ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

www.ijasem.org

Vol 19, Issue 2, 2025

AUTOENCODER BASED SUSPICIOUS TRANSACTION DETECTION

Mr. G Mukesh

Assistant Professor

Computer Science and Engineering (Cyber Security)

Sphoorthy Engineering College,

Nadergul, Hyderabad,501510

g.mukesh@sphoorthyengg.ac.in

Moodu Priya

21N81A6201

Computer Science and Engineering (Cyber Security)

Sphoorthy Engineering College,

Nadergul, Hyderabad,501510

moodupriyap@gmail.com

Ramidi Sowjanya Reddy

21N81A6213

Computer Science and Engineering (Cyber Security)

Sphoorthy Engineering College,

Nadergul, Hyderabad,501510

sowjanyareddy761@gmail.com

Gonella Sai Karthik

21N81A6247

Computer Science and Engineering (Cyber Security)

Sphoorthy Engineering College,

Nadergul, Hyderabad,501510

g.saikarthik.5270@gmial.com

Mohammed Aqib Nazer

21N81A6243

Computer Science and Engineering (Cyber Security)

Sphoorthy Engineering College,

Nadergul, Hyderabad,501510

aqib.nazer223@gmail.com

ABSTRACT: The detection of suspicious financial transactions has been a critical focus in the financial industry for decades. Traditionally, financial institutions employed rule-based systems for identifying potentially fraudulent activities. These systems rely on predefined thresholds and patterns, such as large transactions or frequent deposits, to flag suspicious activities. While effective to some extent, traditional systems face significant limitations. They often generate a high rate of false positives, requiring manual intervention to review flagged transactions. Additionally, these systems struggle to adapt to

ISSN 2454-9940

www.ijasem.org

Vol 19, Issue 2, 2025

evolving fraud patterns, making them less effective in detecting sophisticated financial crimes. The growing complexity and volume of financial transactions in the digital era have heightened the need for advanced detection mechanisms. Traditional systems fail to address the dynamic nature of financial fraud, leading to inefficiencies in preventing financial losses. This creates a pressing need for a more adaptable, accurate, and scalable approach to detecting suspicious transactions. The lack of adaptability in traditional methods, combined with the significant financial and reputational risks posed by undetected fraud, underscores the necessity of a more robust detection framework. The goal is to enhance the ability to detect anomalous patterns in financial data with minimal false positives while maintaining efficiency and scalability. The proposed system introduces an innovative solution that leverages an autoencoder-based model combined with a risk-based assessment strategy. This approach aims to capture subtle anomalies in transaction data that deviate from normal patterns, enabling the identification of suspicious activities. The integration of a risk-based framework ensures that the model considers contextual factors, reducing false alarms and prioritizing high-risk transactions for further analysis. This system addresses the limitations of traditional methods, providing a sophisticated, adaptive, and reliable tool for combating financial fraud.

INTRODUCTION:

Suspicious financial transaction detection is a key area of concern in the financial industry, particularly in combating fraud and money laundering. The concept emerged in the 20th century with the introduction of the first fraud detection systems, relying on manual checks and basic software tools. In India, the Reserve Bank of India (RBI) and financial institutions have been actively working towards improving fraud detection techniques. According to the Financial Intelligence Unit of India (FIU-IND), there was a significant increase in suspicious transaction reports in recent years. In 2020 alone, over 3.5 million suspicious transaction reports were filed, reflecting the growing challenge of financial fraud in the country. Traditional systems failed to keep pace with the complexity of financial crimes, including cyber fraud, phishing, and money laundering. In response, India has adopted various technological advancements to improve detection, including machine learning algorithms. The need for a more automated, accurate, and scalable solution has grown in the face of digital banking, e-commerce, and the increasing number of financial transactions. As the economy becomes more digitized, the financial sector must adapt to these changes by implementing more sophisticated detection modeles.

LITERATURE REVIEW

Singh & Best [1] proposed a method using data visualization to detect suspicious activities for anti-money laundering (AML).

