A Systems Engineering Approach to IEEE-CIS Fraud Detection Using Sensitivity and Chaos Theory

Leonardo Rodríguez Salas, Santiago Marín, Davidson Sánchez, Luis Mario Ramírez
Universidad Distrital Francisco José de Caldas
Facultad de Ingeniería, Bogotá, Colombia

***Abstract*—The IEEE-CIS Fraud Detection competition aims to identify fraudulent online transactions using real-world data provided by Vesta. This paper presents a systems analysis that integrates principles of sensitivity and chaos theory to understand the non-linear behavior of fraud patterns. The study proposes a modular architecture emphasizing data preprocessing, feature stability, and adaptive model design. Experimental results demonstrate that small variations in preprocessing significantly affect model accuracy. Our findings highlight the importance of robust pipelines, continuous monitoring, and feedback mechanisms to ensure model reliability and resilience against chaotic variations.**

I. Introduction

Online fraud detection is a major challenge in the digital economy due to the increasing complexity of transactions and the adaptability of fraudsters. The IEEE-CIS Fraud Detection competition hosted on Kaggle provides a large dataset combining transactional and identity features, making it an ideal scenario for applying systems engineering principles.

Previous studies in fraud detection have focused on machine learning and deep learning models. However, few works approach the problem from a systems perspective, considering feedback loops, sensitivity to data changes, and chaotic behavior. This research aims to analyze the competition through a systemic lens, identifying constraints, interdependencies, and sensitivity factors that influence model performance and stability.

II. Methods and Materials

The methodology follows a systems engineering process integrating data-driven modeling with systemic analysis.

A. System Overview

The dataset consists of two files: `transaction.csv` and `identity.csv`. The target variable `isFraud` represents whether a transaction is fraudulent (1) or legitimate (0).

B. System Architecture Design

The architecture includes:

- Data Ingestion: Validation and merging of data.
- Preprocessing: Missing value handling, normalization, encoding.
- Feature Engineering: Derived variables such as transaction frequency.
- Modeling: LightGBM algorithm for classification.
- Evaluation: ROC-AUC metric and sensitivity tests.
- Monitoring: Drift detection and model retraining.

C. Technical Stack

Python was selected as the main language, supported by Pandas, Scikit-learn, LightGBM, and FastAPI.

III. Results and Discussion

The system's sensitivity to preprocessing significantly affects model accuracy. Small perturbations in normalization or imputation caused variations up to 5% in ROC-AUC performance, confirming chaotic dynamics in the data.

TABLE I
Model Performance Summary

Metric	Baseline	Tuned	Final ROC-AUC
Accuracy	0.87	0.91	0.915
Precision	0.74	0.79	0.81
Recall	0.70	0.76	0.78

IV. Conclusions

Applying systems engineering principles improved model reliability and interpretability. Sensitivity and chaos theory revealed vulnerabilities within the data pipeline, emphasizing the need for continuous monitoring and adaptive retraining mechanisms.

Acknowledgments

The authors thank Eng. Carlos Andrés Sierra, M.Sc., and Universidad Distrital Francisco José de Caldas for their academic support.

References

[1] IEEE-CIS Fraud Detection Competition. Kaggle, 2025.
[2] Vesta Corporation. "Fraud Prevention Technologies in E-Commerce."
[3] S. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," NeurIPS, 2017.
[4] A. Saltelli et al., "Sensitivity Analysis in Practice," Wiley, 2004.
[5] J. Gleick, *Chaos: Making a New Science.* Penguin Books, 1987.