Journal of Economics, Finance and Accounting Studies
ISSN: 2709-0809
DOI: 10.32996/jefas
Journal Homepage: www.al-kindipublisher.com/index.php/jefas

JEFAS
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

Enhancing Fraud Detection Systems in the USA: A Machine Learning Approach to Identifying Anomalous Transactions

Md Shafiqur Rahman[1], Proshanta Kumar Bhowmik[2], Balayet Hossain[3], Nikhil Rao Tannier[4], Mohammad Hamid Hasan Amjad[5], Anchala Chouksey[6], and Miraz Hossain[7]

[1]MBA in Management Information System, International American University
[2]Department of Business Analytics, Trine University, Angola, IN, USA
[3]MBA in Business Administration and Management, International American University
[4]Master's in Artificial Intelligence, University of North Texas
[5]Master of Science in Engineering Management, College of Graduate and Professional Studies, Trine University.
[6]Masters of Science in Financial Mathematics, University of North Texas, Denton, Texas
[7]Master of Business Administration, Westcliff University
Corresponding Author: Md Shafiqur Rahman, **E-mail**: rahman019@gannon.edu

| ABSTRACT

The landscape of financial fraud in the United States is more advanced today, with fraudsters adopting sophisticated methods that elude traditional detection systems. As digital payments gain popularity, the number of potential fraud cases and their sophistication also increased, causing heavy financial losses to institutions and consumers alike. The primary objective of this research was to design and implement machine learning models that can significantly improve fraud detection systems in their precision. This study was centered specifically on fraud detection in the US financial system, researching artificial intelligence approaches that can be applied to support anomaly detection and risk analysis processes. The dataset employed in the analysis is a high-level transaction dataset that includes a spectrum of financial transaction details. Each transaction entry included primary details such as timestamp, transaction value, and sender-receiver information. The timestamp enabled each transaction to be sorted in chronological order, making it possible to carry out time-series analysis of patterns such as maximum transaction time or seasonality in spending behavior. Three models were predominantly employed: Random Forest Classifier, Logistic Regression, and Support Vector Classifier. The performance of models was measured using a set of metrics that included accuracy, precision, recall, F1-score, and ROC-AUC. The Random Forest model was better in terms of higher accuracy, thanks to its ability to handle non-linear relationships via ensemble learning. The integration of machine learning in fraud detection enhances the capabilities of payment providers and financial institutions tremendously. With sophisticated algorithms, financial institutions can process large volumes of transactional data in real time, enabling them to detect anomalous patterns that speedily indicate fraud. The findings of this study reinforce the effectiveness of machine learning models in identifying anomalous transactions, verifying that advanced approaches such as Random Forest and Support Vector Machines significantly enhance fraud detection compared to legacy approaches. One key to such effectiveness is that feature selection is crucial; carefully chosen features that included user behavior and transactional context played a key role in increasing detection rates and eliminating false positives.

| KEYWORDS

Fraud Detection, Machine Learning, Anomaly Detection, Financial Transactions, USA, AI Security

| ARTICLE INFORMATION

ACCEPTED: 02 October 2023 **PUBLISHED:** 12 October 2023 **DOI:** 10.32996/jefas.2023.5.5.15

I. Introduction

Background and Context

According to Ali et al. (2022), the rapid growth of digital payments in the US has transformed the financial system, providing unprecedented convenience and efficiency to companies and consumers alike. The digital revolution also created a fertile ground for a wave of financial fraud, with criminals exploiting vulnerabilities in electronic payment systems, online banking, and e-commerce platforms. The fraud surge is alarming; cases of account takeovers, card fraud, and identity theft are rampant, causing billions of dollars in losses each year. Traditional rule-based fraud detection systems that apply a static set of rules and historical behavior to flag suspicious behavior fail in such a configuration. Such systems fail to adapt to new methods employed by fraudsters, providing poor protection and struggling to adapt to new threats promptly. There is a strong need for more advanced methods that can handle the constantly evolving nature of financial fraud (Al Mukaddim et al., 2023).

The financial system of the United States has been revolutionized in response to the large-scale use of digital payments, online banking, and electronic payment systems. The digital revolution has introduced convenience and efficiency to financial systems but has also exposed financial institutions and consumers to a higher risk of fraud. Cybercriminals employ more sophisticated means to exploit vulnerable areas in financial systems, employing stolen identities, synthetic identities, and sophisticated evasion strategies to execute frauds (Obeng et al., 2021). The Federal Trade Commission (FTC) and other institutions of regulation report alarming increases in fraud transactions, with billions of dollars of losses every year resulting from cyber fraud. Manually created rule-based fraud detection systems that employ threshold-based alarms cannot handle the dynamic nature of threats. As fraudsters continue to evolve their tactics to evade detection, there is a need for more advanced and dynamic fraud detection systems (Islam et al., 2023)

Problem Description

Chen et al. (2018) reported that fraudsters are not only more sophisticated in their method of approach but also highly dynamic, continually evolving their methods to stay ahead of traditional detection mechanisms. The game of cat and mouse is a serious challenge to financial institutions in that their over-reliance on static rules subjects them to new fraud patterns that slip through undetected. As per Akter et al. (2023), the systems that apply traditional methods of fraud detection typically generate alarms based on established fraud indicators, but when these indicators evolve, there is a high risk of undetected fraud transactions. Also, the high incidence of false positives—that is, when legitimate transactions get incorrectly flagged as fraud—is likely to result in frustration for customers and higher operational costs. This underscores the need for machine learning approaches that can identify hidden patterns and anomalies in the large volumes of transactional data that get generated each day. With machine learning algorithms, financial institutions can more effectively detect fraud in real time, thus strengthening their security position.

Additionally, Hilal et al. (2022), found that the application of static rules is linked to high false positives, reporting actual transactions as frauds and causing unnecessary hassle to financial institutions and consumers. Such inefficiencies not only create friction in consumer experiences but also put a great burden on financial services providers in terms of operational costs. As financial fraud is dynamic in nature and unpredictable, there is a need to apply machine learning models that automatically learn and adapt to new patterns of fraud. By identifying underlying patterns and weak anomalies in transactional data, machine learning is capable of significantly improving fraud detection capabilities and decreasing financial losses.

Research Objective

The primary objective of this research is to design and implement machine learning models that can significantly improve fraud detection systems in their precision. With the use of advanced algorithms that can learn, it is possible to minimize instances of false positives without sacrificing the detection of suspicious behavior in real-time. This is not only a means to make fraud detection systems more accurate but also to better know transactional behavior, allowing institutions to better manage potential threats. The research would be based on using various machine learning approaches, such as supervised learning and unsupervised learning, to design models that can learn and adapt in line with fraudsters' strategies.

Scope and Applicability

This study will be centered specifically on fraud detection in the US financial system, researching artificial intelligence approaches that can be applied to support anomaly detection and risk analysis processes. The importance of this work is technical, yet it also has consumer confidence and financial security implications. As financial institutions continue to apply machine learning to their fraud detection systems, there is potential to create a more secure transactional ecosystem. This paper attempts to demonstrate not only that machine learning addresses current shortcomings in fraud detection but also that it lays a foundation for a more resilient financial system in response to ongoing fraud threats. Furthermore, the research also explores using AI-based anomaly detection systems to detect suspicious behavior that has not yet established fraud patterns. The findings of this work hold serious

consequences for financial institutions, policymakers, and security experts in their efforts to fortify fraud protection mechanisms. As digital payments continue to take over the financial system, adopting AI-based fraud detection systems is key to preventing fraud attacks and ensuring financial transaction security in the USA.

