Developing a Machine Learning Model for Real-Time Fraud Detection in Online Transactions

Author: Tejashree R. Kadam
Affiliation: Dr. D. Y. Patil Arts, Commerce & Science College, Savitribai Phule Pune University Course: S.Y M.Sc. (Data Science)
Academic Year: 2024–2025

Abstract

The surge in online transactions has escalated the risks of financial fraud, demanding robust, real-time detection systems. Traditional static systems lack adaptability to evolving fraud tactics. This research proposes a machine learning-based framework for real-time fraud detection, leveraging models such as Random Forest, XGBoost, and Deep Learning, integrated with feature engineering, anomaly detection, and streaming data mining techniques. Results demonstrate high precision and adaptability, offering a scalable solution for dynamic financial environments.

1. Introduction

With the growth of digital commerce, financial fraud has become more sophisticated. Traditional rule- based systems are inadequate for identifying novel fraud patterns. Machine learning provides an adaptive mechanism to detect these behaviors through data-driven insights.

2. Problem Statement

Detecting fraud in real time involves identifying anomalies in massive, fast-moving data streams. Major challenges include class imbalance, evolving fraud patterns, data latency, and system scalability.

3. Objectives

Detect fraud in real time with low latency.

Improve detection accuracy and minimize false positives.

Develop scalable, adaptive machine learning models.

Ensure interpretability and integration into existing systems.

4. Literature Review

The literature highlights the evolution from supervised methods (e.g., logistic regression) to hybrid approaches combining anomaly detection and deep learning. Papers by Sahin et al. (2013), Phua et al. (2010), and Fiore et al. (2019) underscore the value of ensemble models and deep neural networks for fraud detection.

5. Data Collection Data sources include:
Internal logs from payment gateways and user sessions.

Public datasets (Kaggle Credit Card Fraud, IEEE-CIS).

Behavioral and session-level data (device type, IP address).
Data stream inputs for real-time processing.
6. Methodology Preprocessing:
Handling missing values, encoding categories, normalization.

SMOTE for class imbalance.
Feature engineering using transaction time, frequency, and user behavior.

Model Training:

Models: Random Forest, XGBoost, Deep Neural Networks, Autoencoders.
Evaluation: Precision, Recall, F1 Score, ROC-AUC.

Tools: Python, Scikit-learn, TensorFlow, FastAPI for deployment.

7. Results

Random Forest: F1 = 0.85; high accuracy with interpretability.

XGBoost: F1 = 0.86; best overall performance.

Deep Learning: High recall; adaptable to complex fraud patterns.

Autoencoders: Effective in unsupervised anomaly detection.

8. Challenges & Limitations Class imbalance and overfitting.

Evolving fraud techniques (concept drift).

Latency in model inference.

Privacy concerns and compliance (GDPR, PCI-DSS).

9. Future Scope

Use of explainable AI (XAI) for transparency.

Integration with blockchain and edge computing.

Multi-modal data inclusion (biometrics, behavior).

Adaptive learning and cross-platform fraud detection.

10. Conclusion

This research demonstrates the feasibility of using machine learning for real-time fraud detection in online transactions. The proposed models achieve strong accuracy while maintaining scalability and adaptability. Future work will explore edge AI and privacy-preserving computation to enhance deployment at scale.

References

(Condensed list; full references in original document)

Sahin et al., 2013, Computers & Industrial Engineering

Phua et al., 2010, IEEE TKDE

Fiore et al., 2019, Information Fusion

Bifet & Pfahringer, 2010, ACM SIGKDD

Chalapathy & Menon, 2018, ICML Proceedings