International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Detection & Suspicious Financial Flow by Using Autoencoder and RBA

Yerrabogulu Dilip Kumar[1], G.V.S.Ananthnath[2]

[1]Student, Mca 2nd Year, Kmm Institute Of Postgraduate Studies, Tirupati
[2]Associate Professor,Dept Of Mca, Kmm Institute Of Postgraduate Studies, Tirupati

Abstract
Detection and prevention of fraudulent transactions in e-commerce platforms have always been the focus of transaction security systems. However, due to the concealment of e-commerce, it is not easy to capture attackers solely based on the historic order information. Many researches try to develop technologies to prevent the frauds, which have not considered the dynamic behaviors of users from multiple perspectives. This leads to an inefficient detection of fraudulent behaviors. To this end, this project implements the RBA and DNN algorithms by combining internal control risk factors with the existing AML algorithms. Model selection is performed on the base of POC Data, and AE is found to be the most suitable model for unsupervised learning. The predictive model aims to provide accurate predictions for new data, that is, data not used during model training. The objective is to enhance the generalization performance of the predictive model. The predictive model includes hyperparameters that are closely aligned with the training data. Selecting hyperparameters that closely match the training data often leads to overfitting, which causes performance loss. To address this problem, dropout was used during the learning process.

Keywords: Risk-based approach (RBA), anti money laundering (AML), autoencoder, money laundering symptoms, suspicious transaction report (STR)

INTRODUCTION

Financial fraud, including money laundering and unauthorized access, poses significant challenges to financial institutions globally. Traditional rule-based systems struggle to adapt to the dynamic nature of fraudulent activities, often resulting in high false positives and negatives. Machine learning and deep learning techniques offer innovative solutions to these challenges.

Autoencoders, a type of unsupervised neural network, are particularly suited for anomaly detection. By learning compressed representations of normal transaction data, autoencoders can efficiently detect deviations indicative of fraudulent activity. Coupled with a risk-based evaluation framework, the proposed system prioritizes high-risk transactions for further investigation, thereby optimizing resources and reducing manual effort.

This document outlines the development and implementation of a Suspicious Financial Transaction Detection Model using an autoencoder and a risk-based approach, discussing its methodology, advantages, and potential for real-world application.

Financial institutions are at the forefront of combating illicit activities such as money laundering, embezzlement, and unauthorized access to sensitive financial data. With the increasing volume and

complexity of financial transactions, the task of detecting suspicious activities has become more challenging. Existing systems rely heavily on rule-based mechanisms, which, while effective for standard scenarios, fall short when faced with evolving fraud techniques. Moreover, the growing emphasis on regulatory compliance, such as anti-money laundering (AML) and counter-terrorist financing (CTF) laws, underscores the need for robust detection systems.

Deep learning, particularly through the use of autoencoders, offers a promising solution. Unlike supervised learning techniques that require labeled data, autoencoders operate in an unsupervised manner, making them well-suited for environments where fraud patterns are diverse and unpredictable. The incorporation of a risk-based approach ensures that the flagged anomalies are evaluated systematically, prioritizing cases that pose the highest threat. This combination creates a balanced, efficient, and effective fraud detection mechanism that can adapt to the dynamic nature of financial ecosystems.

LITERATURE SURVEY

The rise in digital financial transactions has led to an increased risk of fraud, prompting researchers to explore advanced artificial intelligence (AI) and deep learning (DL) techniques for fraud detection. The reviewed literature highlights diverse approaches, from model adaptation to privacy-preserving analytics and hybrid learning systems, which have demonstrated significant potential in enhancing the robustness and accuracy of fraud detection systems.

Zhang and Chan [1] introduced **Apricot**, a novel weight-adaptation technique designed to correct deep learning models without retraining them entirely. This approach offers a promising direction for maintaining model performance over time, particularly when models encounter unforeseen data distributions, such as fraudulent behaviors not seen during training.

Fang et al. [2] proposed a deep learning-based anti-fraud model tailored for internet loans, addressing the complexities of online financial fraud. Their study emphasized the importance of adapting fraud detection systems to dynamic fraud patterns and integrating domain-specific knowledge.

In the context of privacy, Dhasarathan et al. [3] introduced a homomorphic privacy enforcement approach to analyze COVID-19 health data. Although focused on healthcare, the method illustrates effective strategies for preserving data privacy in sensitive applications, which can be extended to financial fraud detection systems.

Verma and Misra [4] developed a **two-layer deep learning model** combined with a **self-improved Honey Badger algorithm** for detecting financial fraud. Their method demonstrated high accuracy and adaptability in complex institutional environments, showcasing the benefit of bio-inspired optimization in fraud detection.