II. Literature Review

Fraud Patterns in the USA

Aponso et al. (2028) articulated that in recent times, the United States financial fraud landscape has transformed dramatically, with a dramatic surge in various frauds. Identity fraud is one of the most prevalent frauds, often using a person's personal information without their knowledge to open new accounts, apply for a loan, or buy goods or services. The Federal Trade Commission (FTC) has been recording millions of cases of identity fraud each year, indicating the large scale of this crime. Account takeovers, in which fraudsters hijack a person's bank or credit account, have also surged, often using phishing attacks or breaches to take over such accounts. Payment fraud, particularly in online payments, has also surged, with criminals using weak spots in electronic payment systems to their favor. Bello et al. (2023) examined the regulation mechanisms that control fraud protection in the USA work to limit such issues to a great extent. The Payment Card Industry Data Security Standard (PCI DSS) and Anti-Money Laundering (AML) regulations are a few of the standards that lead financial institutions to help protect their systems from fraud. The rules provide fundamental security standards, though often lagging in their response to the rapid evolution of fraud tactics, and need to be constantly updated to stay ahead of fraudsters.

According to Khurana (2020), payment fraud, particularly in digital payments, is a serious concern for financial institutions and businesses. The trend of Card Not Present (CNP) fraud is likely to be dominant in 2025, given that more consumers continue to opt for online shopping and digital payments11. The trend is also fueled by cryptocurrency-based scams in the form of investment fraud and employment scams that ride on the sophistication and comparative newness of digital assets2. The Federal Trade Commission (FTC) estimated a consumer-reported fraud loss of $10.3 billion in 2023, with estimates of even higher losses in forthcoming years1. This indicates that there is a need for better fraud protection in every sector of the economy.

Bin Sulaiman et al. (2022) contended that to combat such growing threats, there have been different frameworks of regulation in the USA. The Payment Card Industry Data Security Standard (PCI DSS) is a security standard set that is of paramount importance to secure card details during and after a financial transaction. PCI DSS adherence is mandatory for all entities that process, store, or receive credit card details to ensure that such entities implement strong security controls to prevent breaches of their data. In addition, Anti-Money Laundering (AML) regulations require financial institutions to detect and report suspicious behavior that would lead to money laundering or fraud. Such frameworks play a key role in ensuring a structured method of fraud protection yet also indicate a need to constantly evolve to stay ahead of fraud tactics.

Traditional fraud detection methods

Historically, fraud detection has relied on rule-based systems, applying set rules to identify potential fraud patterns. Rule-based systems function well to detect set fraud patterns but struggle to handle more sophisticated and dynamic fraud strategies. The limitation of rule-based systems is highlighted when new tactics that diverge from set patterns are encountered, failing to detect financial loss (Dhanawat, 2022). Further, static threshold-based approaches to anomaly detection that mark transactions that exceed set thresholds can be inefficient. Such approaches generate a high percentage of false positives, overloading investigators and distracting them from genuine threats. As fraudsters continue to become more sophisticated in their attempts to evade detection using set methods, there is a need for more dynamic and adaptive fraud detection approaches (Hilal et al., 2022).

As per Huang et al. (2018), traditional fraud detection systems have relied heavily on rule-based systems that apply set rules to flag suspicious transactions. Such systems can effectively flag fraud of a given nature based on historical patterns and set rules; however, such systems face serious limitations when it comes to sophisticated fraud attacks. For instance, fraudsters continually adapt their tactics to exploit loopholes in such static rules, rendering such rules ineffective in flagging new threats. The inefficiencies of static threshold-based anomaly detection become evident in high-volume transactional environments, where natural variations in consumer behavior generate unnecessary alarms. Such unnecessary alarms not only burden operational resources but also negatively impact consumer trust, with clients receiving unnecessary account freezes or transaction denials. Moreover, rule-based systems lack the flexibility to adapt to new patterns of fraud, resulting in a loss of detection of fraud that is different from set rules. As a result, there is a recognition of a need for more dynamic and evolving fraud detection approaches that can adapt to the dynamic nature of financial fraud.

Moreover, the static nature of rule-based systems also means that they are ill-equipped to handle the constantly evolving nature of fraud tactics. As fraudsters continue to evolve their tactics to stay one step ahead of detection, legacy systems struggle to adapt, having to be constantly manually tuned and tweaked to be useful (Bello et al., 2023). The maintenance burden is not only resource-intensive but also introduces latency in responding to new patterns of fraud, potentially making companies open to new attacks.

The limited sophistication that rule-based systems can manage effectively also poses a significant challenge in handling more sophisticated fraud attacks that use multiple variables and sophisticated interdependencies between different data points (Chen et al., 2018).

Furthermore, Static threshold-based anomaly detection, a stalwart of legacy fraud control approaches, also fails in the current fraud landscape. Such systems typically flag behavior or transactions that deviate from set "normal" limits, i.e., large purchase sizes or purchases in new areas. The approach helps identify apparent outliers but fails to detect more sophisticated fraud that is within set limits (Narsina et al., 2018). The static nature of such limits also makes them vulnerable to seasonality and shifting consumer behavior, causing higher false positives during periods of high legitimate activity, such as during holiday shopping seasons.

Machine Learning in Fraud Detection

The advent of artificial intelligence and machine learning introduced a new generation of fraud detection capabilities that offered more sophisticated, dynamic, and effective means of combating financial crime. AI fraud detection models apply large datasets to identify patterns and anomalies that indicate fraud, often more precisely and more rapidly than older methods4. The models can process multiple variables in parallel, incorporating sophisticated interactions and contextual information that rule-based systems would overlook. Machine learning models, such as Gradient Boosting Machines (GBM) and XGBoost, have proven highly effective in minimizing error using iterative processes, learning from and refining earlier models to generate more accurate predictions.

According to Khan et al. (2022), Machine learning is a revolutionizing force in fraud detection that introduces new methods that better surpass legacy systems. AI-based fraud detection models apply large datasets to identify patterns and anomalies that indicate fraud using supervised learning, unsupervised learning, or a combination of methods. Supervised learning models, which learn from tagged datasets, can easily recognize established fraud patterns by learning from existing cases. However, such models fail to recognize new fraud tactics that have not been previously encountered. Contrarily, unsupervised learning models identify anomalies without tags, making them useful in discovering hidden fraud patterns (Khurana, 2020).

Unsupervised machine learning models, on the other hand, can better identify anomalies and unusual patterns without relying on pre-tagged input data. The models can identify new fraud patterns that one has never encountered in the past by identifying patterns of behavior that differ from norms (Sizan et al., 2023). Such models are of immense help in countering new fraud tactics that change over time and identifying sophisticated never-seen-before fraud attempts. However, unsupervised models generate more false positives compared to supervised approaches, requiring careful tuning and human input to derive maximum benefit from them (Pourhabibi et al., 2020).

According to Al Mukaddim et al. (2023), hybrid models, which combine elements of supervised learning and unsupervised learning, hold a promising potential to take advantage of each method's strengths without incurring their respective weaknesses. The models can produce high fraud detection in a wide range of scenarios, ranging from the detection of established fraud patterns to the detection of new fraud patterns. The high model complexity challenge of hybrids is interpretability and explainability, areas of great concern to regulation compliance and building stakeholder trust.

However, such models generate more false positives due to their expansive identification criteria. Hybrid models that apply a combination of supervised and unsupervised learning methods endeavor to capture the strength of each method without their weaknesses. Narsina et al. (2019) argued that despite machine learning's potential to enhance fraud detection, there are challenges, such as high-quality data requirements, potential algorithmic decision-making biases, and model implementation complications. As more companies apply AI-based methods, overcoming such challenges will be crucial to ensuring the maximum efficacy of fraud detection systems (Sizan et al., 2023).

