IEEE-CIS FRAUD DETECTION SYSTEM ANALYSIS & DESIGN

UNIVERSIDAD DISTRITAL
FRANCISCO JOSÉ DE CALDAS

LEONARDO RODRÍGUEZ SALAS - SANTIAGO MARÍN
DAVIDSON SANCHEZ GORDILLO - LUIS MARIO RAMIREZ

DATE: DECEMBER 2025
COURSE: SYSTEMS ANALYSIS & DESIGN

INTRODUCTION

This project explores the IEEE-CIS Fraud Detection challenge from Kaggle, which aims to predict the probability of online transactions being fraudulent.

At this stage, the focus is on the system analysis and the preliminary design of the architecture that will guide the future implementation phase.

GOAL

General Objective:

- To analyze and design a predictive system capable of detecting fraudulent online transactions through systems engineering principles.

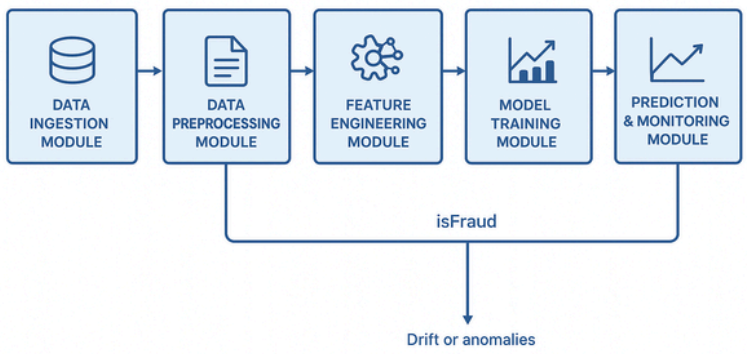
Specific Objectives:

- Understand the system's structure, relationships, and boundaries.
- Identify critical variables, constraints, and sensitivity factors.
- Propose a modular and scalable architecture for future development.

PROPOSED SOLUTION

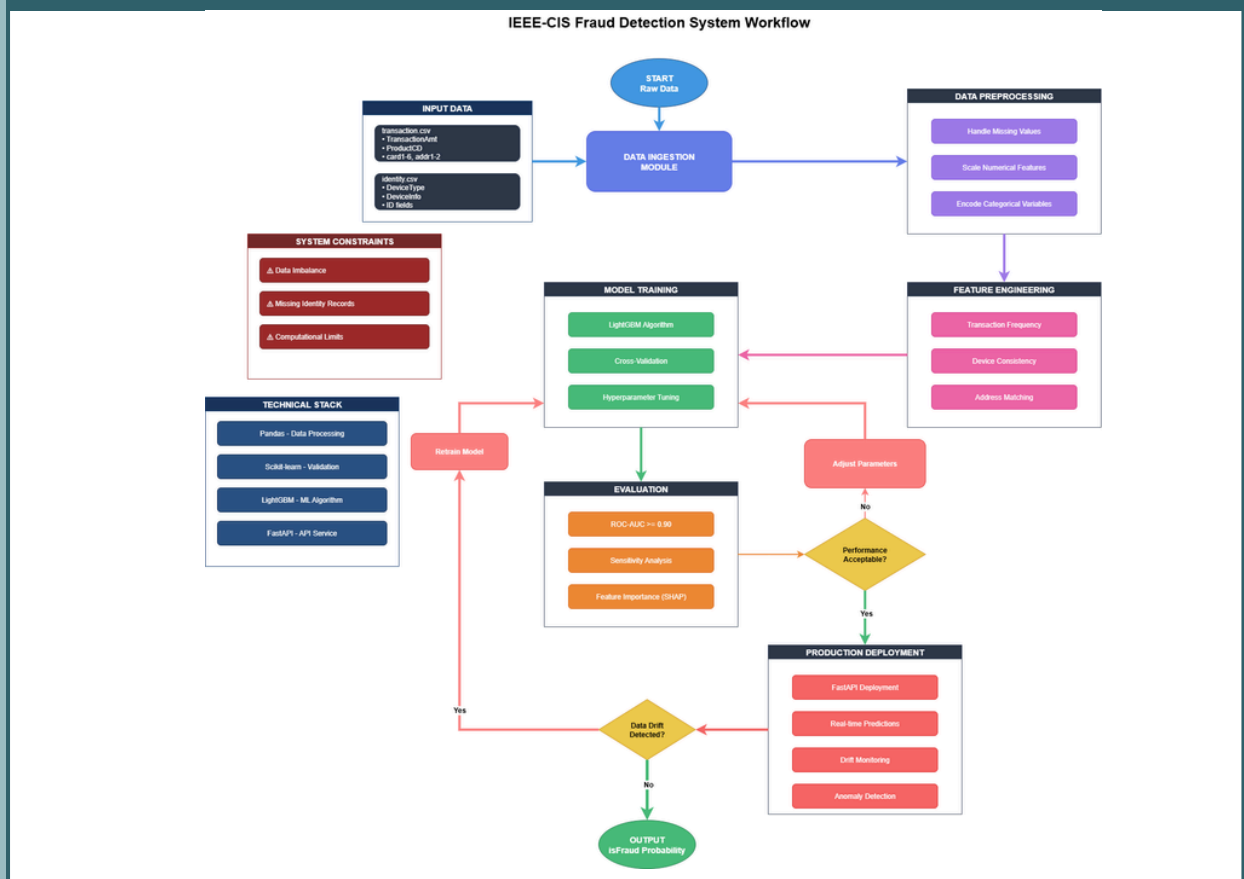
We proposed a modular system that would handle data pre-processing, feature engineering, machine learning model training, prediction & monitoring modules, and results.

This was done using the Python programming language and libraries such as pandas for data processing, scikit-learn for validation, and LightGBM for machine learning.



Kaggle Competition: https://www.kaggle.com/c/ieee-fraud-detection

WORKFLOW



EXPERIMENTS

Two main simulations were performed. The first was a machine learning-based simulation using Random Forest with different dataset sizes (200k, 400k, and 590k) and parameters such as 100–400 trees and depths of 15–25. The sensitivity of the model was evaluated by injecting noise into the data between 0.00 and 0.20.

The second was a simulation with cellular automata, using a 100×100 grid, 70–100 generations, neighborhood radii (1–3), and mutation rates (0.003–0.007). Metrics such as fraud density, entropy, and cluster formation were analyzed to study emerging patterns and chaotic behavior.

RESULTS

After the simulations and the final submission to the Kaggle competition, the results were almost satisfactory:

The internal simulations achieved high ROC-AUC values: 0.887 - 0.924

Final score in the Kaggle competition:

Public score: 0.87 (87%) Private score: 0.83 (83%)

These values represent the final performance, which meets the proposed expectations of ROC-AUC >=90 and a high score in the Kaggle competition.

CONCLUSIONS

- The system analysis identified critical variables and chaotic behavior within fraud detection processes.
- The modular design enables scalability, maintenance, and adaptability to evolving fraud patterns.
- The systems engineering approach provides an integrated view connecting data, models, and behavioral dynamics.
- We achieved the expectations set for the system, demonstrating great viability and performance for this type of problem.