IDENTIFYING SUSPICIOUS TRANSACTION IN FINANCIAL DATA USING MACHINE LEARNING

Chiriki Usha [1], Dadi Divya Lakshmi [2], Alla Teja Vara Prasad [3],Akkena Navya Sri [4]

Assistant Professor [1] ,UG Scholar[2,3,4]
DADI INSTITUTE OF ENGINEERING & TECHNOLOGY
Anakapalle, Visakhapatnam

ABSTRACT

Financial institutions face a constant battle against fraudulent activity. This paper investigates the effectiveness of two unsupervised learning algorithms, Isolation Forest (IF) and One-Class Support Vector Machine (One-Class SVM), in identifying suspicious activity within financial data. Both are trained on historical real time data to learn normal behavior and flag deviations as potential anomalies. The paper's core objective is to compare their accuracy in identifying these anomalies. We utilize a financial transaction dataset and perform data cleaning, feature engineering, and model training for both IF and One-Class SVM. The models then detect anomalies in unseen data, with their effectiveness evaluated through accuracy metrics. Finally, a curve visually depicts the comparative performance of each algorithm.

This paper aims to identify the most accurate and efficient method for detecting suspicious activity in financial data, aiding institutions in combating fraud. This paper explores various methodologies and techniques employed in the detection of fraudulent activities in online transactions. The abstract encompasses an overview of traditional methods such as rule-based systems and anomaly detection which are commonly utilized to analyze transaction data and detect suspicious activity. This project was directly implemented using real-time bank data. However, the outlined approach can be applied to simulated or anonymized datasets to evaluate the comparative effectiveness of Isolation Forest and One-Class SVM in identifying suspicious financial activity.

Keywords: *Unsupervised learning, Isolation Forest, One-Class Support Vector Machine (One-Class SVM), Anomaly detection, financial data analysis, Comparative performance***.**

1. INTRODUCTION

Banks play a crucial role in preventing money laundering, a process where illegally obtained funds are disguised as legitimate income. This can involve criminals depositing illicit money into accounts and then transferring it domestically or internationally through the banking system, often to evade detection by authorities. Suspicious activities encompass a broad spectrum of fraudulent endeavors that can inflict substantial financial losses and disrupt the stability of the financial system. Unlike credit card fraud, these activities directly exploit vulnerabilities within bank accounts, often involving:

Negative Balance Abuse: Malicious actors exploit loopholes or vulnerabilities to manipulate accounts into negative balances and potentially generate fraudulent withdrawals.

Suspicious Deposit and Withdrawal Patterns: Large or frequent deposits or withdrawals exceeding typical account activity may indicate money laundering or other illicit schemes.

The ramifications of these activities are far-reaching, impacting not just individual banks and customers, but the entire financial system:

Destabilizing Financial Markets: Large-scale unauthorized transactions or negative balance exploitation can disrupt the smooth functioning of financial markets.

Facilitating Money Laundering: Undetected suspicious activities can be used to launder illicit funds, further fueling criminal activity.

Traditional methods of detecting suspicious activities often rely on predefined rules that flag transactions exceeding a certain threshold or occurring outside typical business hours. However, these methods struggle to adapt to the evolving tactics of fraudsters who exploit loopholes and manipulate account behavior.

This paper investigates the potential of **machine learning algorithms**, specifically **Isolation Forest (IF) and One-Class Support Vector Machine (OCSVM)**, for identifying suspicious activities in bank transaction data. These algorithms can learn from historical data to recognize anomalies and patterns that deviate from normal user behavior. By focusing on identifying **negative balance withdrawals, unusual deposit/withdrawal patterns, and large or frequent transactions**, we aim to develop a more **adaptable and proactive** approach to detecting suspicious activities within bank accounts. Here, the core principle is to:

Identify Anomalies: Leverage machine learning to identify transactions that deviate significantly from a user's typical behavior, including negative balance withdrawals and unusual deposit/withdrawal patterns.

