Titanic Machine Learning from Disaster

Cepeda Villanueva Andrés M, Carvajal Garnica Julián, Moreno Barragán Jhonatan D, Ramos Rojas Andrés C. School of Engineering

UNIVERSIDAD DISTRITAL
FRANCISCO JOSÉ DE CALDAS

Introduction

The Titanic Kaggle competition aims to build a predictive system to determine passenger survival using real historical data from 1912.

By applying Systems Analysis principles and machine learning algorithms. This project focuses on designing a robust and modular architecture that improves prediction accuracy and system reliability. It emphasizes understanding data sensitivity, feedback, and chaotic variability as key factors influencing model performance.

Goal

Design a robust predictive architecture that reduces instability caused by missing data, biases and chaotic interactions, improving interpretability and accuracy.

System Architecture

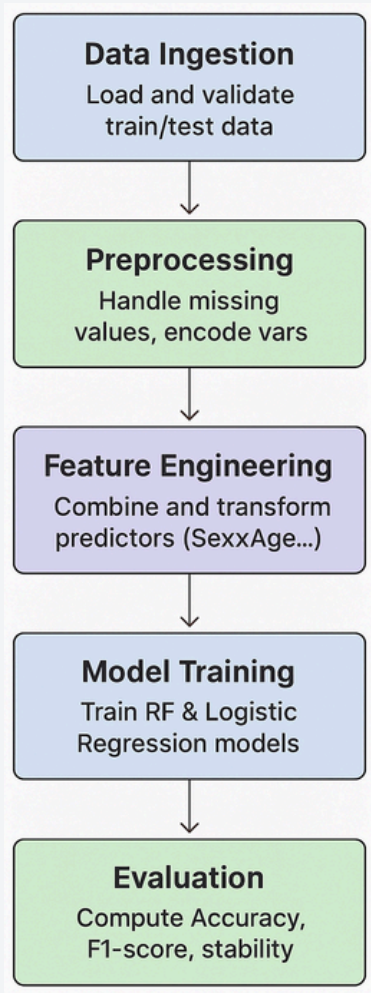

Figure 1. Project Workflow and Timeline Diagram. Source: Workshop 4 (p. 24).

A modular architecture with five stages is proposed: ingestion, preprocessing, feature engineering, training, and evaluation.
Each module improves the model's stability and accuracy, while a feedback loop continuously adjusts the system to ensure reproducible results.
The suggested implementation is based on Python, using Pandas, NumPy, and Scikit-learn under the principles of modularity, sensitivity control, and reproducibility.

Results

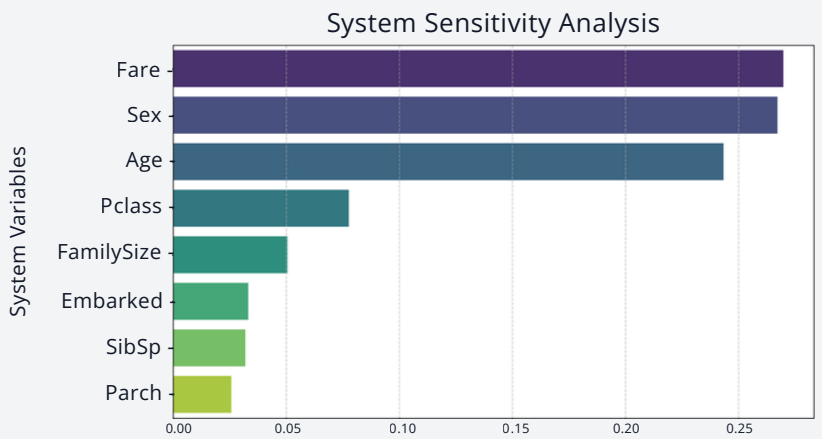

Figure 2. Scenario 1 Results: Feature Importance graph showing that Fare and Sex are the most critical variables driving the system's output. Source: Workshop 4 (p. 32).

Metric	Value
Accuracy	0.83
Precision	0.81
Recall	0.78
F1-score	0.79

Source: Random Forest Validation (Workshop 4).

- Accuracy: >83%
- Feature Importance shows Fare and Sex as dominant predictors
- Emergent segregation: Class 1 agents reach safety faster
- Missing values fully resolved during preprocessing

Conclusion

- A modular architecture improves stability and interpretability
- Sensitivity and chaos analysis enhance system reliability
- The dual simulation approach validates the proposed design