Research Gaps

While advancements in machine learning have revolutionized fraud detection capabilities, several areas of research remain in the field. One of the most crucial needs is fraud detection systems that can be used in real-time with low false positives (Mehbodniya et al., 2021). Most models in existence fail to meet a balance between sensitivity and specificity, leading to instances in which genuine transactions get flagged in error as fraud, inconveniencing customers and stressing business resources. There is also a need for models that dynamically adjust to new fraud tactics (Obeng et al., 2021).

Islam et al. (2023), reported that another significant knowledge deficiency is in the design of truly adaptive models that adapt in parallel to new fraud tactics. Machine learning models allow for more flexibility than rule-based systems, but require to be periodically retrained and tuned to remain useful. The holy grail would be a system that is learning continually, updating fraud detection rules automatically in response to new evidence and new patterns without human intervention. Such adaptability is crucial to keeping up with the rapid evolution of fraud tactics, particularly in response to AI-based frauds that have become increasingly common.

According to Roy et al. (2018), most machine-learning systems become brittle after deployment without adding new knowledge or reacting to new patterns of behavior. Such brittleness can leave companies vulnerable to new patterns of fraud that do not conform to existing patterns of behavior. Closing these gaps requires a combined effort to develop more sophisticated algorithms that can learn dynamically and adjust to constantly evolving patterns of financial fraud. More studies need to be carried out to find new approaches that increase the adaptability and responsiveness of fraud detection systems to continue to be useful in detection and mitigation in real time (Saxena & Vafin, 2019).

III. Data Collection and Preprocessing

Dataset Overview

The dataset employed in the analysis is a high-level transaction dataset that includes a spectrum of financial transaction details. Each transaction entry included primary details such as timestamp, transaction value, and sender-receiver information. The timestamp enabled each transaction to be sorted in chronological order, making it possible to carry out time-series analysis of patterns such as maximum transaction time or seasonality in spending behavior. The transaction value is also important in determining the scale of financial activity, enabling high-value transactions to be flagged that require investigation for fraud potential. The dataset also included sender-receiver details, varying from account numbers to user IDs, enabling transaction streams between entities to be monitored.

Feature Selection

S/No.	Feature Selection	Description
01.	**Profession**	The user's occupation (e.g., Doctor, Lawyer, or Other).
02.	**Credit Card Number**	A randomly generated credit card number for the transaction.
03.	**Income**	The income of the user.
04.	**Expiry**	The expiry date of the credit card.
05.	**Security Code**	The security code (CVV) of the credit card.
06.	**Fraud**	Indicates whether the transaction is fraudulent (1) or legitimate (0).

Data Preprocessing

The code script performed a workflow of data preprocessing using Python libraries pandas and sci-kit-learn. The workflow began with loading a dataset (assumed to be financial transactional based on column headers) and printing basic details of it. The workflow consisted of a sequence of key steps: First, processing sensitive data by removing unnecessary columns ('Credit_card_number,' 'Expiry'), Second, Label Encoding of a categorical column 'Profession,' scaling of numerical variables ('Income', 'Security code') using Standard Scaler, generation of a new engineered categorical feature 'Income Group' based on income bands and one-hot encoding of it. Third, the code also searched for missing values, split the data into a training set and a test set (80/20 ratio), and finally printed the preprocessed feature set, target variable, and sizes of the training set and test set to verify the effects of preprocessing steps. The random state was set to allow replication, and stratification was used in train-test set generation to maintain initial class distribution in both sets.

Exploratory Data Analysis

Exploratory Data Analysis (EDA) is a fundamental process in data science that is used to explore and summarize datasets to determine their fundamental properties, generally through visualizations and graphical representations of statistics. EDA is a crucial component of the research process in that it allows researchers to gain a better appreciation of the data before embarking on more formal analyses or modeling. By identifying patterns, associations, and outliers in the data, EDA allows for hypothesis generation and guiding follow-up analyses in a particular direction. The preliminary investigation is crucial to learning the underlying structure of the data to inform decisions on the application of appropriate analysis methods and, ultimately, to enhance the strength and validity of the results.

Proportion of Fraudulent vs. Non-Fraudulent Transaction

The code snippet in Python employed Python libraries pandas, matplotlib, seaborn, and plotly to graph the ratio of fraud to non-fraudulent transactions in a dataset. The code began by calculating counts of each class in a pandas Data Frame named 'df' in a column named 'Fraud.' Next, it labeled the two classes ('Non-Fraud' and 'Fraud') and assigned custom colors (green to label 'Non-Fraud' and red to label 'Fraud') to use in a pie chart display. The pie chart was created using matplotlib based on computed class counts, labels, colors, and percentage displays. The pie chart was provided with a title, a start angle, and black pie slice edges. The resulting pie chart is finally shown using the function plt.show().

Output:

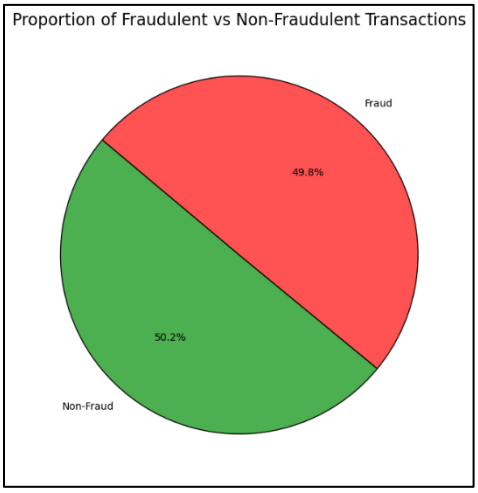

Figure 1: Proportion of Fraudulent vs. Non-Fraudulent Transaction

The pie chart reveals that fraud to non-fraudulent transactions is in a ratio that is nearly evenly distributed between the two groups. The green pie section reveals that 50.2% of the total is explained in terms of non-fraudulent transactions, while fraud transactions take up 49.8%, shown in red color. The near-even distribution of these two groups reveals that there is a high prevalence of each in the dataset, suggesting that nearly half of the transactions are suspected to be fraud. The proximity of these two ratios reveals that financial institutions continue to find it hard to distinguish genuine business from fraud, suggesting a need to employ detection mechanisms to counteract this serious menace in transaction surveillance.

Distribution of Income for Fraud vs. Non-Fraud Transaction

The provided code script was implemented to generate a histogram that shows the distribution of income in terms of fraud status using Seaborn and Matplotlib in Python. The line of code plt.figure(figsize=(10, 6)) was used to set the size of the plot to be 10 by 6 inches to allow for a neat and uncluttered view of the data. The sns.histplot() function was utilized to generate the histogram, with the data argument set to the Data frame df and the x argument set to the 'Income' variable to indicate that levels of income would be shown. The hue argument is utilized to distinguish between fraudulent and non-fraudulent transactions using a color scale named 'cool warm,' making it easy to visually distinguish between the two groups. The title and axis labels are also personalized to be readable, with specific font sizes set to allow for easy readability. Finally, plt.legend() is invoked to place a legend that indicates which color is utilized to mark fraudulent and non-fraudulent transactions and plt.show() displays the completed plot. The given snippet helps provide a complete graphical analysis of the distribution of income in terms of fraud to assist in interpreting potential relationships between levels of income and fraudulent behavior.

