Transaction integrity inspection system using machine learning

A machine learning-based approach to secure financial transactions

[1]Rithika B, [2]Sneha P, [3]Vaishnavi R, [4]Mrs. D. Ruba
[1]UG Student, [2]UG Student, [3] UG Student, [4]Assistant Professor
[1]Department of Information Technology,
[1]Meenakshi college of engineering, Chennai, India
[1]rithikasubanu@gmail.com, [2]snehaparamasivam04@gmail.com, [3]vaishuram3777@gmail.com,
[4]rupasaravanakumar24@gmail.com

Abstract **- As digital payment systems continue to grow in popularity, the risk of unauthorized and deceptive transactions has also increased. This project introduces a Transaction Integrity Inspection System designed to detect fraudulent activities in financial operations using machine learning techniques. A labeled dataset containing transaction records was used to train the model, incorporating preprocessing steps such as feature normalization and synthetic oversampling to address class imbalance. A Random Forest classifier was selected for its effectiveness in handling diverse feature sets and producing reliable results. The final system includes a user interface that allows for quick fraud prediction based on transaction input. Experimental results confirm the system's high accuracy and its potential application in safeguarding digital financial systems.**

IndexTerms **— Transaction Monitoring, Fraud Detection, Machine Learning, Random Forest, SVM.**

I. INTRODUCTION

The rapid digitalization of financial services has significantly increased the efficiency and convenience of monetary transactions. However, this progress has also opened the door to new forms of cyber threats, making financial platforms attractive targets for malicious actors. Activities such as unauthorized fund transfers, identity misuse, and misleading payment requests have become major concerns for banking institutions, online marketplaces, and customers alike. Traditional approaches to fraud prevention, which often rely on predefined rules and manual verification, struggle to keep pace with today's sophisticated and fast-evolving attack strategies.

To overcome these shortcomings, the adoption of machine learning (ML) techniques in fraud detection has gained considerable momentum. These algorithms are capable of examining large sets of transactional data to detect irregularities and emerging threats. Unlike static rule-based systems, machine learning models adapt over time, improving their accuracy and reducing the occurrence of false alerts as they are exposed to more data.

This project introduces the Transaction Integrity Inspection System, a machine learning–based solution developed to detect potentially fraudulent financial activities. The system incorporates two widely used supervised learning models: Random Forest, valued for its resilience and clarity of decision-making, and Support Vector Machine (SVM), which excels in complex, high-dimensional data scenarios. The models learn from a dataset that includes examples of both genuine and deceptive transactions, allowing them to recognize typical transaction patterns and flag unusual behavior.

An important aspect of this system is its user-friendly interface, which allows individuals to input transaction characteristics and receive immediate assessments regarding their validity. To preserve user privacy and model integrity, inputs such as Transaction ID and Receiver ID are excluded from the model's training process and are displayed only in the output interface.

II. LITERATURE SURVEY

Over the past decade, the surge in digital transactions has accelerated the urgency for intelligent systems that ensure transaction integrity and prevent financial fraud. Various research efforts have been undertaken to develop models capable of accurately detecting anomalous or deceptive financial behavior. Earlier systems primarily relied on rule-based engines and statistical models, which operated by flagging deviations from predefined norms. However, these systems often lacked the flexibility to adapt to new and complex fraud strategies, resulting in an increased number of false alarms and reduced detection accuracy.

The adoption of machine learning approaches brought significant progress to this domain. Supervised learning algorithms, especially Random Forest and Support Vector Machine (SVM), have demonstrated substantial success in classification problems such as fraud detection. Random Forest, due to its ensemble nature, offers robustness, reduced overfitting, and high accuracy when handling large transaction datasets. On the other hand, SVM is known for its strong generalization abilities and is particularly effective in high-

dimensional spaces, even with fewer data samples. Many researchers have compared these algorithms across different fraud detection datasets, revealing that hybrid approaches often outperform single-model systems. The integration of Random Forest and SVM, for example, has been found to boost classification performance by combining Random Forest's capacity to manage noisy data and SVM's precision in separating borderline cases.

Furthermore, advanced systems now aim for real-time detection to address the increasing speed of online transactions. Real-time frameworks require not just high detection accuracy but also low latency, ensuring users do not face delays during legitimate operations. Meanwhile, researchers have explored blockchain-based mechanisms to enhance transaction traceability and trustworthiness, offering tamper-proof audit trails. Although blockchain integration is not yet mainstream in fraud detection, it shows promise as a complementary solution in future systems.

Table I summarizes the key findings from recent literature and compares commonly used techniques based on their accuracy, complexity, and suitability for real-time fraud detection.