Flag Suspicious Activity: Based on the anomaly score generated by the model, transactions exceeding a certain threshold or exhibiting a combination of suspicious patterns will be flagged for further investigation. This may include transactions indicative of money laundering attempts.

Key components include:

- **Literature Review:** We will examine existing research on anomaly detection techniques for bank transaction data, focusing on methods that address negative balance exploitation, unusual deposit/withdrawal patterns, and identifying potential money laundering activities.
- **Methodology:** This section will delve into the specific methodologies employed in this research, including data preparation, feature engineering to capture relevant transaction patterns, model selection, training, and evaluation metrics.
- **Results and Discussion:** We will present the findings of our investigation, analyzing the performance of IF and OCSVM for identifying suspicious bank transactions, with a focus on negative balance withdrawals and potential money laundering activities.
- **Conclusion:** The concluding section will summarize the key takeaways of our research, emphasizing the potential of machine learning for safeguarding bank accounts and outlining future directions for exploration.

2. LITERATURE SURVEY

This section delves into the existing research landscape concerning online transaction fraud detection, emphasizing bank transactions and the specific challenges they present.

The document highlights the significance of data mining tools and machine learning techniques, particularly supervised and unsupervised methods, in the fight against banking

fraud. It discusses money laundering and indices associated with suspicious transactions, such as the amount of transaction exceeding a predetermined limit, sources of transfer, date and time of transaction, and change of address. The document also covers the use of fraud detection models, specifically the proposed algorithm consisting of SVM S to check the validity of new transactions.

Key Research Areas:

Machine Learning Techniques: Numerous studies explore various machine learning algorithms for fraud detection. These encompass:
Supervised Learning: Algorithms like Random Forests and Gradient Boosting leverage labeled data to learn patterns that differentiate fraudulent and legitimate transactions.
Data-driven Approaches: Research emphasizes the importance of feature engineering and data pre-processing. This involves extracting meaningful features from transaction data, including:**Temporal features-**Transaction details, date, and frequencies.**Network-based features-** Analyzing relationships between accounts and beneficiaries.**User behavioral features-** Typical transaction amounts, Date, and spending patterns.
Effective feature selection is crucial for improving model performance.

Ensemble Methods: Combining multiple machine learning models can enhance accuracy and robustness compared to relying on a single model. Studies investigate strategies for building effective ensemble models for fraud detection.

EXISTING SYSTEM

In this we already have an existing system which studys on benchmarks for credit card fraud and money laundering and obtained a precision of about 80% using the single class SVM method. Comparing the performance of SVM and BPN models over different data The dataset was classified using SVM and trained with linear regression and logic regression for detecting anomalies using the credit card features. BPN is used for training data and requires setting parameters such as the number of hidden layers, hidden nodes, training epochs, learning rate, and momentum rate.

METHODOLOGY

DATASETS:

"Bank Data Transactions Dataset": The dataset used contains attributes like Account number, Date, Transaction Details, Cheque number, Value Date, Withdrawal Amount, Deposit Amount, Balance Amount.

Sample dataset:

	Unnamed: 0	Account No	DATE	TRANSACTION DETAILS	CHQ.NO.	VALUE DATE	WITHDRAWAL AMT	DEPOSIT AMT	BALANCE AMT
0	0	409000611074	2017-06-29	TRF FROM Indiaforensic SERVICES	0.0	2017-06-29	0.00	1000000.0	1.000000e+06
1	1	409000611074	2017-07-05	TRF FROM Indiaforensic SERVICES	0.0	2017-07-05	0.00	1000000.0	2.000000e+06
2	2	409000611074	2017-07-18	FDRL/INTERNAL FUND TRANSFE	0.0	2017-07-18	0.00	500000.0	2.500000e+06
3	3	409000611074	2017-08-01	TRF FRM Indiaforensic SERVICES	0.0	2017-08-01	0.00	3000000.0	5.500000e+06
4	4	409000611074	2017-08-16	FDRL/INTERNAL FUND TRANSFE	0.0	2017-08-16	0.00	500000.0	6.000000e+06

The system will analyse transactional data from various sources, including financial institutions, e-commerce platforms, and payment gateways, to identify suspicious activities and prevent potential losses.