They demonstrated how data visualization techniques could effectively identify unusual transaction patterns indicative of money laundering, offering a powerful tool for monitoring financial systems and improving compliance procedures in the banking sector. Whisker & Lokanan [2] examined the risks posed by mobile money in the context of anti-money laundering and counter-terrorist financing. They discussed the challenges that mobile money presents, including its anonymous nature and the ease with which it can be exploited for illicit financial activities, and recommended strategies to mitigate these risks. Dobrowolski & Sułkowski [3] presented a sustainable model for implementing anti-money laundering measures aligned with the United Nations' development goals. They proposed an integrated approach that emphasizes sustainable financial systems and the role of AML in promoting global economic stability and development, while addressing the challenges of compliance in varying regulatory environments. Irwin et al. [4] analyzed typologies of money laundering and terrorism financing, focusing on the techniques and methods used to conceal illicit financial activities. They explored patterns and characteristics of these criminal activities, providing a comprehensive overview of how such practices can be detected and prevented through effective AML policies.

Uthayakumar et al. [5] introduced a framework based on swarm intelligence for classification rule induction, applied to bankruptcy prediction and credit risk analysis. They showed how the application of machine learning techniques like swarm intelligence could improve prediction accuracy in financial sectors, particularly for detecting risks associated with money laundering. KOFIU [6] presented the risk-based approach (RBA) standards for anti-money laundering (AML) and counter-terrorism financing (CFT) in financial investment businesses. This work highlighted how financial institutions can apply RBA to identify high-risk activities and tailor their compliance efforts accordingly, thereby optimizing the use of resources in combating financial crimes. Lee & Lee [7] shared their experiences and methodology for deploying and developing South Korea's anti-money laundering (AML) system. They discussed the integration of various regulatory frameworks and the challenges involved in creating a comprehensive system that effectively prevents money laundering while ensuring financial system stability. Pavlidis

[8] addressed the unintended consequences of implementing anti-money laundering standards, specifically those set by the Financial Action Task Force (FATF). The research examined how the FATF's strict compliance requirements may sometimes have negative effects, such as increasing financial exclusion, and suggested ways to mitigate these impacts. Celik [9] explored the impact of the FATF's recommendations on financial inclusion, drawing insights from mutual evaluations and national risk assessments. The study emphasized the balancing act between enforcing AML/CFT regulations and maintaining access to financial services, particularly for underserved communities.

Jayasekara [10] discussed the challenges of implementing an effective risk-based supervision approach for AML and countering the financing of terrorism under the 2013 FATF methodology. The paper focused on the difficulties financial institutions face when adopting these complex systems and provided recommendations to enhance supervisory effectiveness. Raghavan [11] explored the integration of anti-money laundering practices into corporate governance and finance functions, with a focus on compliance with the Bank Secrecy Act (BSA) and anti-money laundering (AML) regulations. The study showed how AML requirements are reshaping corporate governance, emphasizing the need for stricter internal controls to prevent financial crimes. Raghavan [12] examined how AML regulations are influencing corporate governance structures and finance functions in financial institutions. He discussed the integration of AML practices into organizational frameworks and highlighted the evolving nature of compliance requirements, particularly in response to emerging financial crimes. Labib et al. [13] surveyed machine learning approaches for anti-money laundering (AML) and counter-terrorism financing techniques. They reviewed various ML techniques, discussing how they can be applied to detect suspicious transactions, identify potential money laundering activities, and improve overall AML effectiveness.