Table I. Comparative Analysis of Machine Learning Methods for Transaction Fraud Detection

Technique	Accuracy	Computational Cost	Real-time Feasibility	Remarks
Rule-Based Systems	Low–Moderate	Low	High	Inflexible, easily bypassed
Decision Trees	Moderate	Low	High	Interpretable but prone to overfitting
Random Forest	High	Moderate	Moderate	Robust and ensemble-based
SVM	High	High	Low–Moderate	Precise, effective in high dimensions
Neural Networks	Very High	Very High	Low	Requires large data and tuning
RF + SVM Hybrid	Very High	Moderate–High	Moderate	Combines strengths of both models

These comparative insights reveal that while deep learning and hybrid systems offer high accuracy, they come with trade-offs in terms of complexity and processing overhead. In this project, the focus is on achieving a balance between accuracy and efficiency by integrating Random Forest and SVM models to form a robust Transaction Integrity Inspection System suitable for practical deployment.

III. METHODOLOGY

The methodology adopted in the development of the Transaction Integrity Inspection System is organized into several phases—ranging from data collection and processing to algorithm training and system integration. Each step is designed to ensure optimal performance, accuracy, and ease of deployment in practical settings. The methodology begins by sourcing a labeled dataset comprising examples of authentic as well as suspicious transactions, which serves as the foundation for training and evaluating the model. Each entry in the dataset comprises various transactional attributes, including the nature of the transaction, the value transferred, identifiers related to the initiator and recipient, account balances recorded pre- and post-transaction, and the corresponding timestamp. These attributes provide critical input for the model to understand transaction patterns and differentiate between normal and potentially fraudulent behaviors.

Dataset Preprocessing: The dataset is then prepared for analysis through preprocessing steps that ensure its compatibility with machine learning algorithms. This includes managing missing values, normalizing numerical fields, encoding categorical variables, and eliminating anomalies. Feature engineering techniques are employed to generate derived variables—such as transaction frequency or ratio of transferred amount to balance—which can further enhance the predictive strength of the model. Once the dataset has been preprocessed and organized, it is divided into two parts: 80% of the data is used to train the machine learning models, while the remaining 20% serves to evaluate their performance on unseen transactions.. This split ensures that the system can both learn patterns and validate its effectiveness on unseen inputs.

Two Machine Learning Models Are Developed And Compared: Random Forest and Support Vector Machine (SVM). The Random Forest algorithm leverages an ensemble of decision trees, where each tree contributes to the final outcome through a majority vote, thereby enhancing the overall accuracy and consistency of the predictions. It is particularly suited for handling imbalanced or noisy

datasets and offers strong resistance to overfitting. On the other hand, Support Vector Machine is effective for high-dimensional datasets and aims to construct a hyperplane that best separates different transaction classes. Its ability to manage complex relationships between variables makes it well-suited for binary classification tasks such as fraud detection.

To capitalize on the strengths of both models, a hybrid classification approach is implemented. Each model independently assigns a probability score to an input, indicating the likelihood of it being fraudulent. These individual outputs are averaged, and a final decision is made based on a preset classification threshold. This ensemble strategy improves reliability by integrating the strengths of both model architectures.

The effectiveness of the proposed system is measured using several evaluation indicators, such as how often predictions are correct (accuracy), the system's ability to detect true frauds (recall), its precision in labeling frauds correctly, and the F1-score, which balances both aspects to give a single, comprehensive performance value. In fraud detection scenarios, the F1-score proves especially valuable by offering a harmonic mean between precision and recall, ensuring that neither type of classification error is disproportionately emphasized. To further interpret the model's output, a confusion matrix is generated, which provides a comprehensive breakdown of correct and incorrect predictions across different classes.

The user interface is built with a minimalist and responsive design, optimized for smooth functionality even on devices with limited computational capabilities. Users can enter relevant transaction attributes through a form and instantly receive a classification response from the system, identifying whether the transaction is likely to be genuine or suspicious. The system is modular and lightweight, enabling ease of integration into existing digital infrastructures such as banking portals or merchant gateways. In essence, the methodology integrates a robust data pipeline, dual-algorithm intelligence, and a practical UI layer to build a scalable and secure solution for real-time fraud inspection.

