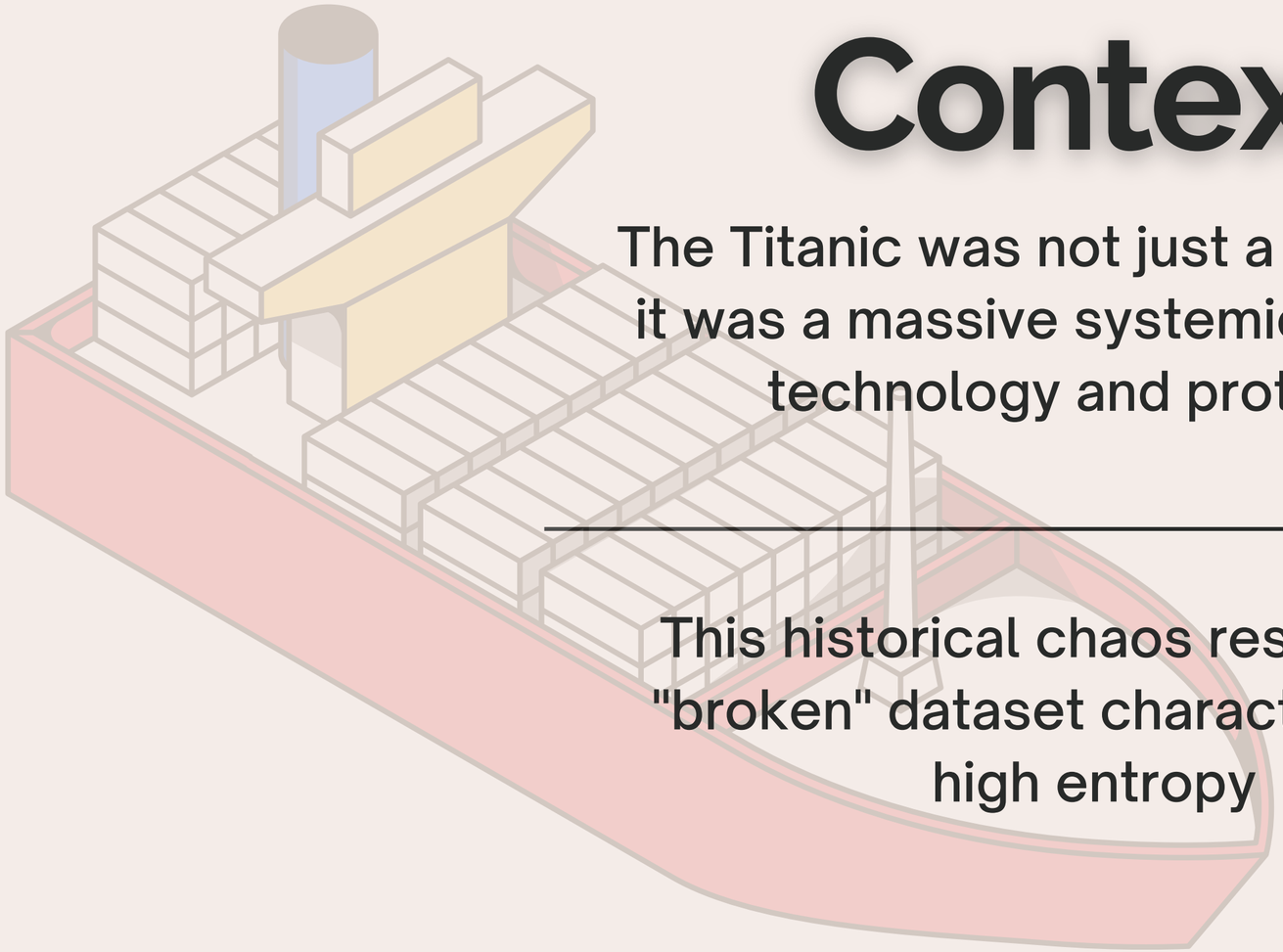
Universidad Distrital
Francisco José de Caldas
School of Engineering

Titanic

Machine learning from disaster

A. M. Cepeda Villanueva 20242020010
J. Carvajal Garnica 20242020024
A. C. Ramos Rojas 20242020005
J. D. Moreno Barragán 20201020094



Context

The Titanic was not just a shipwreck; it was a massive systemic failure of technology and protocol.

This historical chaos resulted in a "broken" dataset characterized by high entropy

Incomplete passenger records.

