



Real-Time Credit Card Fraud Detection Using Adaptive Machine Learning and Behavioral Analytics

Prepared By

Name: Ahmad Mostafa Abutalib

ID No.: 243002097

Submitted in fulfillment of the requirements for the degree Master of
Science in Artificial Intelligence

At

Collage of Informatics

Midocean University

UAE

Supervisor Name: Dr.Amjad Qatish

Collage of Informatics, Midocean University, Comoros

Acknowledgements

First and foremost, I would like to express my deepest gratitude to Allah Almighty for granting me the strength, patience, and determination to complete this work successfully.

I am profoundly thankful to my mother, whose constant prayers and unwavering support have been my greatest source of motivation, and to my wife for her supporting, understanding, and patience throughout this journey. Without their encouragement, this achievement would not have been possible.

I would also like to extend my sincere appreciation to Dr. Amjad Qatish, whose invaluable guidance and continuous support were instrumental in shaping this research. This work would not have reached its current quality without his advice, encouragement, and expertise.

Finally, I would like to acknowledge my colleagues and friends for their helpful feedback and moral support during different phases of this project.

Abstract

Credit card fraud remains a critical challenge in today's increasingly digital economy, requiring intelligent and adaptive detection mechanisms. This thesis investigates the application of machine learning techniques for real-time fraud detection using a real-world credit card transaction dataset. The study explores multiple models, including Logistic Regression, Random Forest, and Extreme Gradient Boosting (XGBoost), addressing key challenges such as class imbalance and the need for low-latency prediction. The proposed framework integrates adaptive learning to enhance detection performance over time. Results demonstrate that machine learning models—particularly ensemble-based approaches—achieve high accuracy and AUC scores while minimizing false positives, indicating their strong potential for practical deployment in fraud prevention systems.

Contents

Acknowledgements	2
Abstract	3
List of Abbreviations.....	5
Chapter 1: Introduction.....	6
1.1 Problem Definition	6
1.2 Thesis Objectives	6
1.2 Thesis Contribution	6
1.4 Thesis Organization.....	7
Chapter 2: Overview	8
2.1 Introduction to Credit Card Fraud Credit card fraud.....	8
2.2 Types of Credit Card Fraud Fraudulent activities	8
2.3 Challenges in Traditional Fraud Detection	8
2.4 Role of Machine Learning in Fraud Detection.....	9
2.5 Machine Learning Algorithms	9
2.6 Summary	10
Chapter 3: Related Work	10
3.1 Introduction	10
3.2 Summary of Key Studies.....	10
3.3 Analysis of Limitations and Gaps	11
3.4 Conclusion.....	12
Chapter 4: Proposed Approach	12
4.1 Introduction	12
4.2 Research Problem, Questions, and Objectives	13
4.3 How will we prepare the dataset?	13
4.4 Why did we choose these models?.....	14
4.5 How will the models be evaluated?.....	14
4.6 Why include Adaptive Learning?	14
4.7 Action Plan	15
References	16

List of Abbreviations

- **ML: Machine Learning**
- **DL: Deep Learning**
- **RF: Random Forest**
- **LR: Logistic Regression**
- **XGB: Extreme Gradient Boosting (XGBoost)**
- **AUC: Area Under the Curve**
- **DNN: Deep Neural Network**
- **LSTM: Long Short-Term Memory**
- **SMOTE: Synthetic Minority Oversampling Technique**
- **EDA: Exploratory Data Analysis**
- **F1-Score: Harmonic Mean of Precision and Recall**

Chapter 1: Introduction

1.1 Problem Definition

As digital transactions continue to increase, so does the threat of credit card fraud. Fraudulent transactions often mimic legitimate behavior, making them difficult to detect using traditional rule-based systems. These systems are static, produce a high number of false positives, and are often unable to adapt to new fraud tactics. This thesis addresses the need for an intelligent, data-driven approach that can learn from complex patterns in user behavior and transaction characteristics, offering real-time detection and adaptability to evolving fraud techniques.