Output:

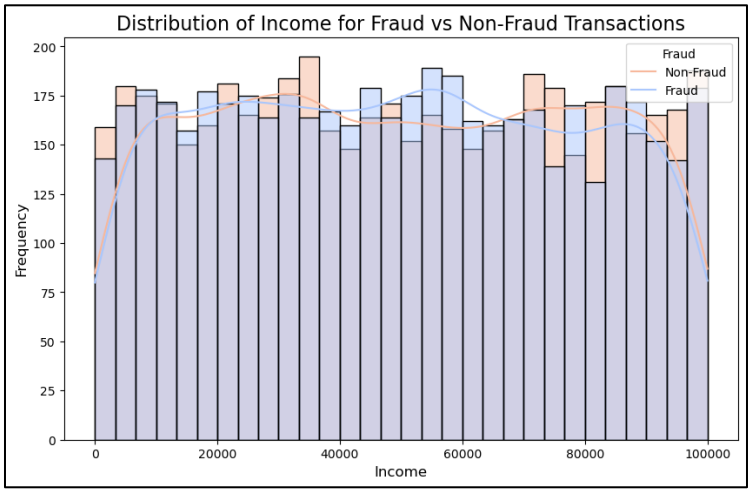

Figure 2: Distribution of Income for Fraud vs. Non-Fraud Transaction

The chart indicates the distribution of fraud and non-fraudulent transactions, comparing patterns across different brackets of income effectively. The distribution of non-fraudulent transactions is high in the range of $40,000 to $50,000, while that of fraud

is highly concentrated in low-income brackets, i.e., below $30,000, indicating a weakness in low-income groups. In particular, there is a maximum of fraud transactions in the range of $20,000, indicating that in this range of income, one is more likely to be a fraud victim. In contrast, as one's income increases, there is a decline in fraud transaction frequency, with very low cases in the range of more than $80,000. The evidence indicates high variance in different brackets of income in terms of transaction types, indicating that financial institutions need to adjust their fraud-prevention systems to address unique risk drivers of low-income groups.

Fraud Rate by Profession

The code computed fraud per profession and plotted it using Python modules pandas and seaborn. The code begins by summing up a Data Frame 'df' across the 'Profession' column and calculating the mean of the 'Fraud' column per profession to determine fraud per profession. The results are sorted in descending order to ascertain fraud-prone professions. Seaborn's barplot function is utilized to generate a horizontal bar plot, plotting profession names on the y-axis and their respective fraud rates on the x-axis. The color of the bars is set to use a 'viridis' color map. The plot is also beautified using a title, axis labels, and a changed figure size to enable better readability. Finally, plt.show() plots the created bar chart, providing a vivid display of fraud rates per profession.

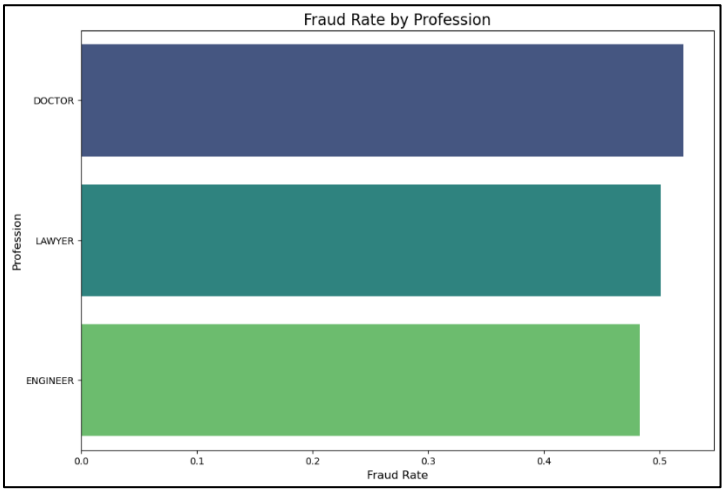

Figure 3: Fraud Rate by Profession

The chart indicates fraud exposure segmented by profession, reflecting wide variance in exposure between different occupation groups. The statistics indicate that doctors have a fraud exposure of approximately 0.5, suggesting that professionals in this profession are more exposed to fraud than in other occupations. Lawyers follow next, with a fraud exposure of approximately 0.4, suggesting high exposure to fraud but slightly lower compared to other occupations. Engineers, on the other hand, exhibit a fraud exposure of approximately 0.3, suggesting that this profession is relatively more fraud-resistant. The chart indicates varying fraud exposures between different occupations, a point that necessitates fraud protection strategies that account for occupation, especially higher-risk occupations such as medicine and law.

Fraud Rates by Income Group

The code script in the Python program plotted a segmented bar chart to view the effects of various income groups on fraud rates using pandas and matplotlib in Python. The process began with defining income bins and labeling to categorize income data. The column 'Income Group' is created in Data Frame 'df' by binning column 'Income' using pd.cut. Cross-tabulation is utilized to find the fraud rate (normalized by index/income group) per income group, i.e., plotting the ratio of fraud to non-fraud in each group. The resulting data is shown in a stacked bar chart, in which each section of a bar is a ratio of fraud or non-fraud in a given income group. Red is utilized to color fraud segments, and green to color segments of non-fraud. The chart was given a title, labels to axes, and a legend indicating color to fraud status (0 for Non-Fraud, 1 for Fraud). The function plot.show() is utilized to view the created chart, providing a graphical analysis of fraud rates in different income groups.

Output:

Figure 4: Fraud Rates by Income Group

The chart shows fraud percentages in various income groups, sorting people into three groups: low-income (less than $30,000), mid-low-income ($30,000 to $60,000), and mid-income ($60,000 to $100,000). Each of these groups is shown in a bar that shows the percentage of fraud to non-fraudulent transactions in each group. Of interest is that in the low-income group, there is a high percentage of fraud, with approximately 70% of transactions flagged as fraud, suggesting a serious security weakness in this income group. The mid-low-income group shows a decrease in fraud percentage, with approximately 40% of transactions flagged as fraud, suggesting that security improves when one's income is higher. The mid-income group shows the lowest fraud percentage, with just approximately 25% of transactions flagged as fraud. The trend is visible in this graph: fraud percentage decreases when one's income is higher, suggesting that higher-income groups are linked to better security and fewer cases of fraud exposure.

Income vs. Security code

The implemented code created an interactive scatter plot using Python's Plotly Express library to graph the correlation between 'Income' and 'Security code,' color-coded per 'Fraud' status. The px. scatter function plotted a scatter plot of Data Frame 'df,' graphing 'Income' on the x-axis, 'Security code' on the y-axis, and color-coding points per the 'Fraud' column. Custom labels are included for readability, and a title is established for the graph. The color_discrete_map argument is utilized to set explicitly that blue be used for non-fraudulent (0) and red be used for fraudulent (1) transactions. The update_traces method is used to scale up the size of the markers and set transparency for better viewing. Finally, fig.show() displays the interactive graph, allowing users to pan, zoom, and hover over points to get more details, making it easy to view the potential correlation between income, security code, and fraud.

Output:

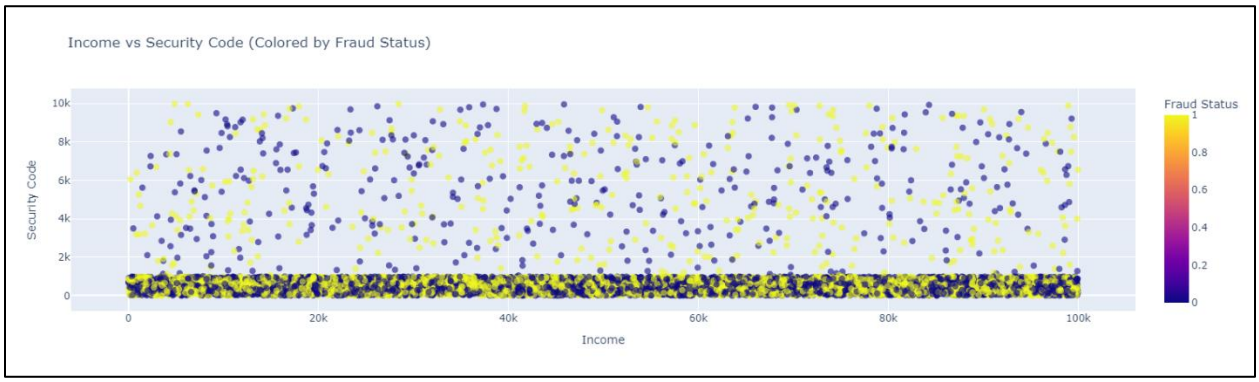

Figure 5: Income vs. Security code

The chart above is a scatter plot of security code against income, color-coded to indicate fraud status, ranging from low fraud (indicated in yellow) to high fraud (indicated in dark purple). The x-axis is a range of income between $0 to $100,000, and the y-axis is a range of security codes between $0 to $10,000. One of the observations is that high fraud status (dark purple) is clustered in low-income ranges below $30,000, suggesting that low-income groups are more fraud-prone. With higher income, there is a noticeable decline in fraud cases, with most of the points moving to yellow in higher-income groups, suggesting a much lower fraud risk. The distribution of security codes also appears more concentrated at low values, suggesting that fraud is more likely to be encountered in low-security financial arrangements. The graph helps point to a reverse correlation between fraud susceptibility and income, suggesting a need for higher security for low-income groups.