Fragmented family groups.

Structural Chaos

19% Missing Age Data , 77% Missing Cabin Data

Approach

We moved beyond the Kaggle competition to model a Closed System

System Components:

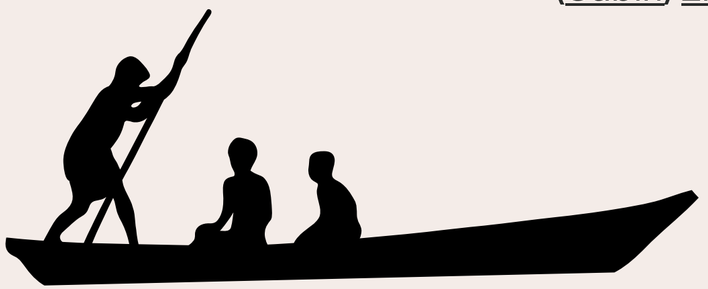
Passengers interacting within strict boundaries.

Social Rules

Stratification by *Pclass* and *Fare*

Physical Rules

Access to lifeboats determined by location (*Cabin*/*Embarked*).



Inputs
(Agents)

Age, Sex, Class

System Rules
(Internal logic)

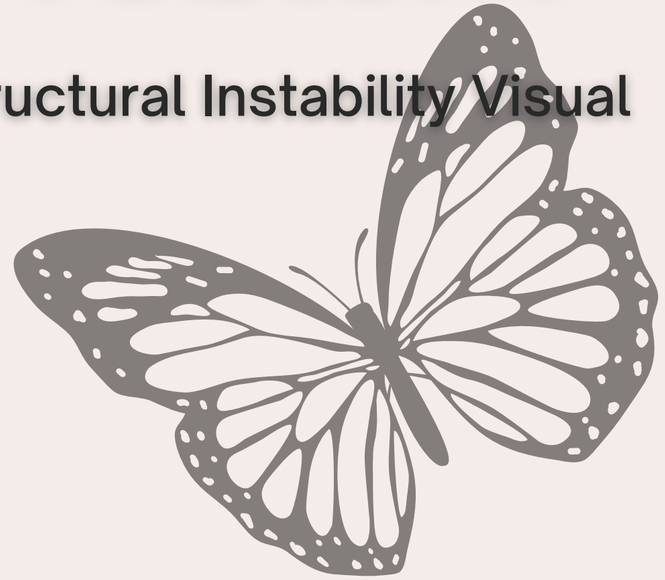
Social stratification, physical access, family groupin

System State
(Outcome)

Survived (1)
Perished (0)

The Core Problem

Structural Instability Visual



01. Extreme Sensitivity:

The "Butterfly Effect". Small input variations (e.g., Age 1) drastically flip the survival output

02. Chaos by Constraints:

Missing data acts as systemic noise, preventing deterministic prediction.

The Challenge

The goal is not just prediction, but Stabilization. We must Engineer order out of chaos.