Data Enhancements:

Missing Data: Analyse your dataset to identify if any of the account number, cheque number, withdrawal amount, deposit amount, or balance amount fields have missing values. Missing data can significantly impact model performance. Missing Value are filled with zeros.
Data Cleaning: Ensure data consistency by checking for typos, inconsistencies in formatting (e.g., decimal placement), or invalid entries (e.g., negative balances). Clean data improves model training and reduces errors.
Feature Engineering:

It's better to choose to focus on withdrawal, deposit, and balance values. This is a good starting point, but consider creating additional features to capture richer information about transaction patterns:
Transaction Amount: Analyse the distribution of transaction amounts. You could create features like "transaction amount above/below standard deviation" to identify unusual transactions.
Transaction Frequency: Calculate features like "number of withdrawals/deposits per day/week/month" to identify deviations from typical user behaviour.
Balance Fluctuations: Features like "percentage change in balance" or "number of days with negative balance" can highlight suspicious activity.

Algorithmic Exploration:

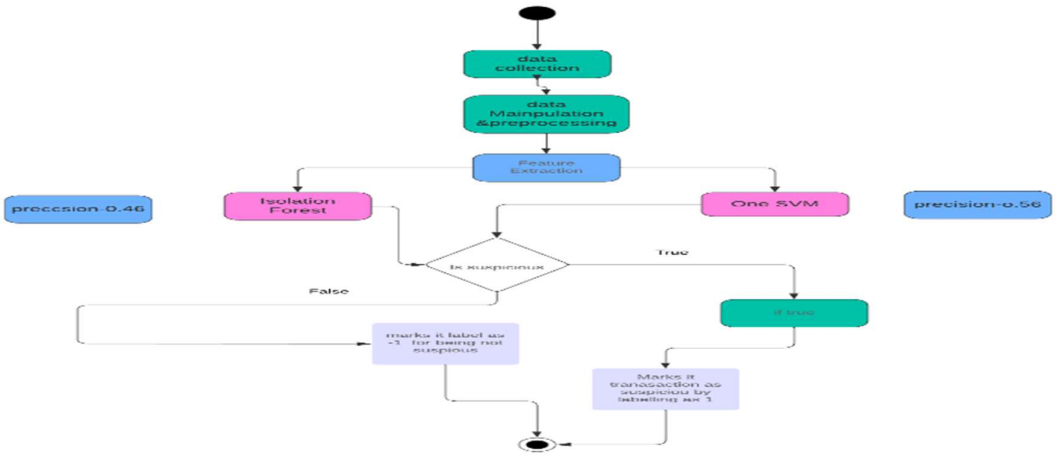
Isolation Forest (IF): This is a well-suited choice for your scenario. IF excels at anomaly detection, isolating transactions that deviate significantly from the learned pattern of legitimate transactions. You can use metrics like isolation score to identify potential outliers.
One-Class Support Vector Machine (OCSVM): Another good option. OCSVM learns a boundary around the expected data distribution (normal transactions). Transactions falling outside this boundary are flagged as suspicious.
Model Training and Evaluation:
Training-Test Split: Experiment with different ratios (e.g., 70/30, 80/20) for training and testing your models. This helps assess model generalizability on unseen data.
Hyperparameter Tuning: Use techniques like grid search or randomized search to find the optimal configuration for parameters like the number of trees in IF or the kernel function in OCSVM. This optimizes model performance.

Evaluation Metrics: Go beyond accuracy. Use precision (identifying true positives), recall (minimizing false negatives), and F1-score (balanced metric) to get a comprehensive picture of your model's effectiveness in flagging suspicious transactions while avoiding false positives.