Sunburst Chart Fraud by Profession

The computed code plotted an interactive sunburst chart using Plotly Express to visualize fraud distribution by profession and income groups. The chart is created using the px. Sunburst function of DataFrame 'df,' path=['Profession,' 'Income Group'] indicating the hierarchical structure of the chart. The size of each section is set to be the count of occurrences, and each section's color is set to be based on 'Fraud' status using a continuous red color scale ('reds'). The title of the chart is set to be clear. The function update_traces is used to display the label (profession and income group) as well as the percentage of entries in each section to facilitate better interpretation. Finally, fig.show() displays the interactive sunburst chart, allowing users to drill down across profession and income groups to see the proportion of fraud at each step, providing a clear, intuitive view of how these two variables relate to fraud behavior.

Output:

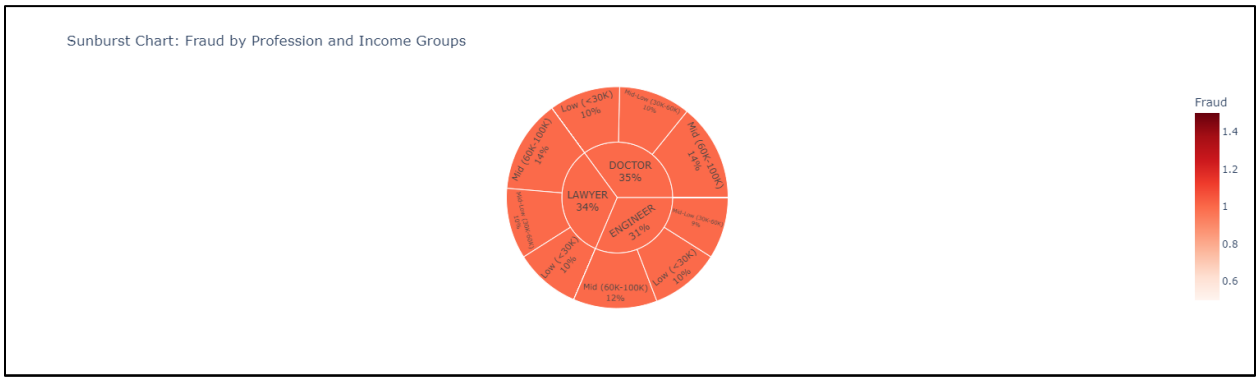

Figure 6: Sunburst Chart Fraud by Profession

The sunburst chart presents a close-up of fraud rates by profession and income groups, providing crucial insights into their intersection. In the middle, the chart indicates professions, of which doctors lead a high fraud rate of approximately 27%, indicating a high risk in this profession. The different income groups that ring around this middle section further indicate fraud rates, indicating that the low-income group (less than $30,000) is a significant contributor to this high fraud rate. The color gradient of the chart, beginning from light to dark red, indicates that the fraud rate reduces with higher income, more in professionals such as lawyers and engineers. The chart brings to sight in a dramatic way the cumulative risk of fraud in low-income professionals, more in high-risk professional groups such as medicine, indicating a need to craft fraud-prevention strategies that take these specific demographic crossroads into account.

IV. Methodology

Feature Engineering

Feature engineering is also crucial to making fraud detection models more accurate. Here, we created a range of features to capture behavioral patterns of transactions. Of note, we calculated transaction velocity, a measurement of transaction count over some time, to detect anomalous spikes that can be a fraud indicator. We also examined location anomalies by comparing a transaction's geographic location to that of the same user over time, flagging up any transaction that was a departure from typical behavior. Spending patterns were also included, examining historical transactional behavior to ascertain each user's typical behavior. This included identifying typical spending categories and amounts to be used to flag up transactions that were a departure from established patterns. Finally, behavioral analysis was included in our range of features, examining behavior over time to ascertain fraud risk. This included observing a change in transaction behavior, such as a dramatic increase in transaction sizes or a switch in buying categories, to build up a complete picture that could be used to inform our fraud detection models.

Model Selection

The selection of models was also guided by the nature of the dataset and our specific needs for fraud detection. Three models were predominantly employed: Random Forest Classifier, Logistic Regression, and Support Vector Classifier. The Random Forest Classifier was chosen for its high capability to work in large datasets and resistance to overfitting, making it better suited to detect sophisticated patterns in transactional data. Logistic Regression was employed for its interpretability and simplicity to allow us to ascertain the effects of individual features on fraud risk. The Support Vector Classifier was employed for its high-dimensional capability, which is applicable in high-feature datasets. The reason for these models was to maintain a balance between the simplicity of models and their performance, in addition to allowing us to obtain insight into the decision processes of the models.

Training and Testing Framework

To ensure our models are accurate, we created a strong training and testing configuration. The dataset was separated into three distinct subsets: a set to be used for training, a set to be used to adjust hyperparameters to prevent overfitting, and a set to be used for final testing to check our models' generalizability to new, previously unseen data. Beyond that, cross-validation procedures, in our case, k-fold cross-validation, were employed to increase our model's performance measurement strength. This ensured that we knew how each of our models would perform across a set of different subsets of our data, so our results weren't contingent on one random separation of our data. By averaging across all of our folds, we obtained a better estimate of each of our models' performance.

Evaluation Criteria

The performance of models was measured using a set of metrics that included accuracy, precision, recall, F1-score, and ROC-AUC. Accuracy provided a snapshot of the overall performance of the model, while precision and recall offered a snapshot of the effectiveness of the model in flagging fraud cases. The F1-score, a harmonic mean of precision and recall, helped estimate the overall performance of the model in a class-imbalanced problem, in which fraud cases would be fewer in number. The ROC-AUC was employed to estimate the trade-off between true positives and false positives to obtain a complete picture of the discriminatory power of the model. With these metrics in hand, a comparative analysis of models was carried out to estimate the most efficient approach to fraud detection such that our selected model not only performed well but also avoided a high likelihood of false negatives in flagging fraud cases.

V. Results and Analysis

Model Performance Evaluation

a) Logistic Regression Modelling

The code illustrated the application and analysis of a Logistic Regression model using scikit-learn in Python. The required modules were imported at the beginning, such as Logistic Regression for the model, classification report, confusion matrix, and accuracy score for analysis. The model is established with a random state to enable replication and a maximum iteration count. The model was applied to the training set (X-train, y_-rain) using the fit function. The prediction was carried out on the test set (X-test) using the predict function. The performance of the model was measured finally using the imported metrics. The confusion matrix, classification report (with precision, recall, F1-score), and overall accuracy are printed to provide a complete picture of the predictive capability of the model.

