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: OCTOBER 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.

SYSTEM ANALYSIS

Main Components:

- Inputs: Transactional and identity data.
- Processes: Data cleaning, preprocessing, training, and validation.
- Outputs: Probability of fraud (isFraud).
- Constraints: Class imbalance, missing identity data, privacy and computational limits.

Sensitivity Findings:

- Small variations in scaling or data handling significantly impact model performance.
- Data quality and consistency are crucial for stability.

Chaotic Behavior:

- Fraud patterns evolve dynamically over time.
- Minor parameter or feature changes can cause major shifts in accuracy.

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

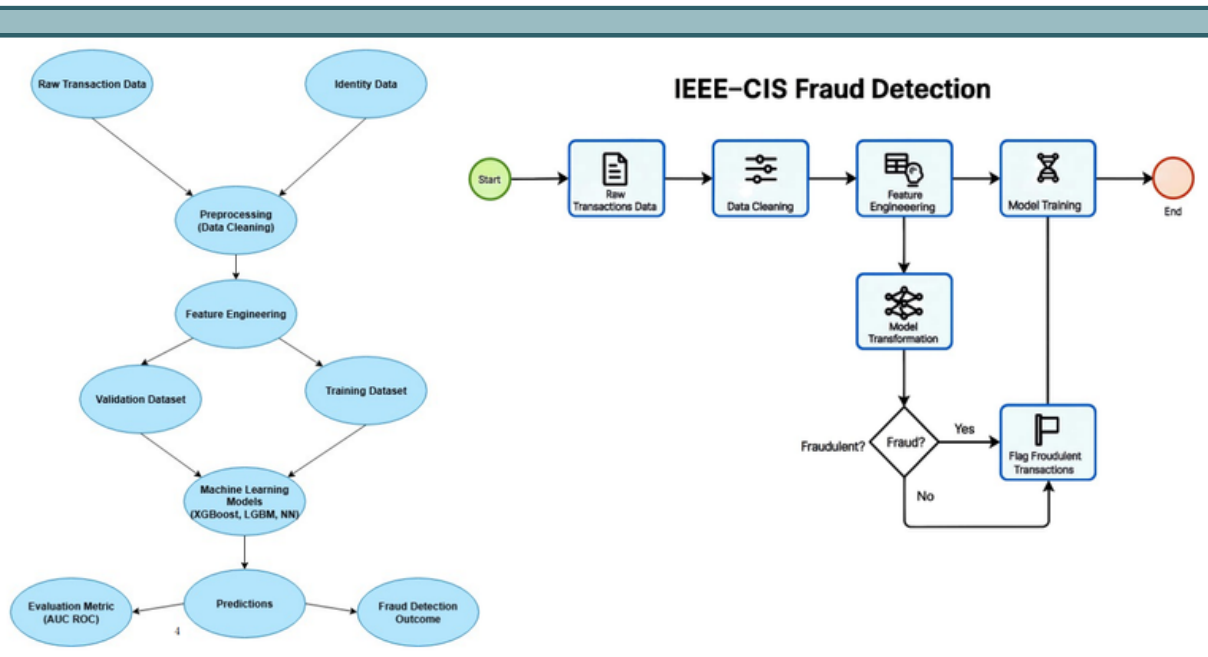
SYSTEM DESIGN

Proposed Modular Architecture:

1. Data Ingestion: Collect and validate raw transactional and identity data.
2. Preprocessing: Handle missing values, scale numeric features, encode categorical data.
3. Feature Engineering: Create derived variables (transaction frequency, device consistency, address matching).
4. Model Training and Validation: LightGBM algorithm proposed for prediction.
5. Future Deployment: Through FastAPI for real-time fraud detection.

System Requirements:

- Performance: ROC-AUC ≥ 0.90.
- Reliability: Consistent performance across datasets.
- Scalability: Ability to handle millions of transactions efficiently.
- Interpretability: Use of SHAP values for feature importance visualization.
- Security: Anonymization and encryption of sensitive data.



NEXT STEPS

The next phase will focus on developing and training the predictive model using the proposed architecture.

Validation will assess the pipeline's robustness, sensitivity to preprocessing, and consistency across data samples.

Future evaluation will include ROC-AUC metrics, feature importance analysis, and drift detection.

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.