Cherif et al. [14] conducted a systematic review of credit card fraud detection methods using disruptive technologies. They focused on how advancements in AI and machine learning have significantly enhanced the accuracy of fraud detection systems, especially in the context of financial

ISSN 2454-9940

www.ijasem.org

Vol 19, Issue 2, 2025

transactions, which are often targeted by money laundering activities. Senator et al. [15] proposed an AI-based system designed to identify potential money laundering activities from large cash transaction reports. Their research introduced the Financial Crimes Enforcement Network's (FinCEN) AI system for detecting suspicious patterns, emphasizing the role of AI in automating the detection of financial crimes. Wang & Yang [16] introduced a decision tree-based method for evaluating money laundering risks. Their model used historical transaction data to classify transactions based on their risk levels, enabling more effective detection and prevention of money laundering activities. Zhang & Zhou [17] explored the use of data mining techniques in financial applications, focusing on how these methods can uncover hidden patterns in transaction data that may indicate illicit activities such as money laundering. They discussed the potential of data mining to enhance financial monitoring systems and improve the identification of financial crimes.

3. METHODOLOGY

Methodologies and Techniques Used to Build Suspicious Financial Transaction Detection System Using Autoencoder and Random Forest Classifier

1. Web-Based Dataset Upload via Django Framework

The system features a **Django-based interface** that enables secure and user-friendly upload of CSV datasets containing financial transactions. The uploaded data typically includes transaction type, origin and destination balances, amount, and a binary label (`isFraud`) indicating whether a transaction is fraudulent.

Key Features:

- File validation and error handling
- Support for large CSV files
- Temporary storage and memory-efficient processing

---

2. Data Preprocessing & Transformation

Before feeding data into machine learning models, a structured pipeline ensures cleanliness and uniformity.

Techniques Used:

- **Null Value Handling**: Missing values are either dropped or imputed using statistical methods (mean, median).
- **Label Encoding**: Converts categorical transaction types (e.g., 'CASH_OUT', 'TRANSFER') into numerical format.
- **Feature Selection**: Key attributes such as `amount`, `oldbalanceOrg`, `newbalanceOrig`, `oldbalanceDest`, and `newbalanceDest` are selected.

ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

- **Standardization**: Features are scaled using `StandardScaler` to ensure zero mean and unit variance across dimensions.
- **Resampling**: Addressing class imbalance through techniques like upsampling minority classes to improve model generalization on rare fraud cases.

---

3. Dimensionality Reduction via Autoencoder (Unsupervised Learning)

An **Autoencoder neural network** is used to extract latent representations of transactional data, emphasizing the detection of anomalies.

Components:

- **Encoder**: Compresses high-dimensional input into a latent space.
- **Decoder**: Reconstructs input from the latent vector.
- **Loss Function**: Mean Squared Error (MSE) measures reconstruction loss to train the network.

Advantages:

- Captures non-linear relationships between features
- Learns compact feature representations that highlight anomalies (potential fraud)

---

4. Fraud Classification via Random Forest Classifier (Supervised Learning)

The encoded output from the Autoencoder is passed into a **Random Forest Classifier (RFC)** to predict fraudulent transactions.

Key Characteristics:

- Ensemble of decision trees
- Bootstrap aggregation to reduce variance
- Hyperparameter tuning (e.g., number of estimators, max depth)

Advantages:

- Robust to overfitting
- Handles non-linear and high-dimensional data effectively
- Provides feature importance scores

---

5. Model Evaluation Metrics

To validate model performance, multiple evaluation metrics are computed:

- **Accuracy**: Overall correctness of the model
- **Precision**: Ability to correctly identify only frauds (low false positives)
- **Recall**: Ability to capture all actual frauds (low false negatives)
- **F1-Score**: Harmonic mean of precision and recall
- **Confusion Matrix**: Breakdown of true/false positives and negatives

Visual feedback is provided using seaborn heatmaps, enhancing interpretability for analysts and stakeholders.

---

6. Real-Time Prediction & Results Display

The trained Autoencoder + RFC pipeline is deployed in a Django interface for real-time prediction on new transactional data.