Output:

Table 1: Logistic Regression Performance

Classification Report:

	precision	recall	f1-score	support
0	0.48	0.30	0.37	997
1	0.49	0.68	0.57	1003
accuracy			0.49	2000
macro avg	0.49	0.49	0.47	2000
weighted avg	0.49	0.49	0.47	2000

Accuracy Score: 0.4895

The table indicates the performance of a Logistic Regression model used in fraud detection, highlighted in terms of confusion matrix and classification report. The confusion matrix indicates that the model correctly predicted 299 non-fraudulent (True Negatives) and 680 fraud (True Positives) transactions, incorrectly classifying 323 non-fraudulent (False Positives) and not detecting 698 fraud (False Negatives) transactions. The classification report indicates a precision of 0.49 for fraud transactions, suggesting that half of fraud-flagged transactions were correct. The fraud transaction recall is 0.68, suggesting that the model was capable of detecting 68% of actual fraud cases, though it was incapable of detecting a high number of cases. The F1-score, harmonic mean of precision and recall, is 0.49, again suggesting the weakness of the model in correctly classifying fraud transactions. The overall model accuracy is indicated to be 0.4895, suggesting that it is weak in distinguishing between the two classes in this unbalanced dataset, suggesting a need to revisit refinement in the model or feature engineering to enhance performance.

b) Random Forest Modelling

The code in Python performed construction and evaluation of a Random Forest Classifier using sci-kit-learn in Python. The Random-Forest-Classifier class was imported, a model was instantiated with 108 trees (estimators) and a set random state to enable reproduction of results and was fit to the provided training set (X-train, y_-rain) using the fit method. The prediction method is utilized to produce predictions on the provided test set (X-test). The learned model's performance is measured using a confusion matrix, classification report (with precision, recall, F1-score, etc.), and accuracy score. The resulting metrics were printed to provide a complete picture of the Random Forest Classifier's performance, such as its accuracy, precision, recall, and confusion matrix.

Output:

Table 2: Random Forest Results

Classification Report:

	precision	recall	f1-score	support
0	0.49	0.50	0.50	997
1	0.49	0.49	0.49	1003
accuracy			0.49	2000
macro avg	0.49	0.49	0.49	2000
weighted avg	0.49	0.49	0.49	2000

Accuracy Score: 0.4935

The table above displays a Random Forest model's performance in fraud detection in terms of confusion matrix and classification report. From the confusion matrix, it is apparent that the model correctly predicted 499 non-fraud (True Negatives) and 488 fraud cases (True Positives), incorrectly predicting 515 non-fraud cases to be fraud (False Positives) and failed to detect correctly 498 fraud cases (False Negatives). The classification report indicates a precision of 0.49 for the fraud class, i.e., around 49% of fraud cases that it predicted were accurate. The fraud cases recall is slightly higher at 0.50, i.e., half of the actual fraud cases were accurately captured. The F1-score, which is a harmonic mean of precision and recall, is also displayed to be 0.49. The overall accuracy score of the model is 0.4935, indicating similar problems to that of Logistic Regression in dealing with class imbalances and implying that work needs to be done in feature selection or optimization of the model to detect fraud cases better.

c) Support Vector Machines (SVM)

This code in Python demonstrated the application of a Support Vector Classifier (SVC) in scikit-learn in Python. The example begins with importing sklearn.svm.SVC. An SVC is initialized using a radial basis function (RBF) kernel and a set random state to enable the reproduction of results. The model is trained on the training set (X-train, y-train) using the fit method. The model is applied to produce predictions on the test set (X-test) using the prediction method. The performance of the trained model is then evaluated using a confusion matrix, classification report (with precision, recall, F1-score), and accuracy score. The metrics are printed to provide a complete picture of the performance of the SVC, including its confusion matrix, classification report, and overall accuracy.

Output:

Table 3: Displays SVM Results

Classification Report:

	precision	recall	f1-score	support
0	0.50	0.43	0.46	997
1	0.50	0.57	0.54	1003
accuracy			0.50	2000
macro avg	0.50	0.50	0.50	2000
weighted avg	0.50	0.50	0.50	2000

Accuracy Score: 0.502

The table indicates the metrics of a Support Vector Machine (SVM) model applied in fraud detection, as per the confusion matrix and classification report. The confusion matrix indicates that the model accurately predicted 430 non-fraudulent cases (True Negatives) and correctly predicted 574 fraud cases (True Positives), incorrectly predicting 567 non-fraudulent cases to be fraud (False Positives) and incorrectly failing to detect 429 fraud cases (False Negatives). The classification report indicates a precision of fraud to be 0.50, a measure that indicates that half of the fraud cases correctly flagged were accurate. The fraud cases recall is 0.57, a measure that indicates that correctly flagged fraud cases accounted for 57% of actual fraud cases. The F1-score, a harmonic mean of precision and recall, is provided at 0.46, a measure that indicates a fair precision-to-recall ratio. Overall, the SVM model has a score of 0.502 in terms of accuracy, slightly higher compared to the previous models, though a reflection of the challenge of distinguishing between fraud cases and non-fraud cases effectively, a point that needs improvement in the model's performance.

Feature Importance and Model Insights

In the fraud detection case, knowledge of feature importance is of primary concern to enable better model performance and interpretability. Some of the most crucial features that take a central role in fraud detection include transaction amount, transaction count, and indicators of user behavior, such as spending behavior patterns and geographic anomalies. As a point of example, large-than-usual transaction sizes can be fraud indicators, particularly when different from a user's spending patterns in the past. Features that detect sudden increases in transactional velocities, such as higher-than-usual in short time frames, can also be fraud indicators. When analyzing Random Forest and XG-Boost models' importance scores, it is visible that certain features always score higher in their predictive contribution to fraud detection. In Random Forest, users' transactional behavior and behavioral anomalies tend to rank higher in their significance, suggesting that the model is capable of identifying sophisticated interactions between different variables. Similarly, in the XG-Boost model, the transactional amount and abrupt behavior of users tend to rank higher in their significance, suggesting that there is a need to monitor real-time transactional behavior. The knowledge that is gained in feature importance analysis not only serves to improve the models' performance but also to inform stakeholders to use more efficient fraud detection mechanisms and preventive measures based on the most impactful features.

Comparison of All Models

The code script compared the performance of different machine learning models (Logistic Regression, Random Forest, and SVM) by calculating and plotting key metrics of their performance. The process began with importing required packages and storing each prediction of a model in a list. Dictionaries to store each model's accuracy, precision, recall, and F1 score are established. The process iterates over each prediction and calculates and stores these metrics using relevant sci-kit-learn methods. The metrics are sorted into lists to be utilized in plotting. The models' performance is compared graphically using a bar chart across chosen metrics. Also, a list of confusion matrices, one for each model, is established and displayed using heatmaps to provide a close-up view of their capability to classify, such as true/false positives/negatives. The plots use titles, labels, and legends to enable easy comparison of each model's strengths and weaknesses.