Chen et al. [5] explored **variational autoencoders (VAEs)** and **Wasserstein generative adversarial networks (WGANs)** to improve anti-money laundering (AML) processes. By generating realistic synthetic data and modeling complex fraud patterns, their model enhanced the detection of subtle fraudulent activities.

Raval et al. [6] proposed **RaKShA**, a trusted and explainable LSTM-based model designed to classify fraud patterns in credit card transactions. The integration of explainability addresses the critical issue of model interpretability, which is crucial for regulatory compliance and trust in AI systems.

Carcillo et al. [7] investigated a hybrid approach by combining **unsupervised and supervised learning** for credit card fraud detection. This method effectively utilizes unlabeled data to enhance the learning process, making it suitable for real-world applications where labeled fraud data is often scarce.
Lastly, Saumya and Singh [8] employed an **LSTM autoencoder** for unsupervised spam review detection, demonstrating the efficacy of autoencoders in detecting anomalies without labeled data. This technique can be repurposed for financial fraud scenarios, where anomalies often manifest as subtle deviations from normal transaction patterns.

EXISTING SYSTEM

Traditional financial fraud detection systems predominantly rely on:

- **Rule-Based Detection**: Predefined rules flag suspicious transactions based on thresholds like transaction amount or frequency.
- **Statistical Models**: Utilize historical data to calculate probabilities and flag anomalies.
- **Manual Monitoring**: Teams manually review flagged transactions for suspicious patterns.
- While these systems have been effective in some cases, they face significant limitations:
- High false positives and negatives due to static rules.
- Inability to adapt to new fraud techniques.
- Resource-intensive manual review processes.
- Moreover, these systems often lack scalability and struggle to process large volumes of transactional data in real time.

DISADVANTAGE:

- High False Positives: Legitimate transactions are frequently flagged, leading to unnecessary delays and customer dissatisfaction.
- Static Rules: Fraudsters can easily circumvent predefined thresholds by adapting their methods.
- Scalability Issues: Rule-based systems are not designed to handle the exponential growth in transaction data.
- Resource Intensive: Manual reviews demand significant time and effort, diverting resources from other critical tasks

PROPOSED SYSTEM

- **Autoencoder for Anomaly Detection**:
- Trains on normal transactional data to learn a compressed representation of legitimate patterns.
- Flags transactions with high reconstruction errors as anomalies.
- **Risk-Based Approach**:
- Evaluates flagged anomalies using a scoring system based on parameters like transaction location, time, frequency, and monetary value.
- Prioritizes high-risk anomalies for further investigation.
- **System Integration**:
- A dashboard provides real-time monitoring and alerts for suspicious transactions.
- Feedback loops update the model with new data to improve accuracy.

Unsupervised Anomaly Detection: The autoencoder model learns the normal patterns in financial transaction data by encoding it into a compressed format and re
constructing it. Transactions with high reconstruction errors indicate deviations from normal behavior and are flagged as potential anomalies.

Risk-Based Transaction Evaluation: Each flagged anomaly is further analyzed using a risk assessment framework. Parameters such as transaction size, frequency, geographic origin, and account behavior history are used to assign a risk score. Transactions exceeding a predefined risk threshold are escalated for manual review or automated alerts.

Feedback Mechanism for Continuous Learning: A feedback loop ensures that the system adapts over time. False positives and negatives are used to fine-tune both the autoencoder model and the risk evaluation framework, making the system increasingly accurate.

Scalable and Secure Architecture: The system is designed to handle high transaction volumes in real-time, leveraging cloud computing for scalability and encryption technologies to ensure data security.

ADVANTAGES:

- Outlier Removal: Detects and removes extreme values in the dataset to avoid skewed model training.
- Feature Normalization: Ensures all transaction features are on a comparable scale to improve model accuracy.
- Dimensionality Reduction: Reduces noise by selecting only the most relevant attributes for transaction analysis.

MODULES

- Data Collection and Preprocessing
- Autoencoder-Based Anomaly Detection
- Risk Scoring Framework
- Alert and Monitoring System
- Feedback and Model Updating

Data Collection and Preprocessing: This module focuses on gathering transactional data from various sources, such as banking systems, online payment platforms, and financial databases. The collected data undergoes thorough preprocessing to ensure it is suitable for training the machine learning model.

Data Sources: Includes structured financial data such as transaction IDs, timestamps, account details, transaction amounts, locations, and categories.

Cleaning: Handles missing values, duplicates, and outliers to improve data quality.