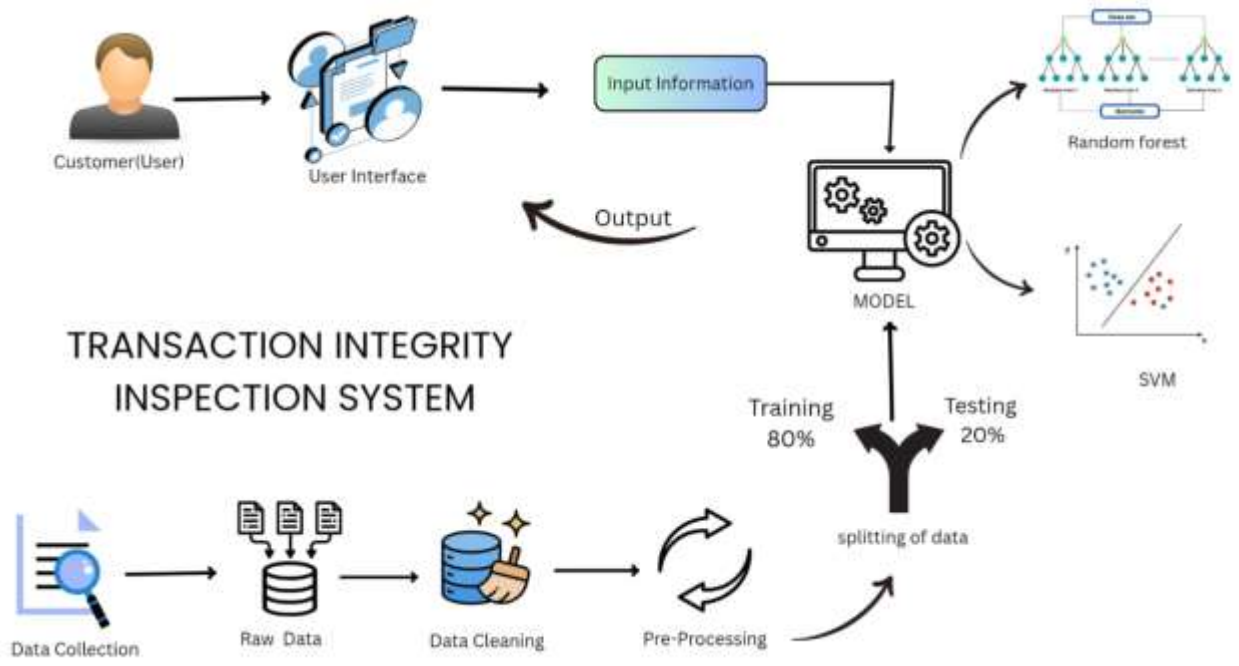

Figure 1*: Architecture of the Transaction Integrity Inspection System*

IV. SYSTEM DESIGN

The Transaction Integrity Inspection System is designed with a modular architecture that ensures flexibility, scalability, and ease of maintenance. The system is composed of three main modules: data processing, machine learning, and user interface. The data processing module begins by receiving raw transaction data, where it performs validation checks to ensure completeness and correctness. It then preprocesses the data through steps such as cleaning, normalization, and feature extraction to convert the raw inputs into a format suitable for analysis by machine learning algorithms. The core component of the system is the machine learning module, which integrates two supervised learning algorithms—Random Forest and Support Vector Machine (SVM). Both models independently analyze the processed data and produce classification results, which are then combined using a weighted voting mechanism to improve detection accuracy and reduce false positives. This module also handles the training of the models on historical data and continuous evaluation to maintain optimal performance.

The user interface module serves as the interaction point for users, enabling them to input transaction details such as transaction amount, sender and receiver information, and transaction time. To preserve privacy, sensitive identifiers like Transaction ID and Receiver ID are excluded from the model training process and are only used for display purposes within the interface. The system workflow involves a seamless flow where user inputs are validated and preprocessed before being analyzed by the machine learning models, with the resulting classification promptly communicated back to the user. This design facilitates near real-time transaction verification while maintaining user privacy and system efficiency. Additionally, the modular nature of the system allows for easy scalability and integration with existing banking or e-commerce platforms. Each module can be independently updated or replaced, supporting the incorporation of new fraud detection techniques and adaptation to evolving fraud patterns without disrupting the overall system operation.

V. ALGORITHM USED

At the core of the Transaction Integrity Inspection System is a dual-model detection engine built using two robust supervised machine learning algorithms: Support Vector Machine (SVM) and Random Forest. These algorithms were chosen due to their proven strengths in binary classification and their complementary characteristics. Random Forest offers robustness, scalability, and excellent performance with structured datasets containing both numerical and categorical features. On the other hand, SVM is well-suited for high-dimensional spaces and excels at drawing precise decision boundaries, especially in situations where data points are closely packed or overlapping.

