Real-Time Credit Card Fraud Detection

By: Gunjan Mukeshbhai Gangwani

Student No.: 501345559

Supervisor: TBA

Date of Submission: September 22, 2025

Toronto
Metropolitan
University

Table of Contents

Abstract

This research investigates adaptive machine learning approaches for real-time fraud detection in streaming financial transaction environments. Using the Kaggle Credit Card Fraud Detection dataset (284,807 transactions), the study compares traditional machine learning models enhanced with adaptive learning capabilities against novel spiking neural networks (SNNs) for fraud pattern recognition. A PySpark-based streaming pipeline simulates real-world transaction flows, enabling evaluation of model adaptation in dynamic environments. The research examines how incremental learning strategies affect traditional algorithms (logistic regression, random forests, LightGBM) and explores whether event-driven SNNs provide superior temporal pattern recognition for evolving fraud detection. Performance evaluation focuses on accuracy, adaptation speed, and computational efficiency in streaming scenarios, contributing to the development of robust, scalable fraud detection systems.

Problem Statement

Financial fraud remains one of the most pressing challenges in today's digital economy, with global losses estimated at over $32 billion each year (Deboran, 2017). The rapid growth of digital payments, online banking, and e-commerce has not only expanded access to financial services but also created new opportunities for increasingly sophisticated fraud schemes. Traditional rule-based detection systems, though still widely used, often struggle to adapt to these evolving patterns and tend to produce high false positive rates, leading to unnecessary costs and friction for legitimate customers.

Advances in machine learning and artificial intelligence now offer more dynamic approaches to combating fraud. Real-time detection has become especially critical, as the opportunity to intercept a fraudulent transaction can be measured in milliseconds (Liu et al.). Techniques such as classification and regression are well-suited for modeling transaction data, while anomaly detection helps identify rare, suspicious activity hidden within large volumes of legitimate transactions. Predictive analytics, including time-series modeling and pattern mining, can uncover evolving fraud strategies. This project explores the application of advanced machine learning methods to design an intelligent fraud detection system capable of identifying suspicious activity quickly and accurately, while reducing false positives and preserving customer trust.

Objective: Develop an intelligent, real-time fraud detection system that can accurately identify fraudulent transactions while minimizing false positives and maintaining processing speeds suitable for real-world financial transaction processing.

The challenge encompasses various dimensions:

- Handling high transaction volumes with sub-second response times.

- Dealing with extreme class imbalance and evolving fraud strategies.
- Striking the right balance between fraud prevention, customer experience, and regulatory compliance.

Research Questions

RQ1: How effective are traditional machine learning models (logistic regression, random forests, LightGBM) when adapted with feedback mechanisms for real-time fraud detection?

Justification: Traditional models are valued in finance for their interpretability and regulatory acceptance. This research explores how they can be extended with adaptive modeling to remain competitive against evolving fraud tactics.

RQ2: How do spiking neural networks (SNN) perform in recognizing fraud patterns compared to adaptive traditional models in regards of false positives and recalls?

Justification: SNNs, inspired by biological neurons, are inherently event-driven and potentially well-suited for transaction streams. This question examines whether they provide measurable gains in detecting subtle, time-dependent fraud behaviors.

RQ3: How does model interpretability compare between adaptive traditional models and event-driven neural networks in real-time financial fraud detection?

Justification: Interpretability is a critical factor in financial fraud detection systems, where decisions often carry regulatory, ethical, and operational implications. Comparing interpretability across these approaches can provide insights into the trade-offs between predictive power and explainability, ultimately informing.

Dataset Selection and Rational

Credit Card Fraud Detection Dataset

Source: Kaggle (*Credit Card Fraud Detection* 2018)
URL: https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud
Size: 284,807 transactions, including 492 fraud cases
Features: 28 anonymized PCA features + time, amount, and class label

Rational:

- Transactions are ordered chronologically, making it suitable for replay in a streaming environment.

- Medium sized data enough to test real-time pipelines.
- Widely used in fraud detection studies, enabling comparison with prior work.
- Provides enough data points for testing adaptation techniques.

Limitations:

- Severe class imbalance requires advanced resampling and evaluation metrics.
- Anonymized features limit interpretability and custom feature engineering.
- Short timeframe (two days) may not capture long-term fraud trends.
- European-only dataset may limit generalization to other markets.