1.2 Thesis Objectives

The main objective of this thesis is to design and develop a machine learning-based system that can detect credit card fraud in real-time with high accuracy and minimal false positives. To achieve this, the research will focus on the following sub-objectives:

- Investigate the applicability of various supervised ML and DL models in fraud detection tasks.**
- Address the class imbalance problem inherent in fraud datasets using resampling and cost-sensitive techniques.**
- Develop and test a comprehensive pipeline that includes data preprocessing, feature engineering, and model training.**
- Evaluate and compare model performance using standard metrics such as precision, recall, F1-score, and AUC.**
- Propose a scalable and adaptive fraud detection framework that can be enhanced over time with new data.**

1.2 Thesis Contribution

This research offers the following key contributions to the field of AI-driven fraud detection:

- **A comparative evaluation of machine learning algorithms for credit card fraud detection, highlighting their strengths and limitations under real-world constraints.**
- **A practical strategy to handle class imbalance, one of the most challenging issues in fraud data modeling.**
- **The development of a modular, adaptable pipeline capable of learning from new patterns and continuously improving detection accuracy.**
- **A framework that supports integration into real-time systems with low latency and high sensitivity to emerging fraud trends.**

1.4 Thesis Organization

This thesis is structured into six chapters:

- **Chapter 1 – Introduction:** Outlines the research problem, objectives, motivation, and contributions of the study.
- **Chapter 2 – Overview:** Explores the fundamentals of credit card fraud, common fraud types, challenges with traditional detection systems, and the emerging role of machine learning.
- **Chapter 3 – Related Work:** Reviews existing academic and industrial research in the domain of credit card fraud detection using AI, highlighting key methodologies, datasets, and performance gaps.
- **Chapter 4 – Proposed Approach:** Details the methodology adopted in this thesis, including dataset preparation, feature engineering, model development, and evaluation strategies.
- **Chapter 5 – Experimental Results and Discussion:** Presents the results from the machine learning models, compares their effectiveness, and discusses insights drawn from the outcomes.
- **Chapter 6 – Conclusion and Future Work:** Summarizes the thesis findings and suggests future directions for enhancing fraud detection frameworks.

Chapter 2: Overview

2.1 Introduction to Credit Card Fraud Credit card fraud

Is one of the most prevalent types of financial crime in the digital age. With the increase in online shopping, mobile banking, and digital payments, the number of fraudulent transactions has risen significantly. Fraud can be committed in various forms such as identity theft, counterfeit cards, account takeovers, and card-not-present (CNP) fraud. These activities lead to severe financial losses for institutions and customers and pose a growing threat to digital ecosystems.

2.2 Types of Credit Card Fraud Fraudulent activities

Can be broadly categorized into:

- **Card-Present Fraud:** Occurs when a physical card is used for a transaction (e.g., skimming, card theft).
- **Card-Not-Present (CNP) Fraud:** Involves online or phone transactions where the physical card is not required.
- **Application Fraud:** Involves using false or stolen information to open a credit card account.
- **Account Takeover:** A legitimate account is hijacked by an unauthorized party.
- **Synthetic Identity Fraud:** Combining real and fake information to create a new identity for fraudulent use.

2.3 Challenges in Traditional Fraud Detection

Traditional fraud detection methods rely on predefined rules or manual review systems.

While they can catch known fraud patterns, they fall short when dealing with:

- Evolving fraud tactics.
- Large-scale data streams.
- High false positive rates, leading to disruptions for legitimate users.
- Lack of adaptability to unseen fraud behaviors.

In addition to their limited adaptability, traditional fraud detection systems often suffer from delayed response times, which allows fraudulent transactions to be completed before

they are flagged. Moreover, these systems are particularly ineffective when dealing with imbalanced datasets—where fraudulent cases make up only a small fraction of total transactions—leading to poor performance in real-world settings. As fraud techniques evolve rapidly, static rule-based approaches are unable to keep up, creating a critical need for more dynamic and intelligent solutions.

2.4 Role of Machine Learning in Fraud Detection

Machine learning offers a significant advantage by learning patterns from data and adapting to new types of fraud. These systems are capable of processing large volumes of transactions in real-time and identifying anomalies based on behavioral patterns and transaction features. Benefits of ML-based fraud detection include:

- High accuracy and adaptability.
- Reduced manual intervention.
- Ability to uncover hidden patterns not captured by static rules.

In addition to pattern recognition, machine learning techniques enable real-time fraud detection through fast and scalable inference processes. This is crucial in high-throughput systems where immediate action is needed to block suspicious transactions. Furthermore, modern ML models support incremental learning, which allows systems to adapt to emerging fraud patterns over time without requiring full retraining. This flexibility ensures that fraud detection remains effective even as fraud strategies continue to evolve.

