Anomaly Detection in Financial Transactions

Group 14 Members
1. Ayana Damtew Tajebe
2. Beamlak Dejene
3. Anwar Mohammed Koji
4. Anwar Gashaw Yimam
5. Amen Zelealem Tadese

LITERATURE REVIEW

Introduction

Anomaly detection in financial transactions has grown more important with the exponential growth in online banking and digital payments. With fraudsters developing more sophisticated methods, traditional fraud detection systems cannot detect complex, novel threats. This research seeks to enhance financial anomaly detection by suggesting a hybrid model that combines rule-based algorithms and machine learning. There is a need to survey the literature to identify successful methods, explore gaps, and inform the creation of more accurate, scalable, and adaptive fraud detection systems.

Organization

This thematic literature review is structured around three general themes: classical rule-based and statistical approaches, classical machine learning techniques, and state-of-the-art deep learning and hybrid models for anomaly detection.

Summary and Synthesis

Chandola et al. (2009) presented a comprehensive survey of anomaly detection techniques across multiple domains. They classified methods into statistical, clustering, and classification-based models. The authors highlighted how statistical methods such as Gaussian models are interpretable but lack robustness in dynamic or high-dimensional environments. Their work set the foundation for anomaly detection taxonomy and comparison of methods across applications.

- **Key Contribution:** Establishes taxonomy and criteria for selecting anomaly detection models.
- **Methodology:** Survey and comparative analysis.
- **Citation:** Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. *ACM Computing Surveys (CSUR)*, 41(3), 1-58. https://doi.org/10.1145/1541880.1541882

Bolton & Hand (2002) examined behavior-based fraud detection in transactional systems using peer group analysis and clustering. They demonstrated that unsupervised techniques, especially profiling user behavior over time, were more effective in identifying anomalies in datasets with limited or no labels.

- **Key Contribution:** Introduced unsupervised behavior-based detection for real-world financial fraud.
- **Methodology:** Peer group analysis, statistical profiling.
- **Citation:** Bolton, R. J., & Hand, D. J. (2002). Statistical fraud detection: A review. *Statistical Science*, 17(3), 235–255. https://doi.org/10.1214/ss/1042727940

Ahmed et al. (2016) provided a focused review on machine learning techniques in network anomaly detection. Although their primary domain was cybersecurity, the relevance of models such as k-means, Random Forest, and Support Vector Machines applies directly to financial transactions. They emphasized issues like class imbalance and scalability.

- **Key Contribution:** Practical evaluation of machine learning techniques for anomaly detection.
- **Methodology:** Comparative review.
- **Citation:** Ahmed, M., Mahmood, A. N., & Hu, J. (2016). A survey of network anomaly detection techniques. *Journal of Network and Computer Applications*, 60, 19-31. https://doi.org/10.1016/j.jnca.2015.11.016

Kiran et al. (2018) discussed deep learning-based anomaly detection in high-dimensional video data. Their insights into autoencoders and convolutional neural networks have been extended to time-series transaction data in financial fraud detection.

- **Key Contribution:** Demonstrated scalability and robustness of deep learning in detecting complex anomalies.
- **Methodology:** Autoencoders, semi-supervised neural networks.
- **Citation:** Kiran, B. R., Thomas, D. M., & Parakkal, R. (2018). An overview of deep learning-based methods for unsupervised and semi-supervised anomaly detection in videos. *Journal of Imaging*, 4(2), 36. https://doi.org/10.3390/jimaging4020036

Conclusion

Literature exhibits a transition from rule-based to machine learning and ultimately deep learning models. While traditional methods are interpretable and simple, they are not adaptive. Machine learning models introduced scalability but at times lag on the interpretability side. Deep learning methods like VAEs offer computationally efficient solutions for high-dimensional and unlabeled data but at the cost of computation. The hybrid approach proposed here brings these together to better locate anomalies, injecting a balanced, practical methodology into the mounting literature on fraud detection.

DATA RESEARCH

Introduction

Financial transaction data are naturally imbalanced, high-dimensional, and sensitive and must be preprocessed with care and investigated with care. This paper examines different data sources and uses analytical methods to derive actionable outcomes for anomaly detection. The research questions addressed are discovering patterns resulting in fraud, evaluating model training data quality, and enhancing real-time detection.

Organization

The data research is structured across three sections: dataset description, data exploration and analysis, and insights derived from anomaly indicators.