Workflow:

ISSN 2454-9940

www.ijasem.org

Vol 19, Issue 2, 2025

- Upload new transaction data
- Apply same preprocessing pipeline
- Encode data using the trained Auto-encoder
- Predict using the Random Forest Classifier
- Display labeled output in a tabular format (fraud / not fraud)

---

7. Model Storage and Reusability

To enhance efficiency and deployment readiness, the models are persisted to disk using:

- `.h5` for Autoencoder via TensorFlow
- `.pkl` for RFC via joblib

On subsequent use, models are loaded without retraining unless explicitly updated, ensuring quick startup and prediction times.

Proposed System:

The **Suspicious Financial Transaction Detection Tool** is a machine learning-powered system designed to assist financial institutions in identifying fraudulent activities within large-scale transactional datasets. It utilizes a **hybrid architecture combining unsupervised anomaly detection and supervised classification**, enabling high-accuracy, low-false-positive fraud detection in real-time environments.

At the heart of the system lies a two-stage pipeline:

1. **Autoencoder-based Feature Extraction**: An unsupervised neural network model trained to learn compressed, latent-space representations of legitimate transaction patterns. The autoencoder flags anomalies based on reconstruction errors, effectively isolating potentially fraudulent behavior that deviates from the norm.
2. **Random Forest Classifier (RFC)**: A robust, ensemble-based supervised learning model that consumes the compressed features from the auto-encoder and classifies transactions as **fraudulent** or **non-fraudulent**. The use of decision-tree ensembles ensures high precision, interpretability, and resistance to overfitting.

This architecture is deployed via a **Django web interface**, allowing secure uploading of transaction data, initiating model training, and displaying prediction results through a user-friendly dashboard. A batch inference module enables real-time scoring of incoming transactions, supporting immediate risk-based decision-making.

The proposed system is designed with modularity and extensibility in mind. It supports easy integration with external data sources (e.g., user behavior logs, device data), and the underlying models can be continuously retrained with new labeled data to adapt to evolving fraud tactics. By combining **deep learning's ability to detect subtle patterns** with the **explainability of ensemble models**, the system offers a balanced solution to the modern challenges of financial fraud detection.

Advantages of proposed system:

Real-Time Anomaly Detection

- The autoencoder enables immediate flagging of suspicious transactions as they occur, minimizing financial exposure and enabling proactive response.

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

✓ Context-Aware Fraud Modeling

- Unlike static rule-based systems, the model adapts to behavioral patterns in the data, detecting emerging fraud trends without manual updates.

✓ Layered Detection Framework

- Combines unsupervised and supervised learning, increasing robustness and minimizing blind spots inherent in single-model systems.

✓ Improved Accuracy & Reduced False Positives

- The integration of autoencoders with Random Forests significantly reduces false alarms, directing analyst attention to genuinely suspicious activity.

✓ Explainability & Interpretability

- Random Forests offer feature importance scores and decision paths, allowing compliance teams to understand why a transaction was flagged.

✓ Scalable & Modular Design

- The system is designed to handle millions of transactions efficiently, with scalable preprocessing, vector transformation, and classification pipelines.

✓ Reusability & Offline Inference

- Trained models are saved and reused via `.h5` (Autoencoder) and `.pkl` (RFC), enabling offline predictions and rapid deployment in production systems.

✓ User-Friendly Interface

- The Django-based UI allows non-technical users to upload datasets, monitor model performance, and retrieve prediction results in a tabular, downloadable format.

✓ Customizable Thresholds

- Risk thresholds can be configured to match institutional risk tolerance, enabling high-sensitivity or high-specificity modes depending on the use case.

✓ Adaptability to Domain-Specific Data

- The modular pipeline can be extended to include additional features such as geolocation, time-of-day, or user metadata for improved context.