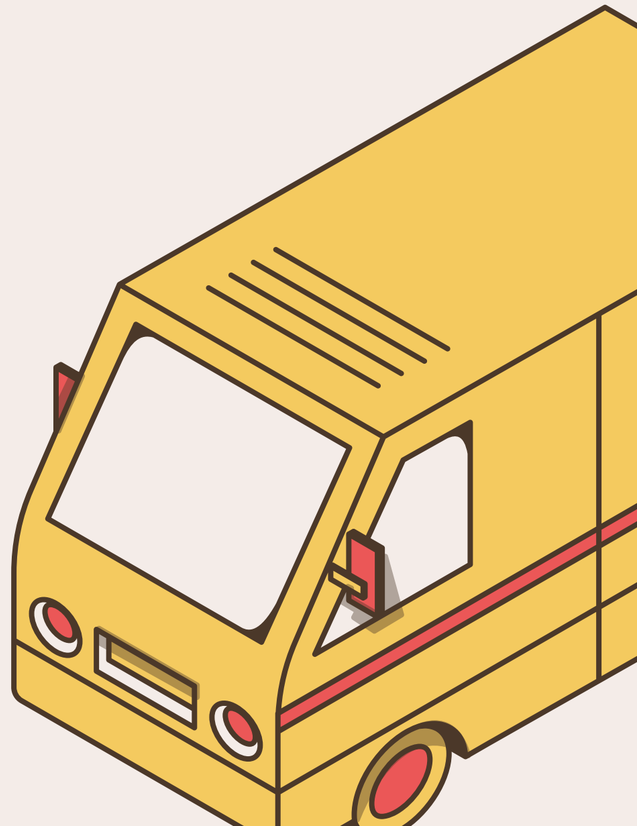
Stabilizing the System

We engineered order out of chaos by applying the Reliability Layer

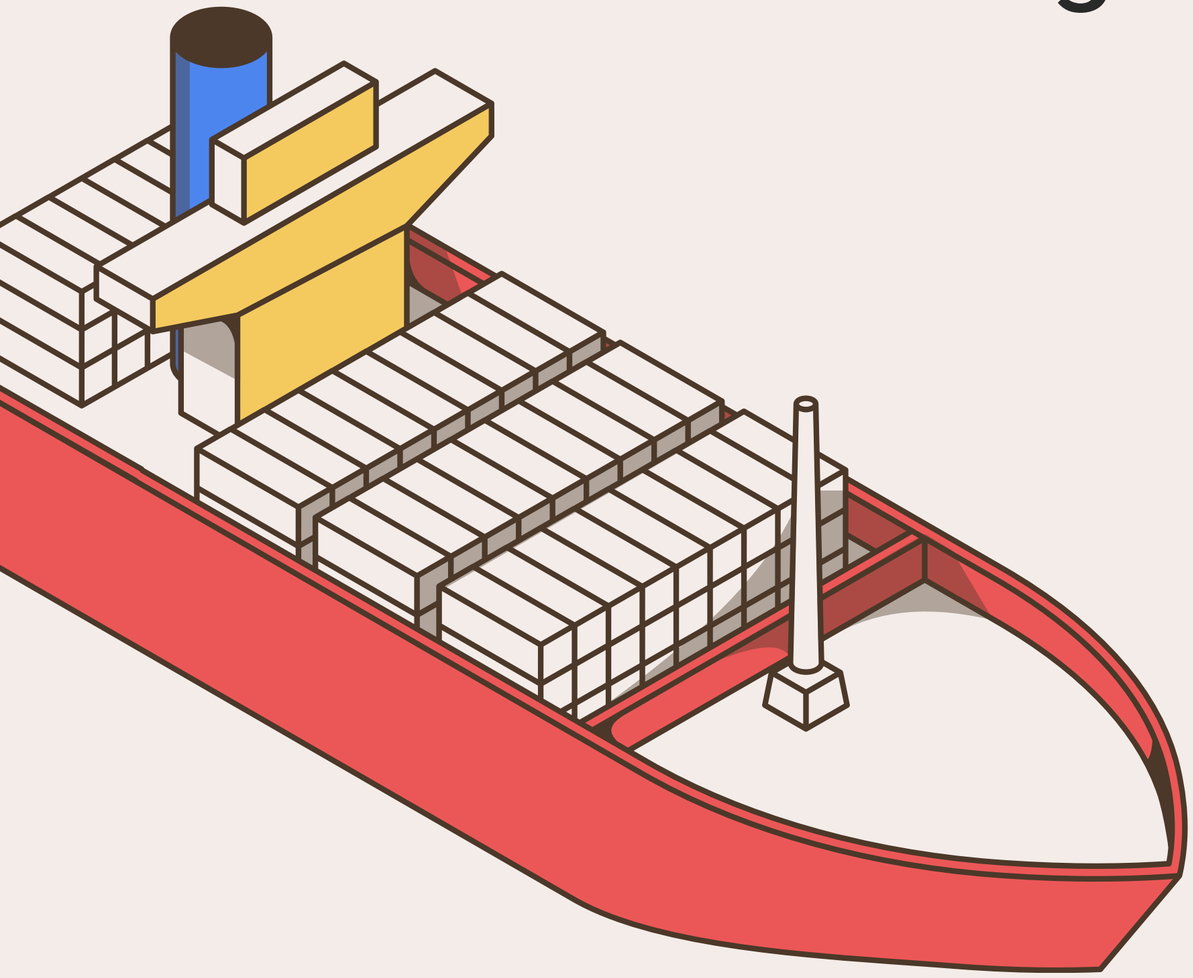
Data Imputation
Age → median
Embarked → mode
Cabin → removed (77% missing)

Noise Reduction
Encoding of Sex, Pclass, Embarked
Validation of all inputs from train.csv

Constraint Control
Elimination of structural instability in the dataset
Zero missing values after preprocessing



Engineering the Internal Logic



We transformed raw data into a coherent, deterministic system.