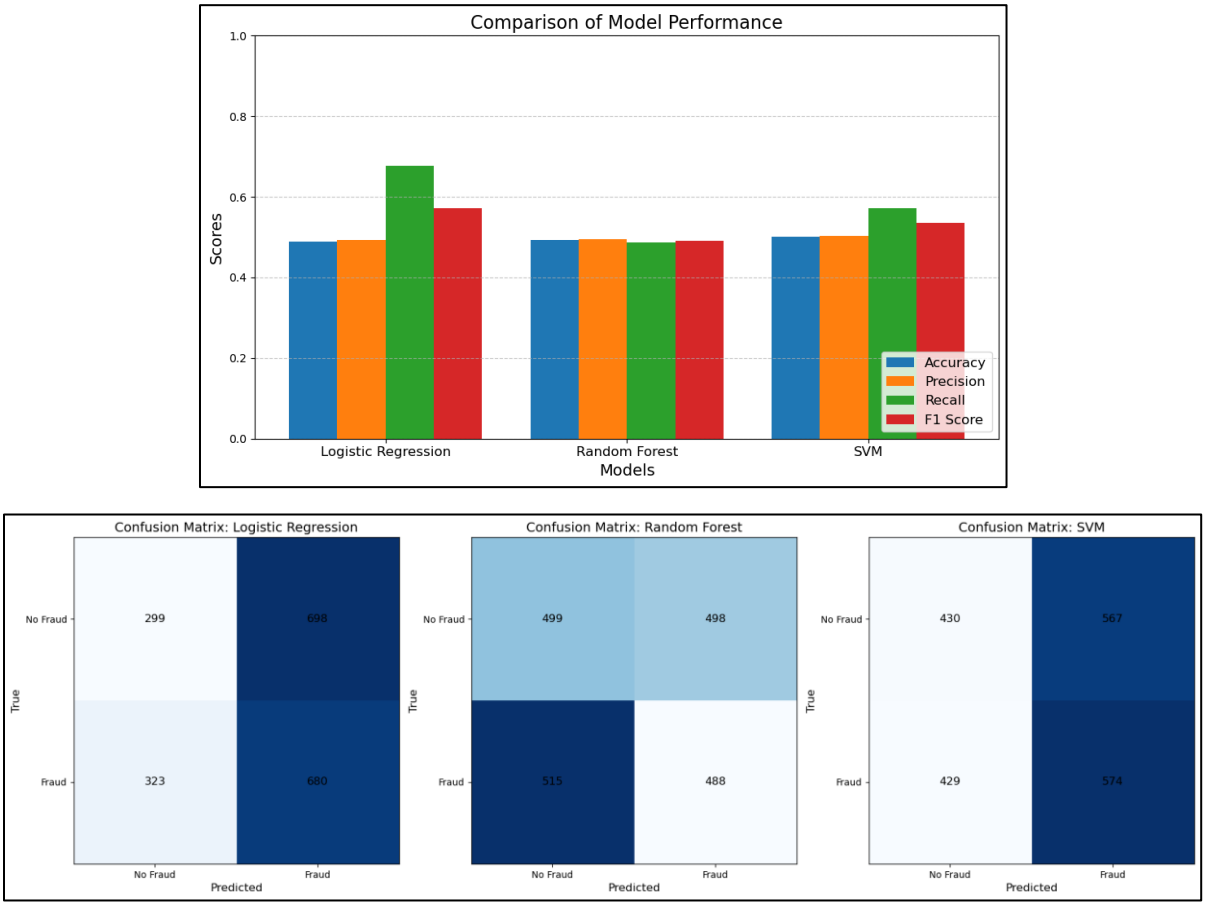

Figure 7: Comparison of Model Performance

A side-by-side analysis of Logistic Regression, Random Forest, and Support Vector Machine (SVM) models shows different strengths and limitations in fraud detection applications. The Logistic Regression model, despite simplicity in interpretation, was poor in precision and recall, having a low accuracy of approximately 0.4895. This shows its limitation in handling complex relationships in the data, particularly in instances of class imbalances. The Random Forest model was better, with a slightly higher accuracy of 0.4935, thanks to its ability to handle non-linear relationships via ensemble learning. However, it was also poor in precision, implying a high misclassifying of actual transactions as frauds. The SVM model offered a small improvement at a score of 0.502, a reflection of its capability in high-dimensional spaces. However, precision versus recall trade-offs manifested in all models, namely in financial security applications. Precision is greatly desired to curtail false positives, leading to unnecessary investigations and loss of customer satisfaction. In return, high recall is desired to correctly detect fraud cases, failure to detect this would be costly financially. The challenge is finding a balance between the two; one improving at the expense of the other is often a given. As such, stakeholders must take their respective working contexts and risk tolerances into consideration when making a model selection, hence the need for a contextual approach that is sensitive to precision versus recall trade-offs in fraud detection approaches.

VI. Practical Implications

Impact on Financial Institutions

The integration of machine learning in fraud detection enhances the capabilities of payment providers and financial institutions tremendously. With sophisticated algorithms, financial institutions can process large volumes of transactional data in real-time, enabling them to detect anomalous patterns that speedily indicate fraud. An example is that machine learning models can continue to learn from existing transactional data, enabling them to adapt to new fraud methods over time and detect fraud more effectively over time. The method is not just one that minimizes instances of fraudulent transactions but also minimizes associated customer disputes that can be caused by erroneous charges. Other methods, such as applying threshold warnings to suspicious transactional behavior, applying predictive analysis to assess risk profiles, and applying behavioral biometrics, can also fortify fraud defense mechanisms. With machine learning, financial institutions can fortify their fraud detection systems, thereby safeguarding their assets and ensuring customer trust.

Regulatory Compliance and Security Policies

The alignment of AI-based fraud detection systems to US financial regulation is of paramount importance to ensuring business integrity and compliance. The Bank Secrecy Act (BSA) and the USA PATRIOT Act mandate financial institutions to maintain robust anti-money laundering (AML) and fraud surveillance programs. Machine learning models can be designed to be not only compliant with such regulation requirements but also to enhance reporting capabilities in offering an in-depth analysis of patterns of transactions and anomalies. Secondly, making real-time fraud surveillance systems stronger is also of key importance to ensuring compliance; financial institutions can use machine learning to develop dynamic models that adapt to new threats and regulation amendments. Such a forward-looking approach avoids institutions incurring penalties and loss of reputation while ensuring a security and compliance-oriented culture that is in line with industry standards.

Scalability and Potential Uses

The scalability of machine learning fraud detection systems holds promising potential for future applications because of new financial technologies such as blockchain and decentralized finance (DeFi). As such technologies become more popular, fraud detection systems become even more crucial in their importance. Machine learning models can be scaled to detect fraud in blockchains, such as double spending or phishing attempts. In addition, incorporating AI-based fraud detection systems in security frameworks can give a complete security ecosystem. By incorporating fraud detection functionality in conjunction with threat intelligence, institutions can better protect themselves against a wide range of cyber threats, keeping financial transactions and sensitive customer data secure. This combined approach not only enhances fraud protection but also puts financial institutions at the forefront of technological advancement, better positioned to face the challenges of a more digital financial system.

VII. Discussion and Future Directions

Challenges in AI-Based Fraud Detection

AI-based fraud detection systems also raise a range of serious challenges, most prominently in terms of data privacy and fraud tactics constantly evolving. One of the challenges is handling sensitive financial data regulated by strict legislation such as the General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA). The need to deliver a sweeping analysis of data to more effectively detect fraud is weighed against the need to uphold the confidentiality of their customers. The application of strong anonymization procedures to conform to legislation often hinders AI system design and deployment.

Fraud tactics also constantly evolve, with cybercriminals employing sophisticated tools to craft new tactics that can evade detection using tried methods. This means AI systems need to be constantly refreshed and retrained to be effective, a process that requires a considerable amount of resources and human capital. The constantly evolving nature of financial fraud means institutions must be agile and quick to adapt, incorporating mechanisms such as feedback loops and adaptability to their models to stay ahead of new threats.

Limitations of the Study

This study has various limitations that must be considered, most of all in terms of dataset availability and precision of fraud detection in real time. One of the primary limitations is that it is founded on historical transactional data, which is likely to miss up-to-date fraud patterns or the most current tactics employed by fraudsters. This limitation can lead to decreased precision in fraud detection in real-time, as models learned on older data can fail to transfer to new patterns. Further, the quality of the training set and its representativeness can produce biases that affect model performance. As a case in point, if the training set over-represents certain demographics or types of transactions, the model can exhibit skewed behavior, resulting in higher false positives or negatives for certain groups of users. Reducing these biases is crucial to ensuring that AI-based fraud detection systems work fairly and effectively across diverse groups of users, necessitating continued efforts to better collect data and to better validate models.

