Fraud Detection Using Machine Learning and Deep Learning Approaches

Abstract

Financial fraud causes billions of dollars in losses annually for banks, insurance companies, and e-commerce platforms. Accurate fraud detection is crucial to mitigate risks in highly imbalanced transaction data. This study investigates classical machine learning algorithms, supervised deep learning (ANN), and unsupervised autoencoder-based anomaly detection on the Credit Card Fraud Detection dataset. Results show ensemble models (Random Forest, XGBoost) and ANN achieve strong performance, while autoencoders are promising for unsupervised detection.

Introduction

Fraudulent activities in transactions pose significant challenges due to rarity of cases and evolving fraud patterns. Machine learning and deep learning methods can identify hidden patterns and anomalies. This research focuses on comparing ML, ANN, and autoencoder methods.

Literature Review

Traditional ML models (Logistic Regression, Decision Trees) offer interpretability but limited power on complex data. Ensemble methods like Random Forest and XGBoost excel. Deep learning ANNs capture non-linear features. Autoencoders reconstruct input data and are useful in anomaly detection.

Methodology

Dataset: Credit Card Fraud Detection (Kaggle). 284,807 transactions with 492 frauds. Features: PCA components, Time, Amount. Preprocessing: Standardization, SMOTE oversampling. Models: Logistic Regression, Decision Tree, Random Forest, Gradient Boosting, AdaBoost, KNN, SVM, XGBoost, ANN, Autoencoder. Metrics: Precision, Recall, F1, ROC AUC.

Results

Tree-based ensembles (Random Forest, XGBoost) and ANN achieved the best results with ROC AUC > 0.98. Autoencoder achieved ROC AUC ~0.91, showing promise for unsupervised detection. Balanced precision-recall tradeoff was observed.

Discussion

Ensembles remain reliable baselines. ANNs provide flexible modeling for large-scale data. Autoencoders enable anomaly detection where labels are scarce. Class imbalance remains a challenge, requiring advanced resampling or cost-sensitive methods.

Conclusion

Hybrid ML + DL approaches improve fraud detection robustness. Random Forest/XGBoost are practical for deployment. ANN and Autoencoder add value in complex and unsupervised contexts. This ensures high recall (catching fraud) and precision (reducing false positives).

Future Work

Explore Graph Neural Networks for fraud rings. Real-time online learning. Deploy as Streamlit/Flask web apps. Investigate federated learning for cross-institution detection.

References

1. Dal Pozzolo et al. (2015). Credit Card Fraud Detection. IEEE TNNLS. 2. Chen & Guestrin (2016). XGBoost: A Scalable Tree Boosting System. KDD. 3. Chalapathy & Chawla (2019). Deep Learning for Anomaly Detection: A Survey. 4. Kingma & Welling (2013). Auto-Encoding Variational Bayes. arXiv.