- **Feature Preparation**
  All passenger attributes standardized
  Consistent processing

- **Structured Pipeline**
  A modular flow:
  Ingestion → Cleaning → Feature Engineering → Simulation

- **System Readiness**
  Inputs stabilized to feed both Simulation Engines

Simulation Engine 01 — Data-Driven Model

01.

- **Random Forest (100 Trees)**
  Trained on cleaned Kaggle dataset
  80/20 split to avoid memorization

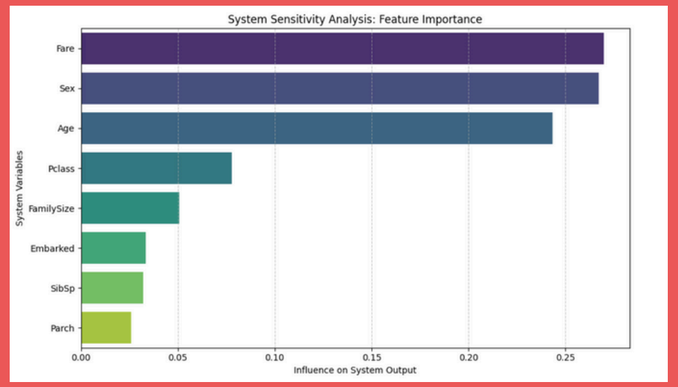
02.

- **Sensitivity Analysis**
  Extracted feature importances
  Fare, Sex, Age → dominant drivers

```python
# Modeling Layer (Data-driven)
# We initialize the model with 100 estimators for robustness
model = RandomForestClassifier(n_estimators=100, random_state=42)

# The model learns the relationship between X (Attributes) and y (Survival)
model.fit(X_train, y_train)

# Sensitivity Analysis extraction to understand system drivers
importances = model.feature_importances_
```



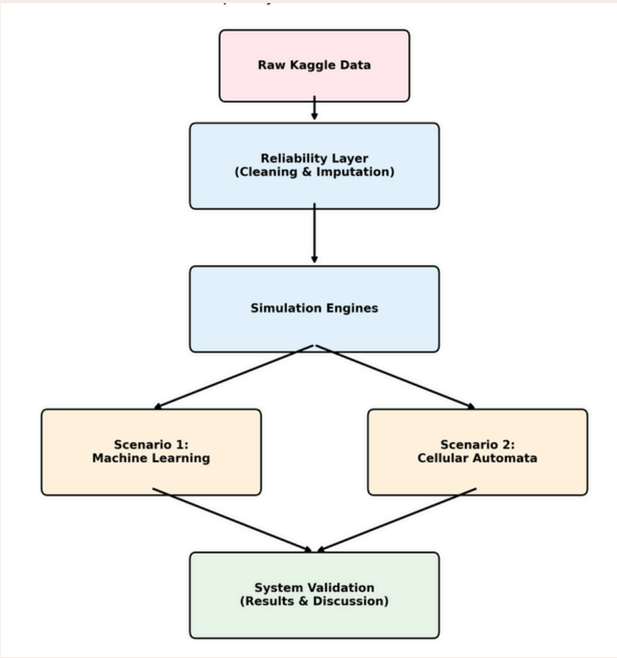
Simulation Engine 02 — Event-Based Model

01

- **Cellular Automata model**

Where each passenger acted as an agent moving on a 20×20 grid with rules based on Class and Sex. This produced emergent patterns such as blocking, bottlenecks, and class-based segregation during the simulated evacuation.

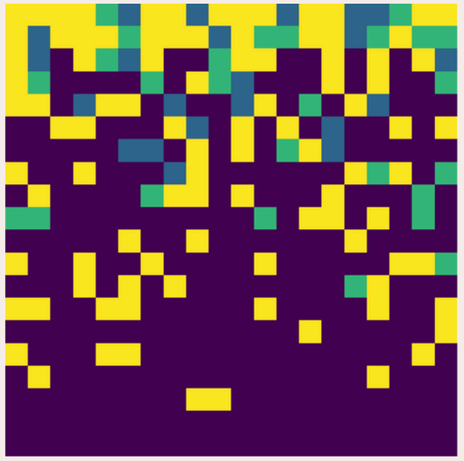
.



