A Review of Machine Learning Models for Fraud Detection in Financial Systems

OPEN ACCESS

Volume: 13

Special Issue: 3

Month: February

Year: 2026

P-ISSN: 2321-788X

E-ISSN: 2582-0397

Citation:
S, Krishna Karthik, et al. "A Review of Machine Learning Models for Fraud Detection in Financial Systems." *Shanlax International Journal of Arts, Science and Humanities,* vol. 13, no. 3, 2026, pp. 419–23.

DOI:
https://doiorg/10.34293/sijash.v13iS3-i2-Feb.10314

Mr. Krishna Karthik S
Assistant Professor, Department of Computer Application
Thiruthangal Nadar College, Chennai

Ms. Abitha S
Assistant Professor, Department of Computer Application
Thiruthangal Nadar College, Chennai

Mrs. Sivasangari S
Assistant Professor, Department of Computer Application
Thiruthangal Nadar College, Chennai

Abstract

In today's banking on digital and payment ecosystem in electronic because of the huge fire growth in online dealings, payment on mobile and agreement systems in real-time. Knowledge of machine using ML provides a volition by learning fraud trends and patterns from history through behavioral attributes. This paper elaborate an analysis of machine knowledge models in a structured manner for discovery of fraud in systems which involves supervising methods, for example Support Vector Machines, Logistic Regression and Random Forest. On the other side, clustering, Isolation forest and One-class SVM approach towards Unsupervised and semi-supervised methodologies. Here Autoencoders with cases where labels are reviewed for fraud which are scare. Artificial Neural Networks (ANN) and Long Short-Term Memory networks are reviewed and examined for detection of fraud in sequential manner and for modelling temporal behavior. This paper discuss with the operational challenges and key research challenges which are concept drift, interpretability, severe class imbalance, privacy, constraints detecting in real-time and adaption which adversarial by fraudsters. Using metrics on performance such as model-wise evaluation and discussions, Comparative analysis are performed. Concluding with the idea of this paper which emerge directions which involves federated learning, online learning method, graph neural networks and explainable AI enhance detection of robustness in frauds in the real time scenario and financial environments.

Keywords: Fraud Detection, Financial Systems, Machine Learning, Anomaly Detection, Class Imbalance, Concept Drift, Explainable AI

Introduction

In today's world, industry of finance has rapidly moved towards digital transactions and services as online method of banking, payments through UPI and settlements systems. Even though it improves the development of that domain, it also increasing then fraud ratio in the huge scale like refund abuse, laundering of transaction, account takeover and card missing issues.

Previously, these detection methodology relies on rules defined by expert in those domains and threshold-based But in this current situation, they cannot adapt the strategies evolving around these fraud. Therefore, Machine Learning comes into play to provide the alternative methodology by finding trends and patterns from historical behavior. In these modern days, systems are using classification using supervised learning, unusual error detection and models which is sequential to identify fraud transactions. So, this paper elaborates an overview of models used under machine learning to identify key problems and errors in financial transaction, their limitations, accuracy, operational challenges and highlights their networks and explainable intelligence.

Formulation on Fraud Detection Problems

Detection of fraud in financial can be formulated by few techniques, namely binary classification - where it is classified into fraud or not, anomaly detection – it is treated as abnormal condition or behavior, sequential detection – fraud decide from a series of transaction of past over time, graph-based detection where each attributes like merchants, IP address and account are models as networks. The decision of model is based on the requirements of the objective, availability and the nature in the target environment.

Data Characteristics and Feature Engineering

These detection models on frauds use some of the attributes - amount, channel, category of merchant, contextual signals – IP, devices, location, customer account history – account age, average spendings. As fraud is all of behavior-driven, features like velocity which is transactions per hour, device indicators, distance should be engineered which makes the model high effective. Dataset is more important here, as it has some privacy constraints, imbalance and noise, so preprocessing is more important for models.

Figure 1 summarizes end-to-end detection of fraud workflow in systems in financial industry. This illustrates how raw data is transformed into decision making insights and model ready, then this will monitor continuously using evaluation metrics.

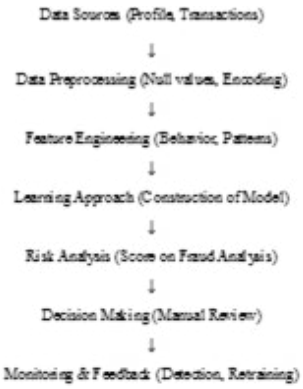

Figure 1. Pipeline for ML-based fraud detection

Supervised Machine Learning Models

This approach of learning is used widely when previous fraud history labels are available. Many approaches are used, the main one is the approach of Logistic Regression, which act as a baseline because it is fast and support needs of the objective. Another well used approach is Support vector machines which models which complex decision boundaries, but the issue is it might not support

for high volume data and transactions. On the other hand Decision Trees are explainable, but it may cause overfit, where Random Forest improves the strength by combining many trees. Gradient Boosting models provide accuracy on structured transactions by learning non linear features, it also support weight optimization. Mostly preferred models are boosted trees for their accuracy.

Unsupervised and Semi-Supervised Approaches

If the transactions are hard to find, delayed, incomplete, in these cases anomaly detection approaches comes into play. We group similar transactions using the approaches of clustering methods like K-means and DBSCAN approaches. It happen where the fraud may appear as outliers with frequent acceptable behavior. Isolation Forest works efficiently through random partitions and scaled for large datasets. One-Class SVM sensitive to parameter settings and expensive, but it learn a limit boundary around normal transactions. These decisions are more valuable as it supports signal in fraud pipelines and valuable in early detection of fraud.