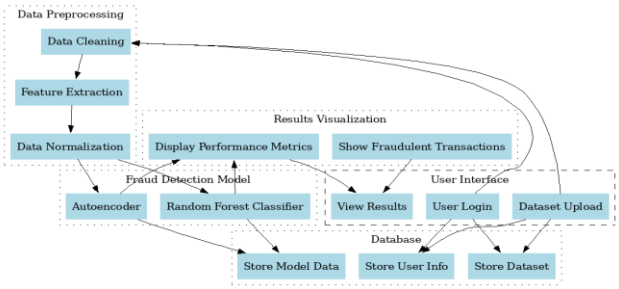


Fig.1: System architecture

1. IMPLEMENTATION

. Step 1: Dataset Upload via Web Interface

- A Django-based frontend allows users to upload `.csv` datasets.
- Uploaded files are stored temporarily and parsed for preprocessing.

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

Step 2: Data Preprocessing

- **Null Value Treatment**: Missing values are either removed or filled using statistical methods.
- **Label Encoding**: Categorical transaction types (e.g., 'TRANSFER', 'CASH_OUT') are converted to numerical format.
- **Feature Selection**: Relevant columns such as `amount`, `oldbalanceOrg`, `newbalanceOrig`, `oldbalanceDest`, `newbalanceDest` are retained.
- **Data Standardization**: StandardScaler is used to normalize numerical features.
- **Resampling**: If the dataset is imbalanced, upsampling or downsampling is used to address class skew.

Step 3: Model Building

A. Autoencoder (Feature Extraction)

- Architecture:
  - Input Layer: Same size as feature vector
  - Encoder Layers: Compress input to a lower-dimensional latent vector
  - Decoder Layers: Reconstruct input from latent vector
- Training:
  - Loss Function: Mean Squared Error (MSE)
  - Optimizer: Adam
  - Output: Encoder model used for feature transformation

B. Random Forest Classifier (Classification)

- Trained using the output of the encoder (latent features).
- Hyperparameters such as number of trees (`n_estimators`) and depth (`max_depth`) are tuned.
- Model is saved using `joblib`.

Step 4: Evaluation & Visualization

- Models are evaluated using:
  - **Accuracy**
  - **Precision**
  - **Recall**
  - **F1-score**
  - **Confusion Matrix** (plotted using seaborn)
- Evaluation metrics are displayed on the web interface along with model name and graphs.

Step 5: Real-Time Prediction

- A separate module enables uploading new unseen transactional data.
- Preprocessing and feature extraction are applied to the new data.
- Trained Random Forest model predicts labels: **Fraudulent** or **Not Fraudulent**.
- Predictions are rendered in a table format via the Django interface.

5. EXPERIMENTAL RESULTS

Fig.1 Home Page of the Financial Transaction Detection

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

www.ijasem.org

Vol 19, Issue 2, 2025

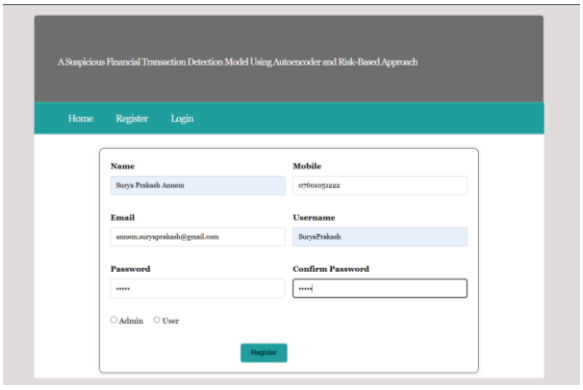


Fig 2: Common Registration for user and admin.