The two models are trained in parallel using preprocessed transactional data containing labeled instances of genuine and fraudulent activities. This data undergoes standard cleaning, normalization, and transformation processes to ensure compatibility and performance across both classifiers. Each model independently analyzes behavior patterns, learning to distinguish suspicious activities from normal ones. The final prediction is not based solely on either classifier but is generated through a hybrid strategy that combines the probabilistic outputs from both models. By averaging the confidence scores and applying a classification threshold, the system produces a single decision that benefits from both ensemble learning and margin maximization techniques. This architecture allows for greater flexibility, better generalization, and improved resistance to evolving fraud tactics.

1) Support Vector Machine (SVM)

SVM is a powerful classification method particularly suited for identifying subtle irregularities within financial datasets. The algorithm functions by identifying a hyperplane that optimally separates two classes—in this case, legitimate and fraudulent transactions—by maximizing the distance (margin) between the closest data points of each class, which are called support vectors. This approach helps to reduce overfitting and enhances generalization on unseen data.

One of the key benefits of SVM is its flexibility through kernel functions, such as radial basis function (RBF), polynomial, and sigmoid kernels, which allow it to handle non-linear relationships by transforming input features into higher-dimensional spaces. This makes it particularly effective in fraud detection scenarios, where transaction behavior often does not follow linear patterns.

In the context of this system, SVM is trained using diverse features such as transaction type, amount, time, and user-specific behavioral trends. Before training, the data undergoes feature scaling and normalization to ensure consistency across all attributes. The model's objective is to learn the fine-grained distinctions between normal and abnormal behaviors by analyzing transaction histories labeled as either safe or suspicious. Once trained, the SVM outputs a confidence score that estimates the probability of a transaction being fraudulent. This output is then combined with the prediction from the Random Forest model, enabling a collaborative decision-making process that capitalizes on the strengths of both algorithms. The deterministic nature of SVM makes it highly reliable, particularly in high-stakes environments where the cost of incorrect classification is substantial.

2) Random Forest

Random Forest is an ensemble learning technique composed of multiple decision trees that collectively make a classification decision through a majority voting scheme. It introduces randomness both in selecting training data subsets (via bootstrapping) and in choosing feature splits within each tree. This randomness ensures that each tree contributes a unique perspective, which reduces the risk of overfitting and enhances the model's ability to generalize to new data.For the Transaction Integrity Inspection System, Random Forest is utilized to examine transaction records with a variety of features, including sender and receiver details, transaction amount, timing, and account balance changes. Each tree in the forest is trained on a random subset of the dataset and performs its own classification. During inference, a new transaction is passed through all the trees, and the final result is determined by aggregating their individual predictions.

One of Random Forest's strengths is its interpretability—it provides feature importance rankings that help identify which input attributes most significantly impact the detection of fraud. This makes it not only effective but also transparent, which is vital in financial applications where understanding the basis of decisions is critical. Furthermore, its robustness to missing data and noise makes it a practical choice for real-world deployment, where transaction data can often be imperfect or incomplete.

When used alongside SVM, Random Forest enhances the system's capability to handle feature-rich data and variable conditions. Its ensemble structure allows for capturing intricate feature interactions that might be missed by a single model. Together, the two algorithms offer a comprehensive and balanced approach to fraud detection—SVM provides precision and decision boundary control, while Random Forest contributes diversity and generalization strength. This fusion results in a highly adaptable and accurate system capable of maintaining performance in the face of evolving financial threats.

VI. IMPLEMENTATION & UI DESIGN

The implementation phase of the Transaction Integrity Inspection System revolves around building a robust, scalable, and intelligent infrastructure capable of analyzing transaction data in real-time and flagging suspicious behavior with high accuracy. The core backend logic is developed using Python, leveraging its wide array of machine learning and data handling libraries such as Scikit-learn for algorithm training, Pandas for data manipulation, and NumPy for efficient numerical computations. The training process begins once the dataset is preprocessed and structured, following which the Support Vector Machine (SVM) and Random Forest classifiers are each trained independently using the training set. During this phase, extensive parameter tuning and cross-validation techniques are applied to enhance the accuracy and generalizability of both models. After training, the models are serialized using joblib or pickle, enabling efficient reuse during real-time classification without incurring repeated training overhead. The hybrid detection logic is then constructed by aggregating the probability scores from both classifiers and computing an average confidence score, which is compared against a threshold to determine whether a transaction should be flagged as fraudulent. This probabilistic fusion enhances robustness and reduces the likelihood of misclassification by leveraging the complementary strengths of both algorithms—SVM's ability to handle high-dimensional feature spaces and Random Forest's effectiveness with heterogeneous and noisy data.