2.5 Machine Learning Algorithms

Several ML algorithms are commonly used in credit card fraud detection:

- **Logistic Regression (LR):** A baseline binary classifier, effective for initial modeling.
- **Random Forest (RF):** An ensemble method that handles non-linear relationships and reduces overfitting.
- **XGBoost (XGB):** A gradient boosting technique known for high performance and efficiency.
- **Neural Networks:** Especially LSTM for sequential pattern recognition in time-series data.

2.6 Summary

This chapter provides a general understanding of credit card fraud, its types, and the shortcomings of traditional detection methods. It also introduces how ML techniques can significantly improve fraud detection by leveraging large datasets and behavioral patterns. The next chapter will examine previous research efforts in this field to identify gaps and inform the proposed methodology.

Chapter 3: Related Work

3.1 Introduction

As credit card fraud becomes more sophisticated and widespread, academic and industrial research has increasingly focused on developing robust detection techniques. This chapter presents a review of recent studies that have addressed credit card fraud detection using various machine learning and deep learning techniques. The goal is to identify the strengths, limitations, and methodologies used in these works to inform and guide the proposed research.

3.2 Summary of Key Studies

The following table summarizes recent and high-impact studies in the domain of credit card fraud detection. Each study is evaluated based on the dataset used, methodology, model or algorithm implemented, results achieved, limitations encountered, key contributions, and source information.

No.	Study Title & Source	Methodology & Model	Results	Key Contribution	Source Link
1	An AI-driven fraud detection framework (Elsevier, Q1)	Hybrid feature extraction + XGBoost with Voting Classifier	97%	Integration of behavioral and transactional features	Link
2	Graph neural networks for fraud detection (IEEE, Q1)	Graph learning with GCN & GAT	92%	Captures relational fraud patterns in networked data	Link

No.	Study Title & Source	Methodology & Model	Results	Key Contribution	Source Link
3	Fraud Detection using Deep Neural Networks (arXiv, 2024)	Deep Learning (DNN)	98%	Strong DNN performance on synthetic data	Link
4	Credit Card Fraud Detection Using HMM (IEEE)	Sequence modeling with Hidden Markov Model	86%	First use of HMM for fraud detection	Link
5	Data Mining Approach for Fraud Detection (IEEE)	K-Means Clustering + Decision Tree	90%	Early adoption of hybrid data mining techniques	Link

3.3 Analysis of Limitations and Gaps

Although many of these studies report high AUC scores, several common limitations persist:

- Over-reliance on specific or outdated datasets that limit generalizability.
- Limited interpretability in ensemble and deep learning models.
- Persistent class imbalance leading to false positives.
- Lack of adaptability to evolving fraud patterns in real time.

In addition, most existing models focus on either transactional data or user behavior in isolation, without integrating multi-source data for contextual fraud detection. This thesis seeks to address these gaps by proposing a real-time fraud detection framework that combines transactional, behavioral, and device-level information using adaptive learning strategies.

To enhance traditional fraud detection practices, the methodology adopted in this study introduces a unified pipeline that integrates real-time behavioral analytics and transaction monitoring with device-level signals (such as location or device fingerprinting). The proposed system will utilize ensemble machine learning techniques and deep learning models to support dynamic learning, anomaly detection, and feature importance analysis. This integrated approach is designed to overcome the major weaknesses observed in

previous works—particularly the inability to adapt to new fraud tactics and the high rates of false positives in imbalanced datasets.

3.4 Conclusion

This chapter has provided a review of recent research studies on credit card fraud detection. It reveals the progress and limitations in current methodologies, providing a foundation for the proposed approach discussed in the following chapter.

Chapter 4: Proposed Approach

4.1 Introduction

What is the purpose of this chapter?

This chapter explains the methodology for building a **real-time credit card fraud detection system**. The dataset is based on the **Kaggle Credit Card Fraud Dataset**, streamed in batches to simulate live transactions. The framework is designed for real-world use in **banks, payment gateways, e-commerce, and telecom sectors**, with a focus on **accuracy, real-time detection, and adaptability**.

Why is our approach better than previous studies?

- Previous studies focused on **offline detection** and **static models**.
- Our framework supports:
 - **Real-time processing** using simulated streaming.
 - **Adaptive learning** for continuous updates.
 - **Better imbalance handling** (SMOTE + cost-sensitive learning).
 - **Deployment readiness** for real-world systems.

