Toronto Metropolitan University | Data Science & Analytics Graduate Studies

2024-2025 M.Sc. in Data Science and Analytics
DEVELOPMENT OF MACHINE LEARNING-BASED ALGORITHM FOR ENHANCING TRANSACTION FRAUD DETECTION

Nguyen Duy Anh Luong

Supervisor: Dr. Shengkun Xie

Objective: This project aims to develop and optimize supervised machine learning models for effective transaction fraud detection. The study focuses on identifying the best-performing model that balances high accuracy with low false positives, leveraging feature engineering, addressing class imbalance with class weighting, and exploring ensemble methods like combining XGBoost with a Deep Neural Network (DNN). The end goal is to deliver a scalable, interpretable, and production-ready fraud detection system.

Background:

Transaction fraud is one of the most pressing issues in today's digital financial systems. With millions of transactions happening every day, identifying fraudulent activity quickly and accurately is a significant challenge. Fraudulent transactions often represent a small fraction of the total volume, which makes the problem not only rare but also complex to detect without generating a high number of false positives. Financial institutions have traditionally relied on rule-based systems for fraud detection. While effective to some extent, these systems struggle to adapt to new and evolving fraud patterns. This has created a growing need for more adaptive and intelligent approaches.

Machine learning offers a promising alternative. By learning patterns from historical data, supervised ML models can automatically distinguish between legitimate and fraudulent behavior. This project investigates several of these models, including Logistic Regression, Random Forest, XGBoost, and Deep Neural Networks, with a focus on improving detection accuracy while minimizing false alarms. The study also explores how combining models (e.g., XGBoost with DNN) and using class weighting techniques can improve performance on highly imbalanced datasets. Alongside model training, explainability tools like SHAP are used to interpret model behavior and uncover key fraud-related features. The goal is to build an accurate, reliable, and production-ready system that supports fraud analysts and reduces risk in real-world applications.

Methodology:

This project used a structured, multi-phase machine learning pipeline to develop an effective fraud detection system. The key stages are as follows:

1. Data Preprocessing & Feature Engineering

The dataset was thoroughly cleaned and filtered to focus on transaction-specific features. Feature engineering was applied to extract informative variables from raw data, including time-based, amount-based, and behavioral features. Identity-based features were intentionally excluded to ensure generalizability and avoid reliance on anonymized or non-critical metadata.

2. Handling Class Imbalance

Given the extreme class imbalance (fraud vs. non-fraud), class weighting was used to penalize misclassification of the minority class and guide model training toward improved recall without overwhelming false positives.

3. Model Development

Five supervised learning models were implemented and evaluated: Logistic Regression (LR), Random Forest (RF), Standard XGBoost (XGB), Enhanced XGBoost (XGB-Adv): With advanced hyperparameter tuning, Deep Neural Network (DNN). Additionally, an XGBoost + DNN ensemble was created by combining their prediction probabilities using a weighted average to leverage both model strengths.

4. Evaluation Metrics

To assess model performance, especially under class imbalance, the following metrics were used: Precision, Recall, F1-Score, ROC-AUC and PR-AUC, Confusion Matrix analysis

5. Model Interpretation

SHAP values were used for the Enhanced XGBoost model to provide explainability and insight into how input features impact prediction. Feature importance rankings were analyzed to guide model trust and potential real-world application.

6. Model Selection & Optimization

Models were compared across all metrics, with emphasis on achieving: High fraud detection (recall), Controlled false alarms (precision), Interpretability and deployment readiness. The Enhanced XGBoost and the Ensemble model emerged as top performers, demonstrating strong balance across all criteria.

Results:

A. Overall Performance Comparison

Table 1 presents a full comparison of all models across evaluation metrics. The XGBoost + DNN ensemble achieved the highest F1-score (0.74), ROC-AUC (0.9638), and PR-AUC (0.7897), offering the best balance between catching fraudulent cases and limiting false positives.

Model	Accuracy	Recall	Precision	F1-Score	ROC-AUC	PR-AUC
Logistic Regression (LR)	77%	0.69	0.10	0.18	0.7959	0.1772
Random Forest (RF)	94%	0.68	0.33	0.45	0.9122	0.5856
XGBoost (XGB)	89%	0.81	0.22	0.35	0.9276	0.6187
Enhanced XGBoost (XGB-Adv)	95%	0.83	0.42	0.55	0.9628	0.7676
Artificial Neural Network (ANN)	98%	0.43	0.91	0.59	0.9160	0.6403
Deep Neural Network (DNN)	98%	0.51	0.82	0.63	0.9182	0.6582
XGB + DNN Ensemble	98%	0.69	0.78	0.74	0.9638	0.7897

Table 1: Performance comparison across all evaluated models

B. External Comparison Benchmark

To evaluate how well the proposed models perform beyond internal experiments, their results were compared with a strong published benchmark: the OLightGBM model by Taha and Malebary.

Model	ROC-AUC
Enhanced XGBoost (XGB-Adv)	0.9628
XGB + DNN Ensemble	0.9638
OLightGBM (Taha & Malebary)	0.9288

Table 2: ROC-AUC comparison between the proposed models and the OLightGBM benchmark

C. ROC and Precision-Recall Curve Analysis

Figures 12 and 13 illustrate the ROC and PR curves for the top-performing models. The ensemble model consistently showed superior curve performance, indicating better ranking ability and precision-recall trade-off.

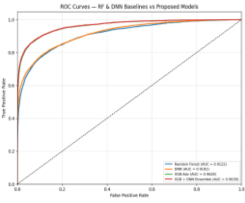

Figure 12: ROC Curves of Baseline and Proposed Models

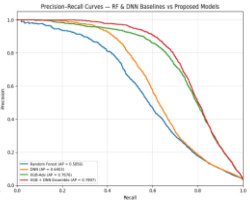

Figure 13: Precision–Recall Curves of Baseline and Proposed Models

D. Confusion Matrices

B.6 Enhanced XGBoost

Confusion Matrix:

	Predicted: Non-Fraud	Predicted: Fraud
Actual: Non-Fraud	109,145	4,830
Actual: Fraud	709	3,424

Table 15: Enhanced XGBoost - Confusion Matrix

B.7 XGBoost + DNN Ensemble

Confusion Matrix:

	Predicted: Non-Fraud	Predicted: Fraud
Actual: Non-Fraud	113,155	820
Actual: Fraud	1,243	2,890

Table 16: XGBoost + DNN Ensemble - Confusion Matrix

Conclusions:

This project evaluated several supervised machine learning models, from Logistic Regression to a hybrid XGBoost + DNN ensemble, for fraud detection. Enhanced XGBoost and the ensemble model achieved the best results, with the ensemble striking the strongest balance between fraud capture and false positives. Feature importance and SHAP analysis also showed that transaction amount, card details, behavioral patterns, and time-based features were most impactful. Identity-based features were excluded due to low added value.

Class weighting addressed class imbalance effectively, boosting fraud sensitivity without overfitting. The models were designed for real-world deployment, with a focus on interpretability, monitoring, and adaptive threshold tuning. In summary, combining tree-based and deep learning models led to accurate, scalable, and production-ready fraud detection systems.