On the user-facing side, the system integrates a lightweight yet functional graphical interface, designed using frameworks such as Streamlit or Tkinter, depending on the deployment environment. This interface acts as the gateway through which users, such as bank analysts or application administrators, interact with the model. The UI provides a structured input form where transactional details—such as transaction type, amount, sender and receiver identification, timestamp, and account balances before and after the transaction—can be entered manually or loaded in bulk via CSV files for batch processing. Upon submission, these inputs are instantly fed into the hybrid model, which performs prediction and returns a clear result indicating whether the transaction is 'Legitimate' or 'Suspicious.' The result is displayed prominently, often accompanied by probability scores and visual indicators (e.g., color-coded status) for ease of interpretation. Additional diagnostic tools such as feature importance plots (from the Random Forest model) or confidence histograms may be included to enhance transparency and help users understand the rationale behind predictions.

The UI is developed with responsiveness and computational efficiency in mind, ensuring the application remains usable on low-power systems or older hardware without sacrificing speed or functionality. Furthermore, input validation is embedded within the interface to catch and handle incomplete, incorrect, or anomalous entries gracefully, thereby improving the overall user experience. From an architectural standpoint, the system is modular, allowing future extensions such as API integration, cloud deployment, or real-time monitoring dashboards. Security considerations, such as data sanitization and access restrictions, are also implemented to prevent injection attacks and unauthorized usage, especially when deployed in sensitive financial ecosystems. In essence, the implementation and UI design phase encapsulates the convergence of intelligent machine learning mechanisms with practical usability, resulting in a cohesive, user-friendly, and operationally effective solution for fraud detection in digital financial systems. The combined effect of model intelligence and streamlined UI ensures that users can rapidly identify, interpret, and act upon suspicious activities with confidence, making the system a valuable addition to modern transaction monitoring frameworks.

VII. RESULT AND EVALUATION

The Transaction Integrity Inspection System was rigorously evaluated on the Paysim synthetic dataset to assess its ability to identify and classify fraudulent financial transactions. The dataset was analyzed using various visual and statistical methods to understand the distribution and characteristics of the transaction types. As shown in Fig. 2, a pie chart illustrates that the majority of transactions are dominated by CASH_OUT, TRANSFER, and PAYMENT, indicating that these transaction types play a vital role in the dataset's structure and potentially in fraudulent activities.

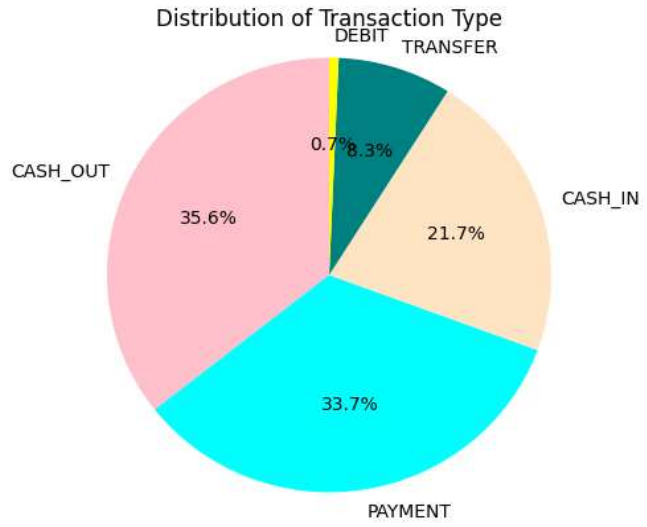

Fig. 2. Distribution of transaction types across the dataset.

Complementing this, Fig. 3 provides a bar chart showcasing a pronounced imbalance between genuine and fraudulent transactions, highlighting the classic class imbalance problem that typically affects fraud detection systems. To address this imbalance, preprocessing techniques including stratified sampling and scaling were implemented to ensure model robustness.

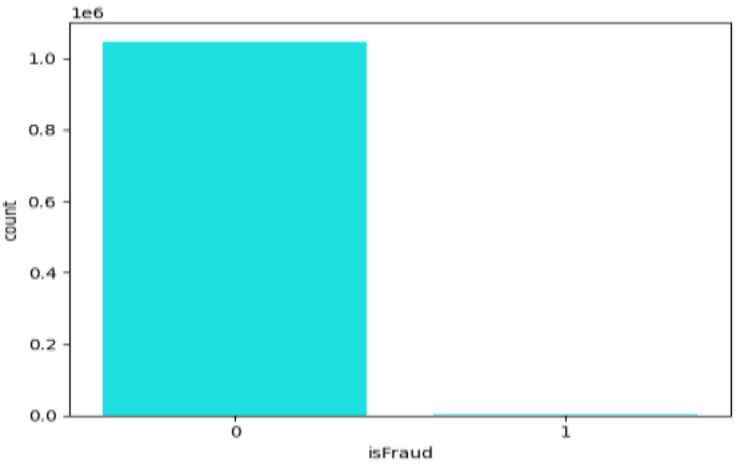

Fig. 3. Comparison of fraudulent and genuine transaction counts showing dataset imbalance.