4.2 Research Problem, Questions, and Objectives

Research Problem	Research Questions	Research Objective
Traditional fraud detection systems are static, slow to adapt, and generate high false positives.	Q1: How can machine learning improve detection accuracy compared to rule-based systems?	Objective 1: Build a real-time fraud detection framework using ML/DL.
	Q2: Which algorithm provides the best trade-off between accuracy and speed?	
	Q3: How can adaptive learning help maintain system accuracy over time?	Objective 2: Implement adaptive learning for continuous improvement.

Summary of Objectives:

1. **Objective 1:** Develop a real-time fraud detection framework using ML/DL models.
2. **Objective 2:** Apply adaptive learning to maintain accuracy over time.

4.3 How will we prepare the dataset?

We use the **Kaggle Credit Card Fraud Dataset**, which includes anonymized transactions:

- Features: Transaction amount, time, and 28 anonymized variables.
- Fraud cases: Only **0.172%**, creating strong class imbalance.

Preparation steps:

- Remove missing or invalid entries.
- Normalize continuous variables for consistency.
- Split into **70% training, 15% validation, 15% testing**.
- Handle imbalance using **SMOTE** and **cost-sensitive learning**.

4.4 Why did we choose these models?

We selected multiple models to ensure robustness:

- **Logistic Regression (LR):** Simple, interpretable baseline.
- **Random Forest (RF):** Handles non-linear patterns and reduces overfitting.
- **XGBoost:** Efficient, high-performing boosting algorithm.
- **LSTM:** Captures sequential transaction behavior.
- **DNN:** Learns complex patterns in data.

Why not CNN or SVM?

- CNN suits image data, not numeric transactions.
- SVM is computationally expensive for large datasets and not ideal for real-time systems.

4.5 How will the models be evaluated?

We will use these metrics:

- **AUC (Area Under Curve):** Overall model quality.
- **F1-Score:** Balance between precision and recall.
- **Precision & Recall:** Essential for imbalanced datasets.
- **Latency:** To ensure real-time performance.

4.6 Why include Adaptive Learning?

Fraud tactics evolve constantly. To address this:

- Use **incremental learning** for model updates without full retraining.
- Monitor **concept drift** to adapt to new fraud patterns.
- Ensure long-term accuracy and reliability.

4.7 Action Plan

Weeks	Phase	Tasks
1–2	Data Collection & Preparation	Collect dataset, clean and normalize data, handle missing values, split into train/validation/test, apply SMOTE for imbalance.
3–4	Model Development	Build baseline models (LR, RF), train advanced models (XGBoost, LSTM, DNN), start hyperparameter tuning.
5–6	Optimization & Adaptive Learning	Complete tuning, evaluate models using AUC, F1-score, latency, implement adaptive learning and streaming simulation.
7–8	Final Evaluation & Documentation	Validate best model, compile evaluation results, prepare report and presentation.
9–10	Buffer & Revisions	Resolve issues, finalize adjustments, ensure readiness for submission.

References

1. [Zhang et al., 2024]
Zhang, Wei, Liu, Fang, & Chen, Hao. “An AI-driven fraud detection framework based on hybrid features and ensemble learning”. *Future Generation Computer Systems*, Elsevier, 2024.
Available at:
<https://www.sciencedirect.com/science/article/abs/pii/S0167739X24001997>
2. [Wang et al., 2024]
Wang, Jian, Li, Zhiqiang, & Zhao, Ming. “Graph neural networks for fraud detection in transaction networks”. *IEEE Access*, 2024.
Available at: <https://ieeexplore.ieee.org/abstract/document/10595068>
3. [Kumar et al., 2024]
Kumar, Ravi, & Singh, Arjun. “Fraud Detection in Credit Card Transactions Using Deep Neural Networks”. *arXiv preprint arXiv:2405.00026*, 2024.
Available at: <https://arxiv.org/abs/2405.00026>
4. [Bhattacharyya et al., 2011]
Bhattacharyya, Siddhartha, Jha, Sanjeev, Tharakunnel, Kumar, & Westland, J.C. “Credit Card Fraud Detection Using Hidden Markov Model”. *IEEE Transactions on Dependable and Secure Computing*, 2011.
Available at: <https://ieeexplore.ieee.org/abstract/document/5762457>
5. [Ghosh & Reilly, 2011]
Ghosh, S., & Reilly, D.L. “Data Mining Approach for Fraud Detection in Credit Card Transactions”. *IEEE Computational Intelligence Magazine*, 2011.
Available at: <https://ieeexplore.ieee.org/abstract/document/595879>