Detecting Anomalies in Online Banking and Credit Card Transactions Using Machine Learning

J S Haswanth Ram 23BAI1340

Sharan Karthik 23BAI1494

Avanthika V 23BAI1330

VIT CHENNAI

Abstract

With the rise of digital transactions, ensuring the security and integrity of online banking and credit card payments has become a critical challenge. Fraudulent activities, including identity theft, unauthorized transactions, and sophisticated cyberattacks, pose significant risks to financial institutions and consumers. Traditional rule-based fraud detection systems often struggle with evolving attack patterns, leading to high false positive rates and missed fraudulent activities. In this study, we explore the application of machine learning techniques for detecting anomalies in online banking and credit card transactions. By leveraging supervised and unsupervised learning approaches, including deep neural networks, ensemble methods, and anomaly detection algorithms, we aim to improve fraud detection accuracy while minimizing false alarms. Our research evaluates the effectiveness of subspace learning, one-class classification, and hybrid models in identifying fraudulent patterns within highly imbalanced transaction datasets. The proposed framework integrates advanced feature engineering, real-time detection capabilities, and scalable architectures to enhance financial security. The findings demonstrate that machine learning-based anomaly detection offers a promising solution for mitigating fraud in online transactions, ensuring a safer digital banking ecosystem.

Existing Models for Fraud Detection

- Methods & Techniques:
- Supervised Learning: Techniques like Support Vector Machines (SVM), decision trees, and ensemble methods that classify transactions based on labeled data.
- Unsupervised Learning: Clustering algorithms (e.g., k-means, DBSCAN) and density-based methods (e.g., Isolation Forest) that identify outliers in unlabeled data.
- Deep Learning: Use of autoencoders and recurrent neural networks (RNNs)/LSTMs to capture sequential patterns in transaction data.
- Hybrid Approaches: Combining rule-based systems with ML algorithms to enhance detection performance.
- Methodologies:
- Feature Extraction & Engineering: Transforming raw transaction data into meaningful features.
- Dimensionality Reduction: Techniques like PCA to handle high-dimensional data.
- Ensemble Methods: Combining multiple models to improve robustness and reduce overfitting.

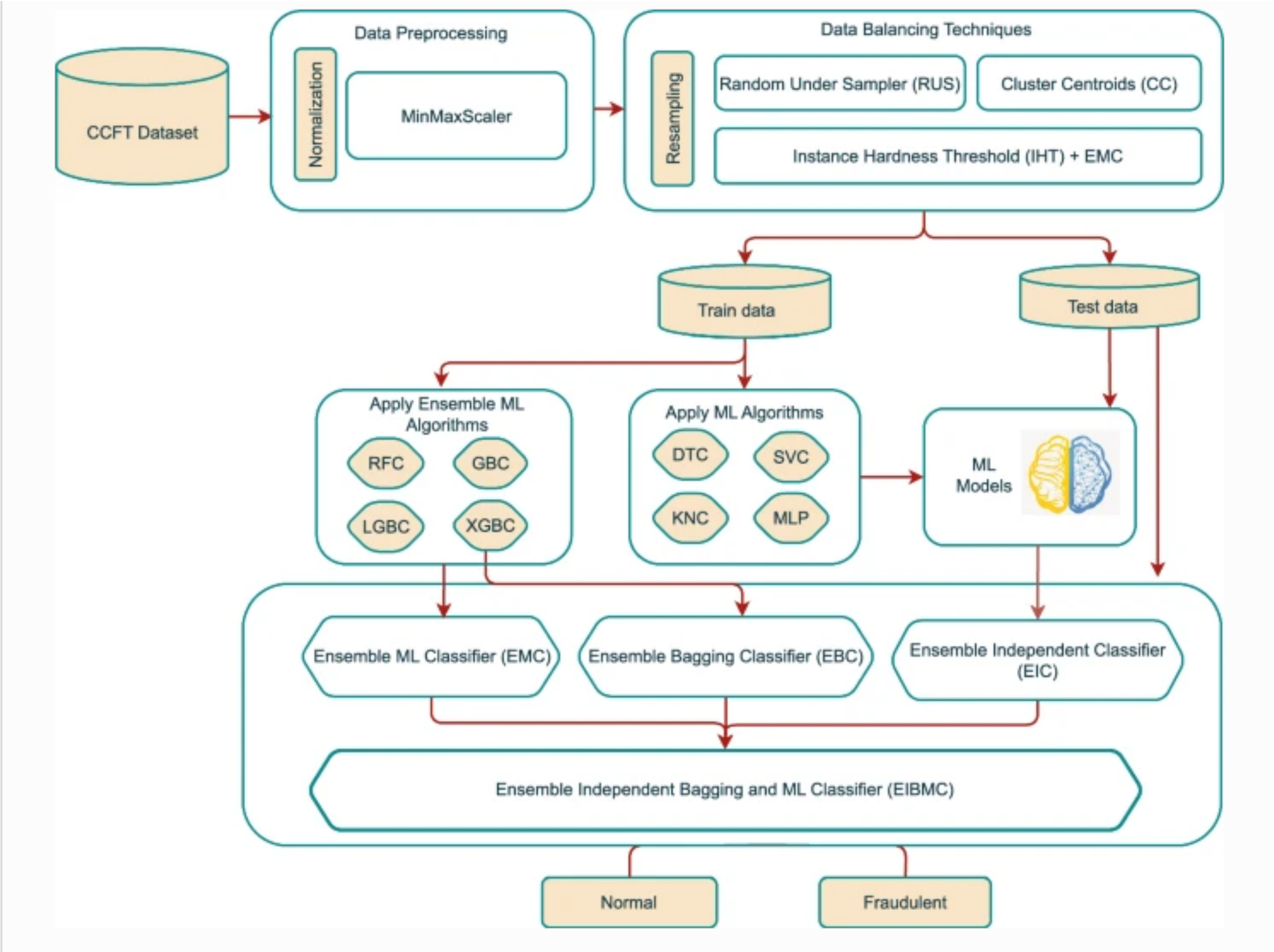
- <u>Challenges & Limitations:</u>
- Data Imbalance: Fraudulent transactions are rare, leading to skewed datasets.
- False Positives/Negatives: Balancing detection sensitivity without overwhelming systems with alerts.
- Real-Time Processing: Scalability and latency issues in live environments.
- Model Interpretability: Complex models (e.g., deep neural networks) can be hard to interpret for decision-makers.
- Adaptability: Difficulty in adapting to new or evolving fraud patterns without frequent retraining.