Suspicious Activity Flagging Strategies:

Thresholding: Based on model outputs (isolation score in IF or distance from the boundary in OCSVM), set thresholds to flag transactions exceeding a certain suspicion level.

Rule-based System (Optional): In addition to models, consider incorporating expert knowledge to create rules based on historical data. These rules could involve thresholds for extreme transaction amounts, sudden large deposits, or frequent withdrawals outside typical times.



Additional Considerations:

Model Monitoring: Regularly monitor model performance over time. As fraudsters evolve tactics, your models might need retraining with new data.

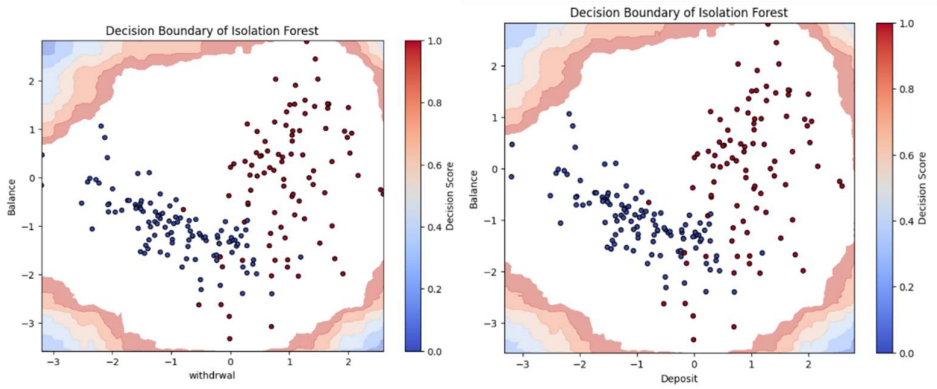
False Positive Reduction: Finding the right balance between flagging suspicious activities and minimizing false positives is crucial. Analyse false positives to understand patterns and refine your models or thresholds.

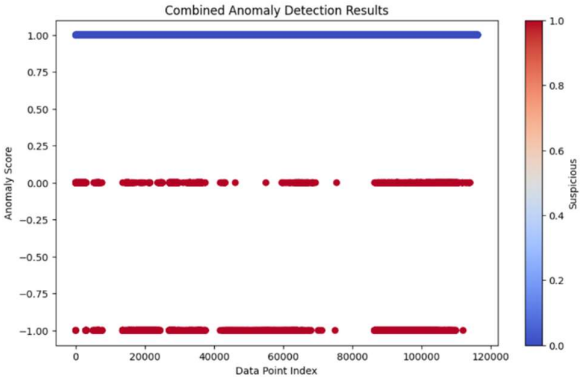
By implementing these enhancements and leveraging your chosen models (Isolation Forest and One-Class SVM), you can significantly improve your system's ability to detect suspicious activities based solely on transaction patterns within your dataset. Remember, fraud detection is an ongoing process requiring continuous adaptation and improvement.