02

- **Emergent Behavior**

Class segregation
Bottlenecks
Top-deck saturation by higher-priority agents



Validating the Architecture

- **Reliability Confirmed**

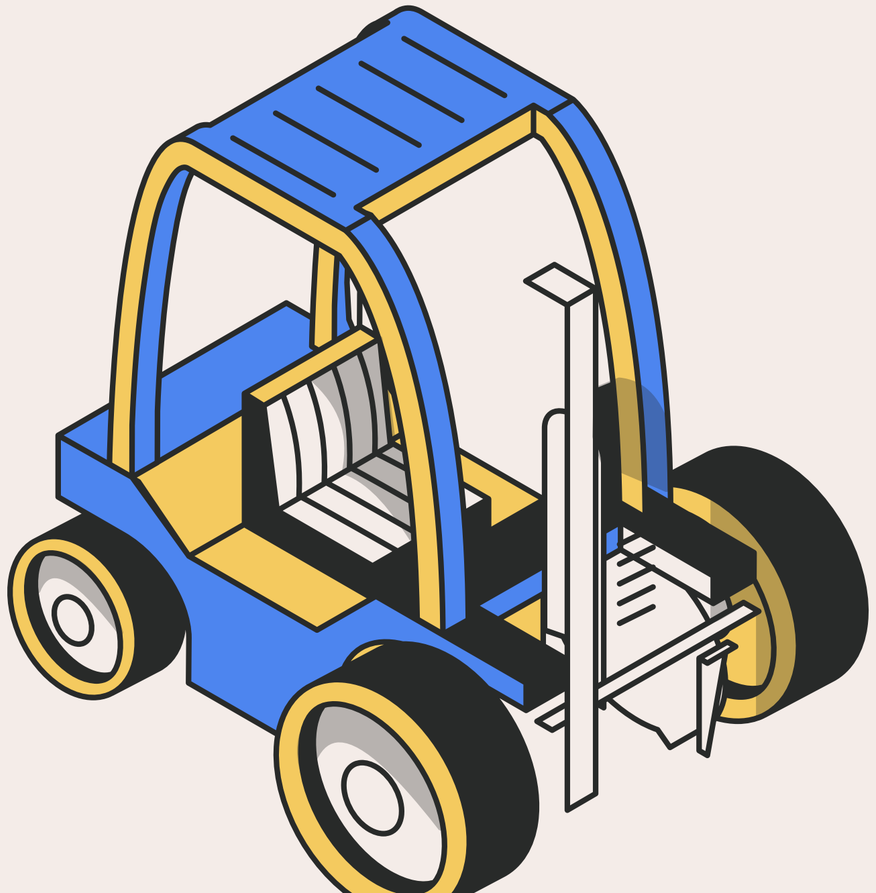
The Reliability Layer successfully absorbed missing data and prevented model failure.

- **Modularity Confirmed**

ML engine and Automata engine ran without changing the pipeline.

- **Chaos-Resistance Confirmed**

Small changes in inputs → large differences in outcomes
System behaves exactly as analyzed



