



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
Methodology & Proposed Approach	4
Data Collection.....	4
Data Preprocessing	6
Model/Method Selection.....	7
Project Steps and Timeline	8

Methodology & Proposed Approach

Data Collection

- Kaggle dataset with anonymized transactions (only 0.172% fraud).

Papers	Year	Methods	Datasets	Results
https://ieeexplore.ieee.org/abstract/document/10595068	2024	MLP	https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud	98%
https://onlinelibrary.wiley.com/doi/full/10.1155/2023/8134627	2023	hybrid CNN-SVM	https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud	91.05%
https://ieeexplore.ieee.org/abstract/document/9480639	2021	MLP-OSCEN results	https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud	98%
https://ccis2k.org/iajit/images/year2023/No.2/20341.pdf	2023	the CNN without Pooling layer	https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud	83.72%

https://www.researchgate.net/profile/Ibraheem-Ibraheem-2/publication/377830394_Credit_Card_Fraud_Detection_Using_Improved_Deep_Learning_Models/links/65d47c8e5fbf6f295177e8a6/Credit-Card-Fraud-Detection-Using-Improved-Deep-Learning-Models.pdf	2024	LSTM	https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud	96.3%
https://www.researchgate.net/publication/340968896_Credit_Card_Fraud_Detection_Based_on_Machine_and_Deep_Learning	2020	BiLSTM-MaxPooling-BiGRU-MaxPooling	https://www.kaggle.com/datasets/muhakabartay/yourallmodelsdata	91.37%
https://journalofbigdata.springeropen.com/articles/10.1186/s40537-024-01048-8?utm_source=chatgpt.com#Abs1	2024	combining CNN, GRU, and MLP architectures	https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud	100%
https://www.researchgate.net/publication/378583709_Enhancing_Credit_Card_Fraud_Detection_A_Neural_Network_and_SMOTE_Integrated_Approach	2024	(NN) and (SMOTE)	https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud	99%
https://www.sciencedirect.com/science/article/abs/pii/S1568494620308218	2020	ANN	https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud	97%

https://www.researchgate.net/publication/348711753_Credit_Card_Fraud_Detection_Using_Artificial_Neural_Network	2021	ANN	https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud	99%
https://ieeexplore.ieee.org/document/8979331	2021	OLight GBM	https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud	98.40 %),
https://www.researchgate.net/publication/362541617_Optimizing_Data_Evaluation_Metrics_for_Fraud_Detection_Using_Machine_Learning	2022	KNN	https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud	90%.
https://www.researchgate.net/publication/343632766_Credit_Card_Fraud_Detection_using_Machine_Learning_Algorithms	2020	ANN	https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud	98.69 %
https://www.researchgate.net/publication/358872914_A_machine_learning_based_credit_card_fraud_detection_using_the_GA_algorithm_for_feature_selection	2022	PF, DT, ANN, NB, LR	https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud	99.98 %
https://www.researchgate.net/publication/336800562_Credit_Card_Fraud_Detection_using_Machine_Learning_and_Data_Science	2019		https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud	99.6%

Data Preprocessing

- Remove missing/invalid values.
- Normalize continuous variables.

- Handle imbalance using **SMOTE**

Model/Method Selection

Traditional ML Models (Baselines)

- **Logistic Regression** – simple interpretable baseline.
- **Decision Trees** – non-linear, easy to understand.
- **Random Forest** – ensemble, handles non-linear relations.
- **SVM (Support Vector Machine)** – effective in high-dimensional data.

Boosting Model

- **XGBoost (Extreme Gradient Boosting)** – efficient boosting, high accuracy for tabular data.

Deep Learning Models

- **CNN (Convolutional Neural Network), RNN (Recurrent Neural Network)** – feature extraction & sequential learning.
- **LSTM (Long Short-Term Memory), GRU (Gated Recurrent Unit), BiLSTM (Bidirectional Long Short-Term Memory), BiGRU (Bidirectional Gated Recurrent Unit)** – capture sequential and temporal transaction patterns.
- **DNN (Deep Neural Network)** – captures complex data interactions.

Optimization Techniques

- Hyperparameter tuning (grid search, learning rate).
- Regularization to reduce overfitting.

Testing & Evaluation

- Metrics: AUC (Area Under the Curve), Precision, Recall, F1-score (F1 Measure), Latency.

Project Steps and Timeline

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.