Data Description

- **Source:** Publicly available synthetic datasets (e.g., Kaggle Credit Card Fraud dataset) and generated datasets via simulation scripts.
- **Format:** CSV
- **Size:** Approx. 1.3MB (284,807 transactions; 492 labeled as fraud)
- **Fields:** Transaction time, amount, anonymized features V1–V28, and class label

Rationale: This dataset closely mirrors real-world banking systems and allows evaluation of anomaly detection algorithms without compromising user privacy.

Data Analysis and Insights

- **Class Imbalance:** Fraudulent transactions represent only 0.172% of the dataset, highlighting the need for anomaly detection over classification.
- **Amount Distribution:** Fraud transactions show higher median values compared to non-fraud, especially during late hours.
- **Correlation Matrix:** Features like V17, V14, and V10 show strong correlation with fraud.

- **Descriptive Statistics:**
    - Median Transaction Amount: $22
    - Fraud Transaction Amounts: Often > $200

Visualizations: Histograms, heatmaps, and boxplots reveal patterns such as clustering of anomalies in specific time windows.

Conclusion

Exploratory analysis confirms that transaction behavior exhibits temporal and monetary patterns which anomaly models can exploit. The data supports the design of hybrid models that do not rely solely on class labels. These findings validate the use of unsupervised learning for fraud detection in highly imbalanced datasets.

Citations

- Dal Pozzolo, A., Caelen, O., Le Borgne, Y. A., Waterschoot, S., & Bontempi, G. (2015). Calibrating probability with undersampling for unbalanced classification. *2015 IEEE Symposium Series on Computational Intelligence*. https://doi.org/10.1109/SSCI.2015.33
- Kaggle. (2016). Credit Card Fraud Detection. https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud

TECHNOLOGY REVIEW

Introduction

The performance of anomaly detection systems is predominantly dependent on the technology stack. This project measures the critical frameworks and tools that enable the robust implementation, performance, and reproducibility of financial fraud discovery.

Technology Overview

- **Python:** Core programming language for data preprocessing, modeling, and evaluation.
- **Pandas & NumPy:** Efficient data manipulation and computation.
- **Scikit-learn:** Traditional ML models like KMeans, DBSCAN, SVM.
- **MLxtend:** For implementing association rule learning (Apriori, FP-Growth).
- **Keras & TensorFlow:** Deep learning library used to build and train Variational Autoencoders (VAEs).
- **Matplotlib & Seaborn:** Visualization libraries to explore trends and results.

Relevance to Project

- **Apriori & FP-Growth:** Identify common transaction rules to filter anomalies that break expected patterns.
- **Clustering Algorithms:** Detect dense regions of normal transactions; deviations are marked as anomalies.
- **VAEs:** Model high-dimensional transaction features and capture low-probability events as outliers.

Comparison and Evaluation

Technology	Strengths	Limitations
Apriori/FP-Growth	Fast rule mining, interpretable patterns	Memory-intensive on large data
KMeans/DBSCAN	Effective unsupervised clustering	Sensitive to initialization and scaling
VAEs	Excellent for modeling complex data distributions	Requires GPU and extensive training time

Use Cases and Examples

- **Paypal:** Uses autoencoders to monitor account behavior and flag fraud.
- **Amazon AWS Fraud Detector:** Combines rule engines with ML to assess transaction legitimacy.
- **Banking Sector:** DBSCAN applied in anomaly detection for real-time fraud alerts.

Identify Gaps and Research Opportunities

- Need for interpretability in deep learning-based systems.
- Difficulty acquiring real-world labeled transaction datasets.
- Potential for federated learning to improve data privacy and model generalization.

Conclusion

This project leverages a mature and scalable technology stack that blends rule mining, clustering, and deep learning. The combination supports an end-to-end pipeline for detecting and explaining anomalies. By integrating interpretable and data-driven models, the system can evolve with transaction behaviors and threats.

Citations

- Aggarwal, C. C. (2017). *Outlier Analysis* (2nd ed.). Springer. https://doi.org/10.1007/978-3-319-47578-3
- Brownlee, J. (2020). *Deep Learning for Anomaly Detection*. Machine Learning Mastery.
- Subash, P. (2021). Anomaly Detection in Financial Transactions. Medium. https://medium.com/@subashpalvel/anomaly-detection-in-financial-transactions-e895847e99d3