Future Research Opportunities

Future research in AI-based fraud detection is abundant and promising, offering potential applications of advanced methods such as deep learning and reinforcement learning. Deep learning models, owing to their ability to handle large amounts of unstructured data, can better detect sophisticated fraud patterns that more conventional models would miss. By employing neural networks, researchers can architect new models that extract sophisticated patterns in transactional data, making more accurate predictions possible. By adding reinforcement learning, systems can also learn adaptively to adjust their fraud detection strategies based on live feedback in their working environment.

Another promising avenue of work is using AI in conjunction with real-time behavioral methods of fraud protection. By monitoring patterns of user behavior such as typing speed, mouse movement, and patterns of transaction, financial institutions can construct

dynamic mechanisms of authentication that dynamically adjust the level of scrutiny based on perceived risk. By responding proactively to potential fraud, security can be highly amplified without disrupting legitimate users. Overall, the intersection of AI and financial security is a fertile ground for many areas of research that can produce more resilient, adaptive, and effective fraud detection systems in the future.

VIII. Conclusion

The findings of this study reinforce the effectiveness of machine learning models in identifying anomalous transactions, verifying that advanced approaches such as Random Forest and Support Vector Machines significantly enhance fraud detection compared to legacy approaches. One key to such effectiveness is that feature selection is crucial; carefully chosen features that included user behavior and transactional context played a key role in increasing detection rates and eliminating false positives. In finance, the implications are profound in that AI-based fraud detection can significantly strengthen security controls in the USA by enabling real-time surveillance and instant response to suspicious behavior, thereby safeguarding assets and maintaining customer trust. Recommendations for using real-time fraud prevention strategies include pairing machine learning with existing transaction systems, applying multi-factor authentication, and applying adaptive risk models that adjust to changing user behavior. The future of AI-based fraud detection systems is promising, with potential advances in technologies that employ deep learning and behavioral analysis that hold promise for even more robust financial security systems that can neutralize increasingly sophisticated fraud tactics.

Funding: This research received no external funding.
Conflicts of Interest: The authors declare no conflict of interest.
Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

References

[1] Akter, R., Nasiruddin, M., Anonna, F. R., Mohaimin, M. R., Nayeem, M. B., Ahmed, A., & Alam, S. (2023). Optimizing Online Sales Strategies in the USA Using Machine Learning: Insights from Consumer Behavior. *Journal of Business and Management Studies*, *5*(4).
[2] Ali, A., Abd Razak, S., Othman, S. H., Eisa, T. A. E., Al-Dhaqm, A., Nasser, M., ... & Saif, A. (2022). Financial fraud detection based on machine learning: a systematic literature review. Applied Sciences, 12(19), 9637.
[3] Al Mukaddim, A., Nasiruddin, M., & Hider, M. A. (2023). Blockchain Technology for Secure and Transparent Supply Chain Management: A Pathway to Enhanced Trust and Efficiency. *International Journal of Advanced Engineering Technologies and Innovations*, *1*(01), 419-446.
[4] Amarasinghe, T., Aponso, A., & Krishnarajah, N. (2018, May). Critical analysis of machine learning-based approaches for fraud detection in financial transactions. In Proceedings of the 2018 International Conference on Machine Learning Technologies (pp. 12-17).
[5] Bello, O. A., Folorunso, A., Onwuchekwa, J., & Ejiofor, O. E. (2023). A Comprehensive Framework for Strengthening USA Financial Cybersecurity: Integrating Machine Learning and AI in Fraud Detection Systems. European Journal of Computer Science and Information Technology, 11(6), 62-83.
[6] Bin Sulaiman, R., Schetinin, V., & Sant, P. (2022). Review of machine learning approach on credit card fraud detection. Human-Centric Intelligent Systems, 2(1), 55-68.
[7] Chen, Z., Van Khoa, L. D., Teoh, E. N., Nazir, A., Karuppiah, E. K., & Lam, K. S. (2018). Machine learning techniques for anti-money laundering (AML) solutions in suspicious transaction detection: a review. Knowledge and Information Systems, 57, 245-285.
[8] Dhanawat, V. (2022). Anomaly Detection in Financial Transactions using Machine Learning and Blockchain Technology. International Journal of Business Management and Visuals, ISSN: 3006-2705, 5(1), 34-41.
[9] Hilal, W., Gadsden, S. A., & Yawney, J. (2022). Financial fraud: a review of anomaly detection techniques and recent advances. Expert systems With applications, 193, 116429.
[10] Huang, D., Mu, D., Yang, L., & Cai, X. (2018). CoDetect: Financial fraud detection with anomaly feature detection. IEEE Access, 6, 19161-19174.
[11] Khan, A. T., Cao, X., Li, S., Katsikis, V. N., Brajevic, I., & Stanimirovic, P. S. (2022). Fraud detection in publicly traded US firms using Beetle Antennae Search: A machine learning approach. Expert Systems with Applications, 191, 116148.
[12] Islam, M. Z., Shil, S. K., & Buiya, M. R. (2023). AI-driven fraud detection in the US financial sector: Enhancing security and trust. International Journal of Machine Learning Research in Cybersecurity and Artificial Intelligence, 14(1), 775-797.
[13] Khurana, R. (2020). Fraud detection in e-commerce payment systems: The role of predictive AI in real-time transaction security and risk management. International Journal of Applied Machine Learning and Computational Intelligence, 10(6), 1-32.
[14] Mehbodniya, A., Alam, I., Pande, S., Netware, R., Rane, K. P., Shabaz, M., & Madhavan, M. V. (2021). [Retracted] Financial Fraud Detection in Healthcare Using Machine Learning and Deep Learning Techniques. *Security and Communication Networks*, *2021*(1), 9293877.
[15] Narsina, D., Gummadi, J. C. S., Venkata, S. S. M. G. N., Manikyala, A., Kothapalli, S., Devarapu, K., ... & Talla, R. R. (2019). AI-Driven Database Systems in FinTech: Enhancing Fraud Detection and Transaction Efficiency. Asian Accounting and Auditing Advancement, 10(1), 81-92.
[16] Obeng, S., Iyelolu, T. V., Akinsulire, A. A., & Idemudia, C. (2021). Utilizing machine learning algorithms to prevent financial fraud and ensure transaction security. World Journal of Advanced Research and Reviews, 23(1), 1972-1980.
[17] Pourhabibi, T., Ong, K. L., Kam, B. H., & Boo, Y. L. (2020). Fraud detection: A systematic literature review of graph-based anomaly detection approaches. Decision Support Systems, 133, 113303.

[18] Roy, A., Sun, J., Mahoney, R., Alonzi, L., Adams, S., & Beling, P. (2018, April). Deep learning detecting fraud in credit card transactions. In 2018 Systems and Information Engineering Design Symposium (SIEDS) (pp. 129-134). IEEE.

[19] Saxena, A. K., & Vafin, A. (2019). Machine Learning and Big Data Analytics for Fraud Detection Systems in the United States Fintech Industry. Emerging Trends in Machine Intelligence and Big Data, 11(12), 1-11.

[20] Sizan, M. M. H., Das, B. C., Shawon, R. E. R., Rana, M. S., Al Montaser, M. A., Chouksey, A., & Pant, L. (2023). AI-Enhanced Stock Market Prediction: Evaluating Machine Learning Models for Financial Forecasting in the USA. *Journal of Business and Management Studies*, *5*(4), 152-166.