Deep Learning Models for Fraud Detection

To learn complex representations from transaction data, this approach of deep learning is used. Here the main concept of Artificial Neural Network (ANN), which captures the non linearity in relationships and can transforms high dimensional data into dense numerical vectors, for high cardinality categorical variables like IDs. For semi-supervised, Autoencoders are preferred mostly as it detect by learning to reconstruct transactions normal. Here high reconstruction error indicate the potentiality of fraud. Temporal behavior are captured by models like Long Short Term Memory and it also effective for detecting suspicious transactions and account takeover. But unless carefully optimized, these deep learning are less interpretable and costly due to the low-latency deployment.

Evaluation Metrics

Detection of these frauds are highly imbalanced, therefore accuracy is not that much reliable. Precision and recall are central, as precision measures the relation of transactions that are truly fraud, while recall measures the relation of detected fraud. These two are balances by precision and recall. ROC-AUC evaluates the quality in ranking but that cant be that much optimistic under huge imbalance, whereas PR-AUC is more informative. In financial systems, because as missed fraud have different operational and monetary impacts, cost-sensitive evaluation is crucial. Under business constraints, Threshold selection often tuned to reduce expected loss.

Challenges in Real-World Deployment

Detection of fraud has some several challenges, like imbalance of class which reduces the effectiveness and requires sampling strategies or cost-sensitive training. Because of customer behavior change and fraud tactics, concept drift occurs, which requires retraining, monitoring of drift or online learning. Interpretability is necessary for regulatory compliance and trust, motivating tools such as SHAP and LIME. Constraints demand feature methodologies, which attains through screening systems. There are some privacy constraints which limits sharing of data, motivating federated learning and privacy analytics. Adversarial adaption occurs when fraudsters modifies the behavior to bypass detection, which needs monitoring and strong feature design.

Comparative Summary

Table I. Comparative Summary of Machine Learning Models for Fraud Detection

Model	Type	Key Strength	Key Limitation	Typical Use
Logistic Regression	Supervised	Interpretable, Fast	Linear Boundary	Baseline, Compliance
SVM	Supervised	Strong margins	Scaling	Medium datasets
Decision Tree	Supervised	Explainable rules	Overfitting	Rule-like decisions
Random Forest	Supervised	Robust, accurate	Less interpretable	Tabular fraud
Gradient Boosting	Supervised	High accuracy	Complex tuning	Industry standard
Isolation Forest	Unsupervised	Scalable anomalies	Flags rare legit	Early warning
One-Class SVM	Semi-supervised	Normal boundary	Slow, sensitive	Limited labels
Autoencoder	Semi-supervised	Few labels needed	Threshold selection	Anomaly detections
ANN	Supervised	Non linear learning	Black-box	Large datasets
LSTM	Sequence	Temporal behavior	Compute heavy	Sequential fraud

Emerging Research Directions

Research on detection of fraud in future cases focuses on improving strength, privacy security and adaptation of the data and models. Institutions enables federated learning to train the models which are shared without sharing raw data and privacy restrictions. Continuous learning can reduce drift by altering and updating the models as per the new patterns which appear. Patterns like fraud transactions and fraud rings are determined by Graph Neural Networks. Explainable Al is important for making decisions and minimize investigation time. Systems combine models, rules and graph-based signals are expected for deployments due to their balance of control and accuracy.

Conclusion

Financial fraud detection requires explainable and accurate decision making under high imbalance of class and fraud strategies. This review summarized Machine Learning models, detection techniques and deep learning approaches which applies in today's fraud detection systems. Ensemble tree models and boosting methods were strong for structured data, while limited scenarios are helped by autoencoding models. Concept drift, privacy constraints, adversarial adaptation and interpretability are addressed by practical deployment. Online learning, explainable Al, federated learning and graph neural networks are emerging approaches which provide directions for building real-time detecting systems in financial industry.

References

1. S. Barbon Jr., R. S. Miani, J. P. Papa, and C. O. N. Silla Jr., "A survey on credit card fraud detection," IEEE Access, vol. 8, pp. 133–156, 2020.

2. S. Tabassum, A. Ullah, and A. H. Khan, "Deep learning-based financial fraud detection: A systematic literature review," IEEE Access, vol. 8, pp. 190541–190558, 2020.
3. A. Dal Pozzolo, G. Boracchi, O. Caelen, C. Alippi, and G. Bontempi, "Credit card fraud detection: A realistic modeling and a novel learning strategy," IEEE Transactions on Neural Networks and Learning Systems, vol. 29, no. 8, pp. 3784–3797, 2018.
4. J. Gama, I. Žliobaitė, A. Bifet, M. Pechenizkiy, and A. Bouchachia, "A survey on concept drift adaptation," ACM Computing Surveys, vol. 46, no. 4, 2014.
5. T. Fawcett, "An introduction to ROC analysis," Pattern Recognition Letters, vol. 27, no. 8, pp. 861–874, 2006.
6. J. Davis and M. Goadrich, "The relationship between Precision-Recall and ROC curves," in Proc. ICML, 2006.
7. F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation forest," in Proc. IEEE ICDM, 2008, pp. 413–422.
8. T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. ACM SIGKDD, 2016, pp. 785–794.
9. G. Ke et al., "LightGBM: A highly efficient gradient boosting decision tree," in Proc. NeurIPS, 2017.
10. S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Proc. NeurIPS, 2017.
11. M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should I trust you? Explaining the predictions of any classifier," in Proc. ACM SIGKDD, 2016.
12. Z. Wu et al., "A comprehensive survey on graph neural networks," IEEE Transactions on Neural Networks and Learning Systems, vol. 32, no. 1, pp. 4–24, 2021.