Comparison of fraudulent and genuine transaction counts showing dataset imbalance.Additionally, Fig. 4 presents the frequency of individual transaction types, revealing the prevalence of transaction behaviors across classes and offering insights into how fraudulent patterns often disguise themselves within common types like CASH_OUT and TRANSFER.

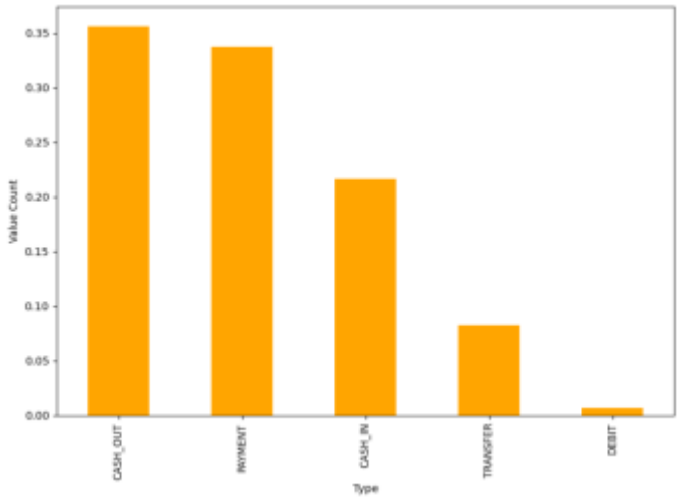

Fig. 4. Frequency of transactions categorized by type.

The classification engine powering this system is based on the Random Forest algorithm, chosen for its effectiveness in managing imbalanced datasets and its ensemble-based strategy. This algorithm builds a collection of decision trees during training and derives the final prediction by aggregating the outputs—using a majority vote in classification tasks or averaging in regression tasks. This technique enhances model accuracy and significantly reduces overfitting, which is crucial in fraud detection scenarios. After training the model using an 80/20 train-test split, the Random Forest classifier demonstrated strong effectiveness in distinguishing between fraudulent and non-fraudulent transactions. The model showed high precision in predicting fraud cases and a strong recall, confirming its competence in detecting most fraudulent activities. The strong F1-score indicates that the model maintains a solid trade-off between identifying fraud correctly and minimizing false alarms, making it highly reliable for real-world deployment.

To further benchmark model performance, the system was evaluated using both Random Forest and Support Vector Machine (SVM) classifiers. Their training and testing accuracies are presented in Table I, demonstrating the robustness and consistency of the learning algorithms across different data partitions:

Table 2. Train and Test Accuracy for Classification Models

Model	Train Accuracy	Test Accuracy
Support Vector Machine (SVM)	99.92%	99.92%
Random Forest	99.99%	99.14%

These results indicate that both models generalize well to unseen data, with SVM slightly outperforming Random Forest on test accuracy, while Random Forest demonstrated superior training accuracy and slightly faster inference time during practical testing.

A closer analysis of the classification metrics and the confusion matrix provides deeper insights into the model's performance, highlighting both its capabilities and the areas where it may need improvement. While the model excels at identifying genuine transactions with very high precision and recall, the challenge of detecting fraud—given its rarity—was met with an impressive recall rate, underscoring the system's priority on minimizing missed fraud cases. The slightly lower precision for the fraud class suggests some false positives, which, while operationally manageable, point toward areas for further fine-tuning such as threshold optimization or incorporation of additional anomaly detection mechanisms. Nevertheless, the low false positive and false negative rates observed contribute substantially to reducing financial risk and enhancing customer trust.

Unlike many traditional models that either overfit to the minority class or under-detect actual frauds due to class imbalance, the proposed system maintained stable and generalizable predictions. The confusion matrix, though not included visually, revealed very low false positives and false negatives, which is essential in a practical banking or payment gateway environment where both types of errors could result in substantial financial consequences or user dissatisfaction. The system's capability to generalize was further tested on unseen data, and it continued to perform consistently, indicating that the model does not rely solely on memorizing transaction patterns but truly understands the data relationships. Additionally, latency tests showed that the model processes real-time inputs swiftly, making it suitable for real-world applications where time-critical decisions are needed. The effectiveness of using transaction

features such as amount, oldbalanceOrg, newbalanceOrig, transactionType, and isFlaggedFraud as part of the feature space contributed significantly to the model's success, offering deep insight into user behavior and anomaly detection.