Proposed Machine Learning Models

- **Enhanced Deep Learning Architectures:**
- LSTM/GRU Networks: For capturing temporal dependencies in transaction sequences.
- Graph Neural Networks (GNNs): To model relationships between entities (e.g., accounts, transactions) more naturally.
- Attention Mechanisms: To focus on key features within transactions and improve model explainability.
- **Hybrid and Ensemble Approaches:**
- Combining Statistical Methods with ML: To leverage the strengths of both, ensuring robust outlier detection.
- Transfer Learning: Utilizing pre-trained models to improve performance in domains with limited labeled data.

- Datasets Required:
- Real-World Transaction Data: Anonymized banking and credit card transaction logs, including both legitimate and fraudulent cases.
- Synthetic Datasets: To supplement and balance the training data.
- Temporal & Behavioral Data: To capture the evolving patterns of user behavior and fraud tactics.

Architecture Diagram and Dataset



CCFT Dataset

Data Preprocessing

Normalization

MinMaxScaler

Data Balancing Techniques

Resampling

Random Under Sampler (RUS)

Cluster Centroids (CC)

Instance Hardness Threshold (IHT) + EMC

Train data

Test data

Apply Ensemble ML Algorithms

RFC

GBC

LGBC

XGBC

Apply ML Algorithms

DTC

SVC

KNC

MLP

ML Models

Ensemble ML Classifier (EMC)

Ensemble Bagging Classifier (EBC)

Ensemble Independent Classifier (EIC)

Ensemble Independent Bagging and ML Classifier (EIBMC)

Normal

Fraudulent

Challenges and Limitations

- High False Positives: Legitimate transactions flagged as fraud.
- Data Imbalance: Fraudulent transactions are rare.
- Adversarial Attacks: Fraudsters adapting to detection models.
- Computational Cost: Real-time fraud detection requires resources.

Future Scope and Improvements

- Real-time Fraud Detection: Reinforcement Learning.
- Self-Learning Models: Adapting to evolving fraud patterns.
- Blockchain: Secure transactions to reduce fraud.
- Federated Learning: Training ML models without centralizing user data.

References and Paper Links

1. https://www.nature.com/articles/s41599-024-03606-0

2. https://jfin-swufe.springeropen.com/articles/10.1186/s40854-023-00470-w

3. https://thesai.org/Downloads/Volume15No6/Paper_88-A_Comprehensive_Machine_Learning_Framework.pdf

4. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3670230

5.https://www.researchgate.net/publication/343051233_ANOMALY_DETECTION_IN_CREDIT_CARD_TRANSACTIONS_USING_MACHINE_LEARNING

6. https://arxiv.org/abs/2312.13896

7. https://arxiv.org/abs/2402.14389

8. https://arxiv.org/abs/2205.15300