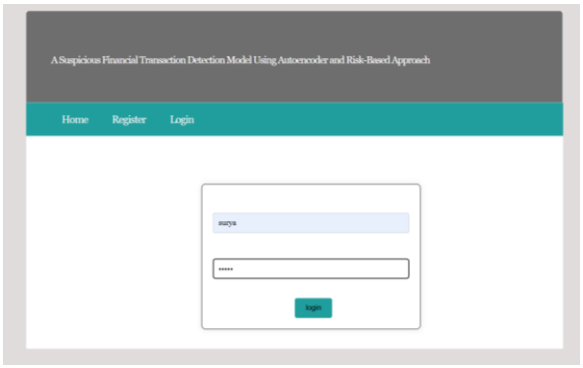
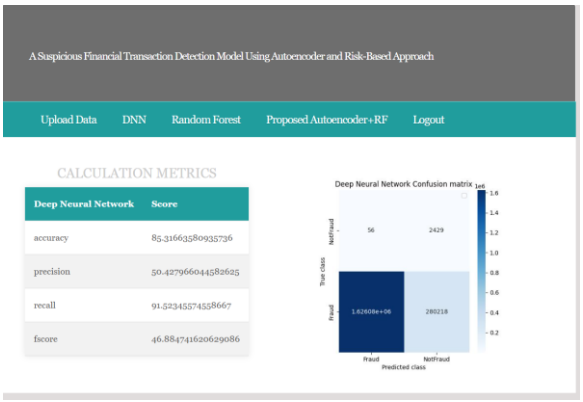
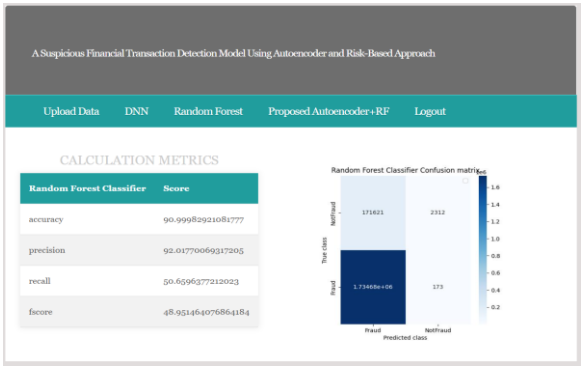


Fig 3: : User login for using Transaction detection

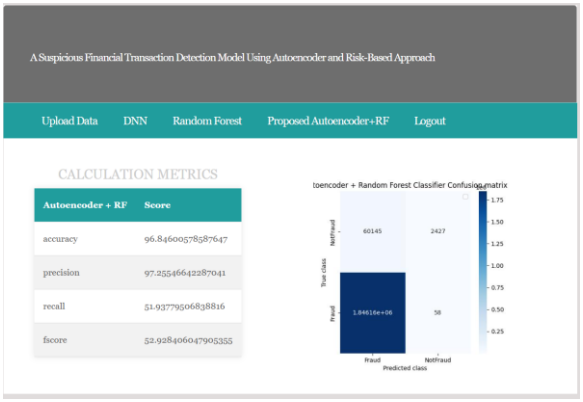
Fig.4: Sample Fraud Transaction Uploaded Dataset



Performance metrics of the Existing DNN model



Performance metrics of the Existing RFC model



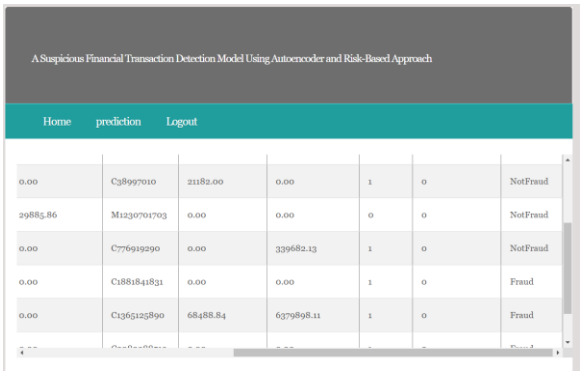
Performance metrics of the Proposed Auto Encoder + RFC model

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

www.ijasem.org

Vol 19, Issue 2, 2025



Proposed model prediction on user uploaded test data.

6. CONCLUSION

The Research combines the strengths of **Autoencoders** for feature extraction and dimensionality reduction with the **Random Forest Classifier (RFC)** for robust classification to detect fraudulent transactions in financial datasets. This hybrid approach leverages the unsupervised learning capabilities of Autoencoders to identify hidden patterns and anomalies in transaction data while utilizing RFC's high accuracy and interpretability for classification tasks. The system's architecture ensures scalability, efficiency, and accuracy in fraud detection, addressing challenges posed by imbalanced datasets and complex transactional behaviors. The project significantly contributes to reducing financial losses and enhancing trust in financial systems.