Statistical Results

Sensitivity

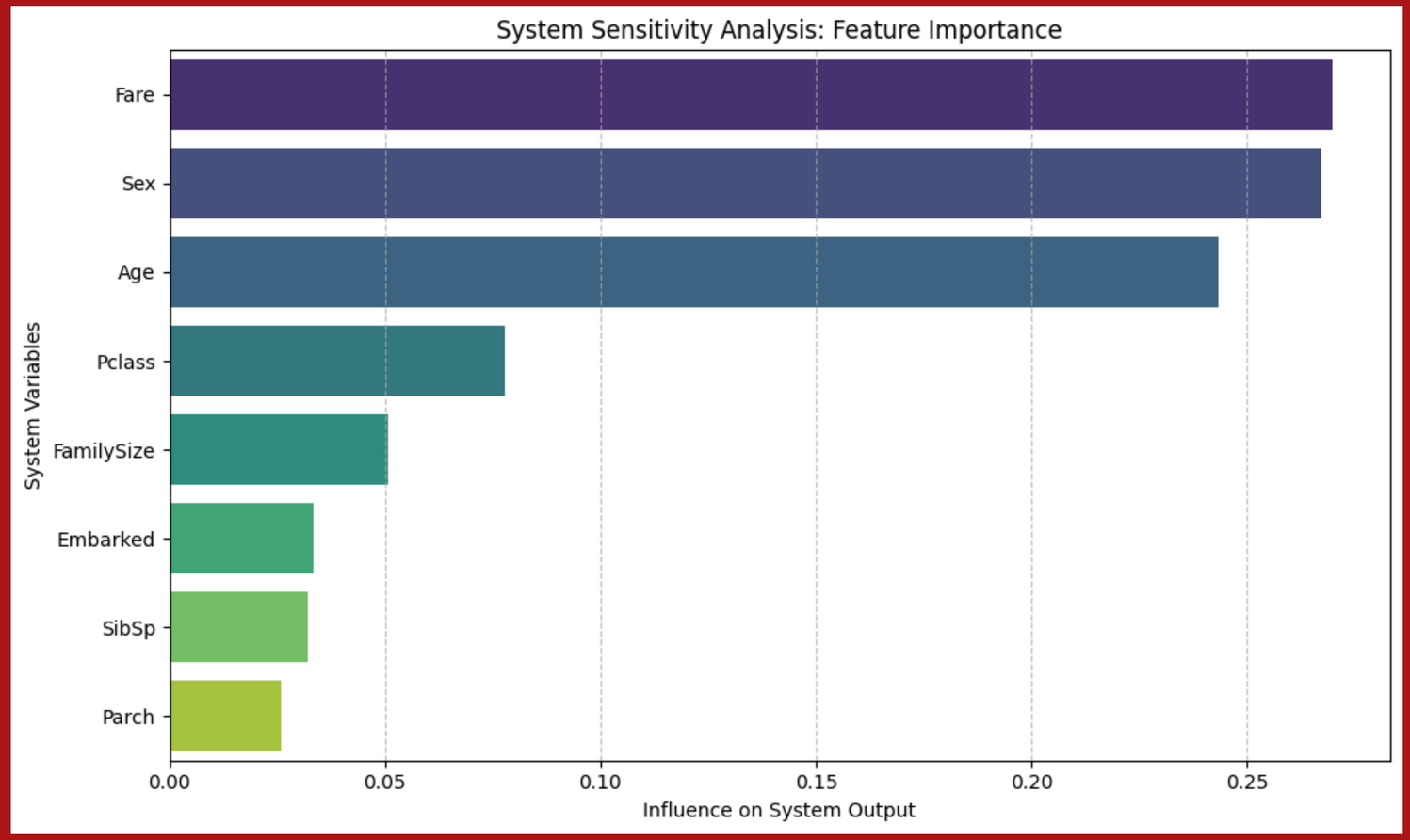
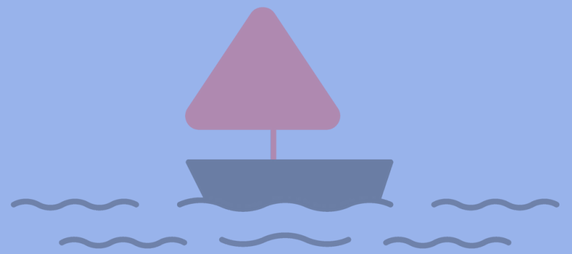


