this did not exhibit better performance. Further scaling of the dataset was done and cross validation included to improve the accuracy score

Conclusion.

Key Factors Influencing Survival:

- Gender was the most significant predictor of survival, with females having a notably higher chance of survival compared to males, aligning with the historical "women and children first" protocol.

- Passenger Class (Pclass) showed that first-class passengers were more likely to survive than those in lower classes, likely due to their proximity to lifeboats and greater access to assistance.

- Age also influenced survival, with children (especially those under 10) having a higher survival rate, reinforcing the priority given to younger passengers.

Conclusion.

2. Model Performance and Accuracy:

- The model achieved an accuracy of 100% on the test set, indicating strong predictive performance .

- Cross-validation results demonstrated consistency across different folds, with low variance in accuracy, suggesting that the model generalizes well to unseen data.

- - Metrics such as precision, recall, and F1 score indicated balanced performance across survival classes, with no significant bias toward any class.

Thanks!

- Do you have any questions?