Methodology and Tools

Phase 1: Literature Review and Data Exploration

Literature Review Focus:

- Streaming machine learning architectures
- Adaptive learning algorithms for fraud detection
- Spiking neural networks in financial applications
- Performance evaluation and comparison methods

Data Exploration Techniques:

- Exploratory Data Analysis (EDA) using statistical summaries and visualization
- Transaction temporal pattern and correlation analysis
- Class distribution analysis and imbalance assessment

Tools: Pandas, NumPy, Matplotlib, PySpark

Phase 2: Streaming Pipeline Setup and Feature Engineering

PySpark Streaming Pipeline Development:

- Real-time data streaming simulation using PySpark Structured Streaming
- Stream processing pipeline design for feature computation
- Sliding window operations for temporal feature engineering

Feature Engineering Techniques:

- Automated feature selection using mutual information and correlation analysis
- Time-based feature extraction (day of week patterns, time since last transaction)
- Amount-based categorization and binning

Tools: PySpark (Structured Streaming, MLlib), Feature-selection, Scikit-learn

Phase 3: Traditional Models with Incremental Learning Adaptations

Baseline Traditional Models:

- Logistic Regression
- Random Forest
- LightGBM

Adaptive Learning Enhancements:

- Stochastic Gradient Descent (SGD) adaptations
- Incremental learning methods
- Online machine learning through River library
- MLFlow for streaming scenarios

Tools: Scikit-learn, River, PySpark MLlib, MLFlow

Phase 4: Temporal Encoding and SNN Implementation

Temporal Encoding Strategies:

- Time-to-first spike encoding for temporal features
- Population vector encoding for categorical features
- Event-driven encoding schemes for transaction sequences

SNN Training and Optimization:

- Evolutionary algorithms for network topology optimization
- Unsupervised learning through STDP
- Hybrid supervised-unsupervised training approaches

Tools: Brian2, BindsNET, PyTorch

Phase 5: Comparative Analysis and Evaluation

Comparative Analysis Framework:

- Traditional vs. incremental traditional models
- Traditional incremental vs. SNN approaches
- Different online learning strategies comparison
- Computational efficiency analysis across all methods

Performance Evaluation Metrics:

- Traditional metrics: Precision, Recall, F1-Score, AUC-ROC, AUC-PR
- Traditional incremental vs. SNN approaches
- Different adaptive learning strategies comparison
- Spike pattern analysis and visualization

Tools: PySpark, Scikit-learn, MLFlow

Phase 6: Real-time Dashboard and Documentation

Interactive Dashboard Development:

- Build a real-time fraud monitoring dashboard with Streamlit
- Fraud alerts for flagged transactions.
- Precision, recall, FPR/FNR updated in near real time.
- Feature importance (traditional models) and neuron activations (SNNs)

Documentation

- Technical notes on pipeline setup, streaming integration, and model deployment.
- Model cards: purpose, performance, interpretability, limitations.
- Reproducibility guide: code, dataset preprocessing, and instructions for running the dashboard.
- Ethical and regulatory considerations (interpretability, fairness, transparency).

Tools: Streamlit, Collab Notebooks, GitHub

Implementation Timeline

Weeks 1-2: Literature review and data exploration/preprocessing

Weeks 3-4: Streaming Pipeline Setup and Feature engineering

Weeks 5-7: Traditional model development and adaptive learning modeling

Weeks 8-10: Temporal encoding for SNNs and spiking neural network implementation

Weeks 11-12: Comparative analysis, performance evaluation, and documentation

References

Borgne, Y.-A. L. (2019). Foreword — Reproducible Machine Learning for Credit Card Fraud detection - Practical handbook. https://fraud-detection-handbook.github.io/fraud-detection-handbook/Foreword.html#acknowledgments

Deboran, A.-M. (2017, June 20). Credit Card Fraud in Canada. https://mfacc.utoronto.ca/management/media/727/download?inline

Jeyachandran, P., Akisetty, A. S. V. V., Subramani, P., Goel, O., Singh, D. S. P., & Shrivastav, Er. A. (2024). Leveraging Machine Learning for Real-Time Fraud Detection in Digital Payments. Integrated Journal for Research in Arts and Humanities, 4(6), 70–94. https://doi.org/10.55544/ijrah.4.6.10

Kumar Tambi, V. (2022). AI-Powered Fraud Detection in Real-Time Financial Transactions. https://philarchive.org/archive/VARAFD-2

Liu, C., Tang, H., & Yang, Z. (n.d.). Big Data-Driven Fraud Detection Using Machine Learning and Real-Time Stream Processing. https://www.arxiv.org/pdf/2506.02008

Machine Learning Group - ULB (2018) *Credit Card Fraud Detection*, *Kaggle*. Available at: https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud (Accessed: 20 September 2025).

Perdigão, D., Antunes, F., Silva, C., & Ribeiro, B. (2024). Improving Fraud Detection with 1D-Convolutional Spiking Neural Networks Through Bayesian Optimization. Lecture Notes in Computer Science, 14969, 127–138. https://doi.org/10.1007/978-3-031-73503-5_11