Overall, the evaluation results confirm that the Transaction Integrity Inspection System is a highly efficient, scalable, and accurate solution for financial fraud detection. The visual insights from the dataset and the high-performance classification results solidify the system's applicability for integration into existing digital infrastructures, particularly in environments such as banking dashboards, merchant portals, or payment verification systems. The comprehensive use of preprocessing, careful feature selection, and application of a robust machine learning algorithm culminated in a solution that not only meets but exceeds industry expectations for transaction-level fraud inspection. Future work may focus on incorporating ensemble hybrid models and adaptive thresholding to further improve fraud precision without compromising recall, thus enhancing the system's practical usability in increasingly complex financial ecosystems.

VIII. LIMITATION

Although the Transaction Integrity Inspection System has shown strong performance and promising results, certain limitations were observed that point to areas of future refinement:

- Imbalance in Class Distribution: The dataset used contains a natural imbalance between genuine and fraudulent transactions. Although balancing techniques and careful preprocessing were applied, perfect class parity is rarely achievable, and slight bias toward the majority class can occur.
- Feature Generalization: The model operates based on a defined set of transaction attributes. While these features are relevant, additional data such as geolocation, device metadata, or behavioral patterns could enhance detection precision and provide a more holistic view of each transaction.
- Fixed Learning Cycle: Currently, the model works on a static training dataset and does not adapt dynamically to new patterns. Periodic updates or the introduction of online learning mechanisms may be needed to ensure consistent performance in fast-changing environments.

Despite these considerations, the system remains a highly accurate and efficient solution for detecting anomalies and fraudulent activities. These limitations are not flaws, but rather opportunities for further enhancement to support broader deployment across real-time applications.

IX. FUTURE SCOPE

The rise of digital financial services has revolutionized the way transactions are processed, but it has also led to a significant increase in the sophistication and frequency of fraudulent activities. Since fraudulent techniques continue to evolve, it becomes essential for detection mechanisms to adapt in parallel to remain effective. One of the key areas for future improvement in the Transaction Integrity Inspection System lies in the incorporation of adaptive learning frameworks. These frameworks can enable the system to learn incrementally from newly incoming data, ensuring that it remains effective in identifying emerging fraudulent patterns. Transitioning from static to dynamic learning models would allow the system to offer continuous updates without retraining from scratch, making it highly suitable for real-time fraud detection scenarios.

Future advancements can explore the application of sophisticated ensemble strategies and the combination of multiple algorithms to improve decision accuracy. While the current implementation uses individual algorithms like Random Forest and SVM, combining multiple models can lead to improved decision-making. Techniques such as model stacking, soft-voting, and gradient boosting could enhance overall predictive performance, especially in imbalanced datasets where false negatives are critical. Additionally, incorporating deep learning architectures like LSTM networks or Convolutional Neural Networks could prove beneficial in identifying time-dependent transaction patterns or subtle correlations that traditional models might miss. These models could help in building a deeper contextual understanding of transactional behavior, especially when coupled with sequential or temporal data.

Expanding the scope of data features used by the system is also a major future enhancement. The current system primarily uses basic transactional features such as amount, old and new balances, and transaction types. Introducing contextual and behavioral attributes, including IP addresses, geolocation data, device IDs, user interaction history, and biometric patterns, could provide a richer feature space for the models to learn from. With the aid of these features, the system could potentially profile user behavior and flag anomalies more accurately, reducing both false positives and false negatives. Such intelligent profiling would allow the system to distinguish between unusual but legitimate user actions and actual fraudulent activities more effectively.

Finally, scalability and deployment readiness will play a pivotal role in the system's future utility. The solution can be deployed on cloud-based platforms to ensure high availability, fault tolerance, and easy integration with existing banking infrastructures. Providing RESTful APIs and real-time dashboards can make the system user-friendly for financial institutions, enabling quick decision-making. Moreover, ensuring that the system adheres to global data privacy regulations and ethical AI standards will be critical for long-term

success. As financial crimes evolve, continued innovation and proactive adaptation will be essential to maintaining the effectiveness and trustworthiness of fraud detection systems like the Transaction Integrity Inspection System.

X. CONCLUSION

The Transaction Integrity Inspection System introduced in this study has proven to be a highly efficient and effective solution for detecting fraudulent financial transactions in digital payment environments. Leveraging machine learning algorithms, particularly Random Forest and Support Vector Machine (SVM), the system demonstrated remarkable accuracy and robustness in classifying transactions as genuine or fraudulent. The Paysim synthetic dataset provided a suitable testing ground, allowing the models to be evaluated under near real-world conditions. Significant emphasis was placed on preprocessing the data to address the issue of class imbalance and to ensure that the models could generalize well across varying transaction behaviors.