REFERENCES

[1] Singh, K., & Best, P. (2019). Anti-money laundering: Using data visualization to identify suspicious activity. International Journal of Accounting Information Systems, 34, 100418.

[2] Whisker, J., & Lokanan, M. E. (2019). Anti-money laundering and counter-terrorist financing threats posed by mobile money. Journal of Money Laundering Control, 22(1), 158-172.

[3] Dobrowolski, Z., & Sułkowski, Ł. (2019). Implementing a sustainable model for anti-money laundering in the United Nations development goals. Sustainability, 12(1), 244.

[4] Samantha Maitland Irwin, A., Raymond Choo, K. K., & Liu, L. (2011). An analysis of money laundering and terrorism financing typologies. Journal of Money Laundering Control, 15(1), 85-111.

[5] Uthayakumar, J., Vengattaraman, & Dhavachelvan, T. P. (2022). Swarm intelligence based classification rule induction (CRI) framework for qualitative and quantitative approach: An application of bankruptcy prediction and credit risk analysis. Journal of King Saud University - Computer and

ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

Information Sciences, 32(6), 647-657.

[6] KOFIU. (2017). Risk-Based Approach (RBA) Processing Standards for AML/CFT in Financial Investment Businesses. Institutional Operations Division, Financial Intelligence Unit.

[7] Lee, C.-J., & Lee, J.-C. (2013). Experiences and methodology of Korea's anti-money laundering system deployment and development. Knowledge Sharing Program: KSP Modularization.

[8] Pavlidis, G. (2023). The dark side of anti-money laundering: Mitigating the unintended consequences of FATF standards. Journal of Economic Criminology, 100040.

[9] Celik, K. (2021). Impact of the FATF Recommendations and their Implementation on Financial Inclusion: Insights from Mutual Evaluations and National Risk Assessments.

[10] Jayasekara, S. D. (2018). Challenges of implementing an effective risk-based supervision on anti-money laundering and countering the financing of terrorism under the 2013 FATF methodology. Journal of Money Laundering Control, 21(4), 601-615.

[11] Raghavan, K. R. (2006). Integrating anti-money laundering into the compliance structure: How the requirements for compliance with BSA/AML are changing the emphasis of corporate governance and finance functions. Bank Accounting & Finance, 19(6), 29-37.

[12] Raghavan, K. R. (2006). Integrating anti-money laundering into the compliance structure: How the requirements for compliance with BSA/AML are changing the emphasis of corporate governance and finance functions. Bank Accounting & Finance, 19(6), 29-37.

[13] Labib, N. M., Rizka, M. A., & Shokry, A. E. M. (2020). Survey of machine learning approaches of anti-money laundering techniques to counter terrorism finance. In Internet of Things—Applications and Future: Proceedings of ITAF 2019 (pp. 73-87). Springer.

[14] Cherif, A., Badhib, A., Ammar, H., Alshehri, S., Kalkatawi, M., Imine, A. (2023). Credit card fraud detection in the era of disruptive tech-

ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

nologies: A systematic review, Journal of King Saud University - Computer and Information Sciences, 35(1), 145-174.

[15] Senator, T. E., Goldberg, H. G., Wooton, J., Cottini, M. A., Khan, A. U., Klinger, C. D., Llamas, W. M., Marrone, M. P., & Wong, R. W. (1995). Financial crimes enforcement network AI system (FAIS) identifying potential money laundering from reports of large cash transactions. AI magazine, 16(4), 21-21.

[16] Wang, S.-N., & Yang, J.-G. (2007). A money laundering risk evaluation method based on decision tree. 2007 international conference on machine learning and cybernetics, January.

[17] Zhang, D., & Zhou, L. (2004). Discovering golden nuggets: Data mining in financial application. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 34(4), 513-522.