Normalization: Scales features like transaction amounts and frequencies to a uniform range, ensuring that no single feature dominates the model.

Feature Engineering: Extracts and transforms raw data into meaningful attributes such as transaction velocity, time of day, and account usage history to enhance model learning.

Autoencoder-Based Anomaly Detection: Employs an unsupervised learning model to learn patterns in legitimate data.

Flags transactions with high reconstruction errors.

Risk Scoring Framework: Assigns risk scores based on transaction parameters.

Ranks flagged anomalies for prioritization.

Alert and Monitoring System: Displays suspicious transactions on a user-friendly dashboard.

Sends real-time alerts to relevant stakeholders.

Feedback and Model Updating: Incorporates user feedback into the training process for continuous improvement.

Autoencoder-Based Anomaly Detection

This module utilizes a deep learning autoencoder to identify anomalies in financial transactions by learning the normal transactional patterns and flagging deviations.

Model Architecture: The autoencoder consists of an encoder that compresses input data into a latent representation and a decoder that reconstructs the original input. Reconstruction errors are analyzed to identify anomalies.

Training: The model is trained exclusively on legitimate transactional data to learn the patterns and reduce false positives.

Detection: Transactions with high reconstruction errors are flagged as suspicious, indicating deviations from normal behavior.

ALGORITHM

Autoencoder:

Architecture: Encoder compresses input data; decoder reconstructs it.

Loss Function: Measures reconstruction error to identify anomalies.

SMOTE-ENN (Synthetic Minority Oversampling Technique – Edited Nearest Neighbors):

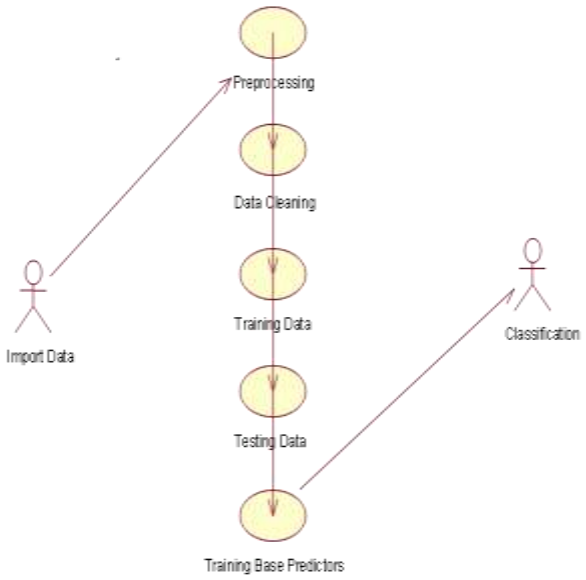
Balances the dataset by oversampling minority classes and removing noise.

Risk Scoring Algorithm:

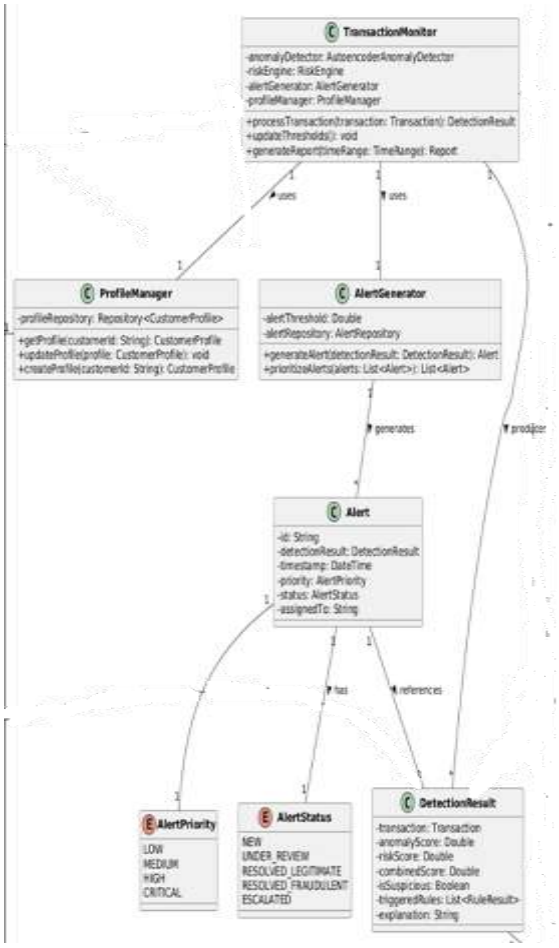
Weighted scoring system evaluates flagged transactions based on predefined parameters.