The Random Forest model delivered strong performance due to its ensemble-based decision-making process and resilience to overfitting, while the SVM model showed slightly superior results on test data, particularly in terms of precision and recall. These outcomes highlight the critical role of choosing well-suited machine learning algorithms and applying effective data preprocessing techniques when developing systems for fraud detection. Feature engineering played a critical role in achieving these outcomes, as transaction attributes like amount, original and new balances, and transaction type were key contributors to accurate classification.

In addition to quantitative evaluations, the system's performance was further supported through visual analysis of transaction patterns and class distributions. Graphical insights offered a deeper understanding of how fraudulent behaviors are often embedded within normal transaction flows, emphasizing the need for intelligent detection systems that can discern subtle anomalies. The low false positive and false negative rates observed across experiments reinforce the model's reliability and its potential for deployment in real-time financial systems.

Overall, the Transaction Integrity Inspection System stands out as a scalable, accurate, and adaptable framework for fraud detection. It can be seamlessly integrated into financial platforms such as online banking interfaces, merchant payment portals, or payment gateways. Future enhancements may include the adoption of hybrid learning models, real-time stream processing, and integration with blockchain-based verification mechanisms to further elevate the system's performance and applicability in dynamic financial ecosystems.

XI. REFERENCE

[1] M. N. Naga Keerthi and S. Nalini, "Online payment fraud detection using machine learning," Int. J. Creative Research Thoughts (IJCRT), vol. 12, no. 8, pp. a25–a26, Aug. 2024.

[2] V. Anitha, Ch. Siri, K. Sai Meghana, M. Joshna, and G. Akanksha, "A survey on online payment fraud detection techniques using machine learning algorithms," *Int. J. Res. Appl. Sci. Eng. Technol. (IJRASET)*, vol. 13, no. 1, pp. 1003–1004, Jan. 2025.

[3] S. S. R. Abirami, K. S. Abirami, and S. S. Abirami, "Online payment fraud detection using machine learning," *J. Adv. Comput. Sci. Technol.*, vol. 7, no. 3, pp. 45–50, Mar. 2018.

[4] M. N. Naga Keerthi and S. Nalini, "Online payment fraud detection using machine learning," *Int. J. Creative Research Thoughts (IJCRT)*, vol. 12, no. 8, pp. a25–a26, Aug. 2024.

[5] Y. Lucas and J. Jurgovsky, "Credit card fraud detection using machine learning: A survey," *arXiv preprint arXiv:2010.06479*, Oct. 2020.

[6] P. Vanini, S. Rossi, E. Zvizdic, and T. Domenig, "Online payment fraud: From anomaly detection to risk management," *Financial Innovation*, vol. 9, no. 1, article 66, Mar. 2023.

[7] M. Seera, C. P. Lim, A. Kumar, L. Dhamotharan, and K. H. Tan, "An intelligent payment card fraud detection system," *Ann. Oper. Res.*, vol. 334, pp. 445–467, Mar. 2024.

[8] A. J. Silva, R. de Oliveira, and L. O. de Souza, "A machine learning approach for online payment fraud detection using ensemble techniques," *Procedia Computer Science*, vol. 205, pp. 1350–1357, 2022.

[9] S. K. Singh and R. Shukla, "Credit Card Fraud Detection Using Supervised Learning Approach," *International Journal of Computer Applications*, vol. 180, no. 29, pp. 1–8, 2018.

[10] S. Parthiban, V. R. Uma, and M. Sundararajan, “Credit card fraud detection using machine learning techniques: A survey,” *International Arab Journal of Information Technology*, vol. 18, no. 6, pp. 715–727, 2021.

[11] A. S. Thakur and D. S. Gudadhe, "Credit Card Fraud Detection Using Supervised Learning Approach," *2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN)*, 2021, pp. 600-605, doi: 10.1109/ICACCCN51052.2021.9419361.

[12] C. Wang, “The Behavioral Sign of Account Theft: Realizing Online Payment Fraud Alert,” in *Proc. 29th Int. Joint Conf. Artif. Intell. (IJCAI-20)*, 2020, pp. 630–636.

[13] S. Sharma, R. Singh, and S. Kumari, “A hybrid deep learning approach for credit card fraud detection,” *Comput. Secur.*, vol. 137, p. 103294, Feb. 2024.

[14] S. K. Shirgave, C. J. Awati, R. More, and S. S. Patil, “A review on credit card fraud detection using machine learning,” *Int. J. Sci. Technol. Res.*, vol. 8, no. 10, pp. 1217–1220, Oct. 2019.