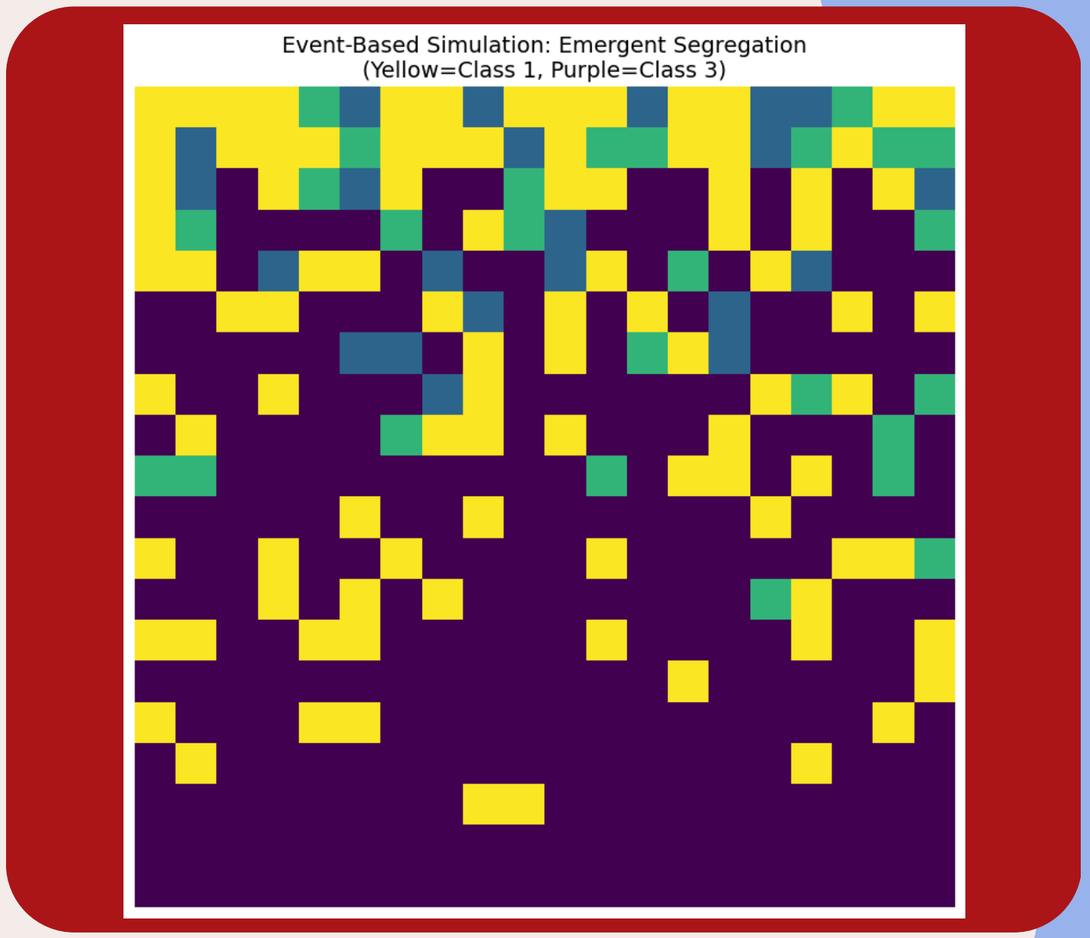
Fig 1. Feature Importance: Validating Socioeconomic Stratification

- System Stability: Achieved 83.24% Accuracy on unseen data .
- Deterministic Behavior: Survival is not random; it follows strict rules.
- Key Findings: Fare and Sex identified as the dominant system drivers.



Statistical Results

Emergency



- The Physics of Bias: Modeled evacuation with simple priority rules (PClass1) > (PClass3).
- Emergent Behavior: Class 1 agents naturally formed a physical blockade at the exits.
- Systemic Failure: Class 3 was trapped by design geometry, not by chance.

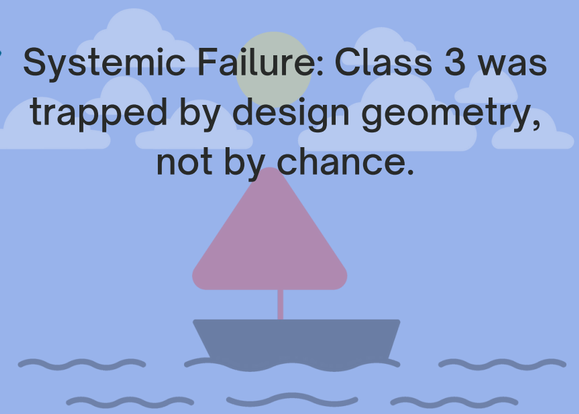
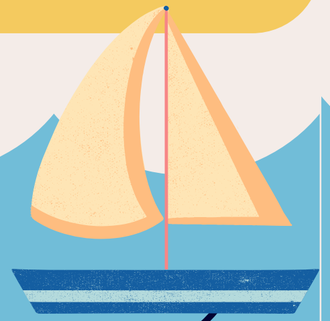


Fig 2. Emergent Segregation: Physical Blockade created by Social Priority

Titanic

Machine learning from disaster



Conclusions

Our project yielded three engineering findings:

Validated Architecture:

We demonstrated that it is possible to stabilize a chaotic system (with 19% data loss) using an ISO-25010 architecture.

Determinism:

The system is deterministic but responsive; the initial conditions (Class/Fare) dictate the survival trajectory.

Value:

Value: Applying Systems Analysis allowed us to transform a noisy dataset into a robust and explainable model.

Titanic

Machine learning from disaster

THANKS!