UML DESIGNS

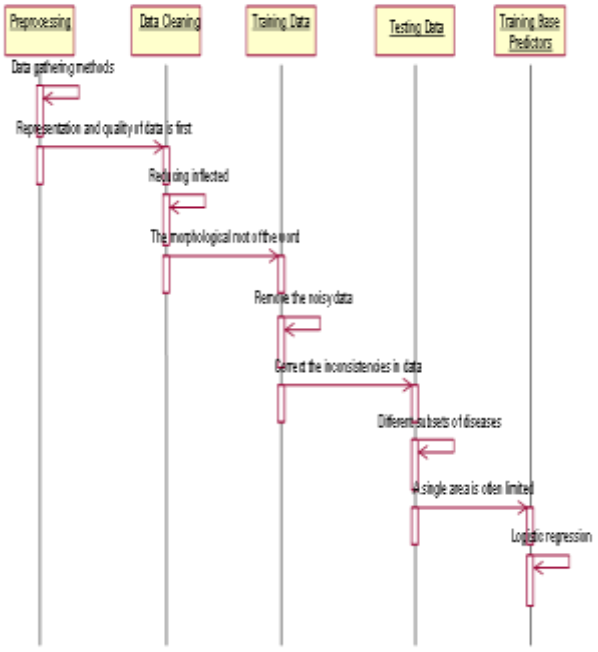
USE CASE DIAGRAM:



CLASS DIAGRAM:



SEQUENCE DIAGRAM:



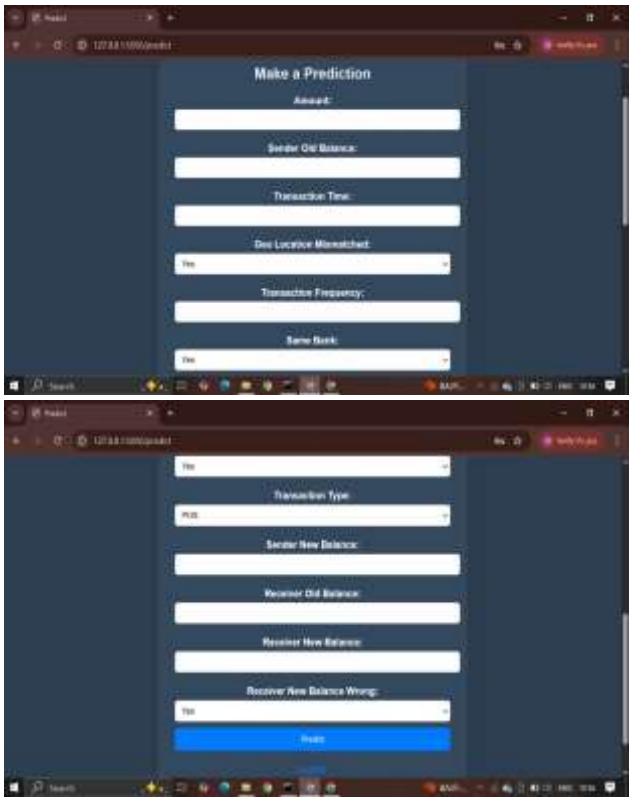
Screenshots:

Home Page

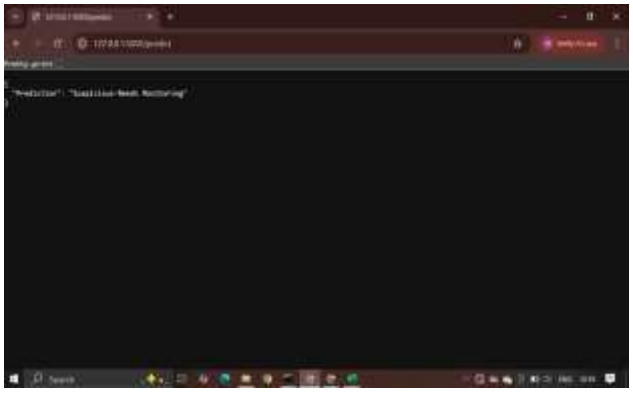
Registration Page

Login Page

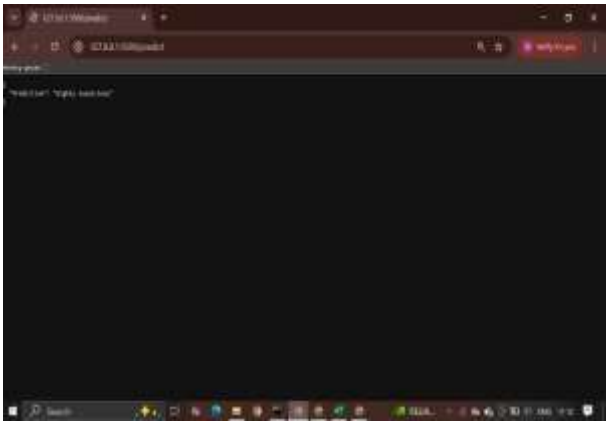
Prediction Page



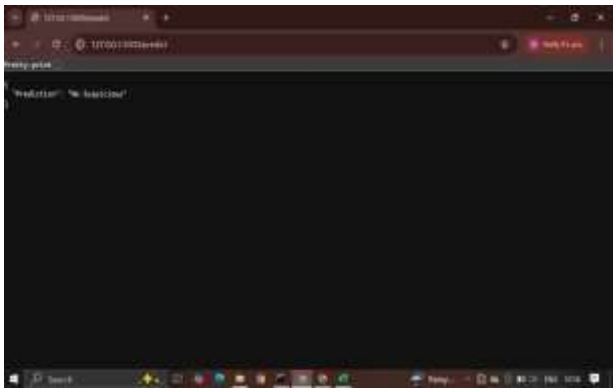
Result 1



Result 2



Result 3



CONCLUSION

The proposed model effectively addresses the limitations of traditional fraud detection systems by integrating autoencoder-based anomaly detection with a risk-based approach. This hybrid system ensures accurate, real-time detection of suspicious financial transactions while optimizing resources and minimizing manual efforts. By leveraging machine learning, the model not only adapts to evolving fraud tactics but also enhances scalability and efficiency, providing a robust solution to combat financial fraud. Future advancements, including explainable AI and blockchain integration, promise to further elevate the system's capabilities, ensuring its relevance in the dynamic financial landscape.

The fusion of autoencoder-based anomaly detection with a risk-based evaluation framework presents a groundbreaking approach to identifying suspicious financial transactions. Unlike traditional systems, the proposed model offers a dynamic, scalable, and intelligent solution capable of adapting to new fraud techniques. Its ability to operate in real time ensures immediate action can be taken, reducing financial and reputational risks for institutions.

REFERENCES

1. H. Zhang and W. K. Chan, ‘‘Apricot: A weight-adaptation approach to fixing deep learning models,’’ in Proc. 34th IEEE/ACM Int. Conf. Automated Softw. Eng. (ASE), Nov. 2019, pp. 376–387.
2. W.Fang,X.Li,P.Zhou, J.Yan, D.Jiang andT.Zhou,‘‘Deeplearninganti fraud model for Internet loan: Where we are going,’’ IEEE Access, vol. 9, pp. 9777–9784, 2021.
3. C. Dhasarathan, M. K. Hasan, S. Islam, S. Abdullah, U. A. Mokhtar, A. R. Javed, and S. Goundar, ‘‘COVID-19 health data analysis and personal data preserving: Ahomomorphic privacy enforcement approach,’’ Comput. Commun., vol. 199, pp. 87–97, Feb. 2023.
4. T. Verma and A. Misra, ‘‘Financial fraud detection in financial institutions usingtwo-layer-deeplearningandself-improvedhoneybadgeralgorithm,’’ J. Int. Finance Econ., vol. 23, no. 3, pp. 30–54, Oct. 2023.
5. Z. Chen, W. M. Soliman, A. Nazir, and M. Shorfuzzaman, ‘‘Varia tional autoencoders and Wasserstein generative adversarial networks for improving the anti-money laundering process,’’ IEEE Access, vol. 9, pp. 83762–83785, 2021.
6. J.Raval,P.Bhattacharya, N.K.Jadav, S.Tanwar, G.Sharma,P. N. Bokoro, M. Elmorsy, A. Tolba, and M. S. Raboaca, ‘‘RaKShA: A trusted explainable LSTM model to classify fraud patterns on credit card transactions,’’ Mathematics, vol. 11, no. 8, p. 1901, Apr. 2023.

7. F. Carcillo, Y.-A. Le Borgne, O. Caelen, Y. Kessaci, F. Oblé, and G. Bontempi, ‘‘Combining unsupervised and supervised learning in credit card fraud detection,’’ Inf. Sci., vol. 557, pp. 317–331, May 2021.
8. S. Saumya and J. P. Singh, ‘‘Spam review detection using LSTM autoencoder: An unsupervised approach,’’ Electron. Commerce Res., vol. 22, no. 1, pp. 113–133, Mar. 2022.