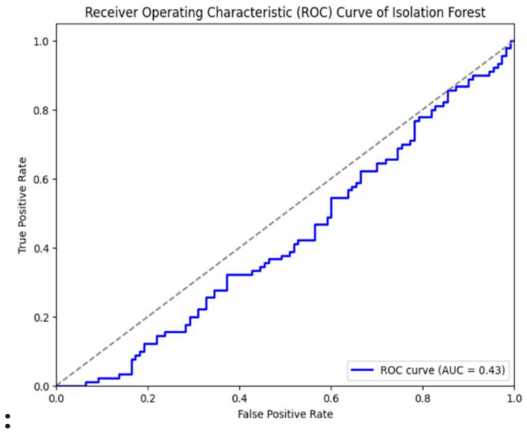
RESULTS

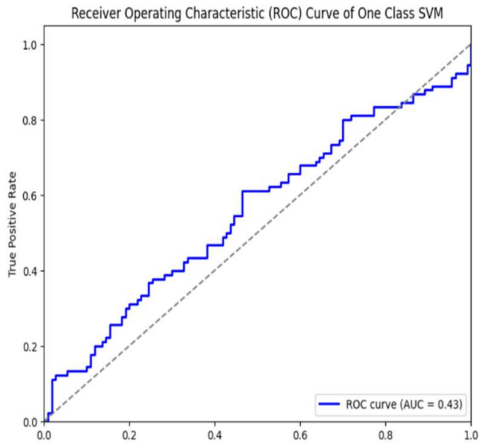
The result of detection of suspicious transaction in data using machine learning typically includes metrics such as accuracy, precision, recall, AUROC curves along with a visual inspection to evaluate the model's performance. These metrics help assess how well the model is identifying suspicious transactions compared to legitimate ones.

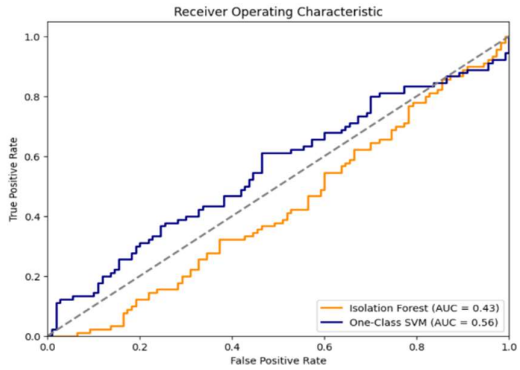
Understanding patterns of Bank data attributes and their relations:











4.FUTURE SCOPE:

This includes advancements in deep learning algorithms for better pattern recognition, integration of real-time data streams for most accurate detection, utilization of explainable AI techniques to enhance model transparency, and incorporation of blockchain technology for secure transactions and immutable records**.**

CONCLUSION

This paper explored the potential of machine learning (ML) algorithms, specifically Isolation Forest (IF) and One-Class Support Vector Machine (OCSVM), to combat suspicious activities in bank transactions. By continuously refining transaction pattern analysis methods and exploring new avenues, financial institutions can build robust and adaptable fraud detection systems.

REFERENCE

[1] Detecting Anomalies using one SVM and Isolation Forest: This explains how anomaly detection is done using one SVM and Isolation Forest. **https://ai.plainenglish.io/detecting-anomalies-a-comprehensive-guide-with-one-class-svm-and-isolation-forest-230336f0988a**

[2] **"Predicting Suspicious Activities in Bank data using online Methods (2021) ":** Predicting Financial Suspicious Activity Reports with Online Learning Methods like unsupervised neural networks to predict suspicious activity https://ieeexplore.ieee.org/document/9671716

[3] **"Online Transaction Fraud Detection using Backlogging on E-Commerce Website"** (2022): This paper explores the use of backlogging, a technique that temporarily delays transactions for further analysis, to improve fraud detection in e-commerce. Gyanangshu Misra || https://www.ijert.org/research/online-transaction-fraud-detection-using-backlogging-on-e-commerce-website-IJERTV11IS050319.pdf

[4] **"Online payment fraud: from anomaly detection to risk management" (2022):** This paper reviews the use of machine learning for online payment fraud detection, highlighting the shift from anomaly detection to risk management approaches. Paolo Vanini, Sebastiano Rossi, Ermin Zvizdic & Thomas Domenig || https://jfin-swufe.springeropen.com/articles/10.1186/s40854-023-00470-w

[5] **"Online Transaction Fraud Detection using Machine Learning"** (2023): This paper provides a comprehensive overview of machine learning techniques used in online transaction fraud detection, including decision trees, support vector machines, and neural networks. Deepakshi Mahajan || https://www.geeksforgeeks.org/online-payment-fraud-detection-using-machine-learning-in-python/