International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 12 Issue: 2
Article Received: *25 November 2023* ***Revised:*** *12 December 2023* ***Accepted:*** *30 January 2024*

Generative AI in Financial Fraud Detection

Shenson Joseph
Graduated from Texas tech
shenson.joseph@gmail.com

Abstract**—**the objective of this exploration is to utilize AI calculations to recognize Visa misrepresentation. Because machine learning algorithms can examine enormous datasets and uncover patterns suggestive of fraud, they present a promising method for detecting fraudulent transactions. An experimenta design is used in the research approach to gather, prepare, and analyze data. The study's dataset, which included 23 variables linked to financial transactions, was sourced from Kaggle. Feature engineering, addressing missing values, and removing superfluous columns were all part of the data pre-processing step. Three machine learning models — the Random Forest Classifier, K-Nearest Neighbor (KNN), and Support Vector Machine (SVM) — were prepared and evaluated. The models were evaluated utilizing performance measures such area under the curve (AUC), review, exactness, accuracy, and F1-score. The review's discoveries show that Random Forest and KNN beat SVM in the distinguishing proof of Mastercard fraud. With an exactness of 99.63%, accuracy of 99.53%, review of 99.45%, and F1-score of 99.80%, Random Forest performed very well. With an exactness of 99.45%, accuracy of 99.87%, review of 99.89%, and F1-score of 99.80%, KNN performed all around well. With a F1-score of 99.61%, an exactness of 99.58%, accuracy of 99.74%, and review of 100 percent, SVM performed all around well. As per the examination, machine learning calculations might be helpful in recognizing Mastercard fraud. The capacity of machine learning calculations to recognize Visa theft is shown by this review. The results exhibit that Random Forest and KNN perform preferable on this task over SVM. The outcomes offer critical viewpoints for financial foundations and endeavors seeking to further develop their fraud detection systems. To expand the exactness of fraud detection, future examination could focus on exploring elective machine learning calculations and systems.

Keywords— Financial fraud detection, Generative Adversarial Networks (GANs), Machine learning, Random Forest, K-Nearest Neighbor, Support Vector Machine, Data preprocessing, Performance evaluation

I. INTRODUCTION

The dynamic sphere of modern finance, we are inextricably linked to the rapid rate of technology progress and a reliance on the intangible world of digital transactions. But within this symphonic choreography, an enigmatic and complex force emerges—a thing known as financial fraud, an elusive force that haunts the virtual worlds. Cybercriminals, akin to enigmatic spirits, are ever-changing in their tactics, delving into intricate vulnerabilities within financial systems to execute an array of illicit operations. Identity theft and unauthorized access are two of the many challenges financial organizations face in preserving the integrity of transactions [1]. Traditional fraud detection techniques are losing efficacy as technology is increasingly incorporated into financial systems. Due to the volume and complexity of financial transactions, creative solutions are required to manage the intricate dance of always evolving fraudulent behavior. The successful combination of machine learning (ML) with artificial intelligence (computer based intelligence) is viewed as a significant breakthrough in fraud detection frameworks, expanding efficiency.

A. Understanding Generative AI

A group of machine learning models known as "generative man-made intelligence" can create new data tests that are equivalent to those in the preparation dataset. In contrast to conventional AI models that concentrate on classification or prediction tasks, generative models understand the data's underlying distribution and produce new samples that closely resemble the original data [2].

Variational Auto encoders (VAEs):

A class of generative models known as variational autoencoders (VAEs) learns a minimal portrayal of information data in a lower-layered dormant space. An encoder and a decoder are the two essential pieces of a VAE. The

Manuscript received October 9, 2001. (Write the date on which you submitted your paper for review.) This work was supported in part by the U.S. Department of Commerce under Grant BS123456 (sponsor and financial support acknowledgment goes here). Paper titles should be written in uppercase and lowercase letters, not all uppercase. Avoid writing long formulas with subscripts in the title; short formulas that identify the elements are fine (e.g., "Nd–Fe–B"). Do not write "(Invited)" in the title. Full names of authors are preferred in the author field, but are not required. Put a space between authors' initials.
F. A. Author is with the National Institute of Standards and Technology, Boulder, CO 80305 USA (corresponding author to provide phone: 303-555-5555; fax: 303-555-5555; e-mail: author@ boulder.nist.gov).
S. B. Author, Jr., was with Rice University, Houston, TX 77005 USA. He is now with the Department of Physics, Colorado State University, Fort Collins, CO 80523 USA (e-mail: author@lamar.colostate.edu).
T. C. Author is with the Electrical Engineering Department, University of Colorado, Boulder, CO 80309 USA, on leave from the National Research Institute for Metals, Tsukuba, Japan (e-mail: author@nrim.go.jp).

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 12 Issue: 2
Article Received: *25 November 2023* ***Revised:*** *12 December 2023* ***Accepted:*** *30 January 2024*

information data is planned into an inert space by the encoder, and the first data is reproduced from the idle space portrayal by the decoder. By improving both the reproduction mistake and an idle space regularization term, variational approach preparing is utilized to prepare VAEs.

Generative Adversarial Networks (GANs):

An extra popular strategy for generative demonstrating is the utilization of Generative Adversarial Networks, or GANs [3]. The generator and discriminator neural networks make up a GAN. The discriminator assesses the created samples' legitimacy in relation to real data, whereas the generator creates synthetic data samples. GANs are taught in a competitive fashion, where the discriminator improves its ability to discern between real and false data, and the generator aims to produce realistic data to trick the discriminator [4].

B. Benefits and Mechanisms of Generative AI for Fraud Detection

With a number of benefits and tools that more conventional approaches did not have, Gen AI has emerged as a crucial weapon for identifying and stopping fraudulent conduct. Adaptive learning, data augmentation, and complex algorithms are hallmarks of its application, which result in notable increases in precision and decreases in false positives.

Machine learning algorithms can detect up to 94% of fraudulent transactions in real-time, according to the Arab Monetary Fund Study (AMF), greatly lowering financial losses for businesses.

C. Machine Learning-Based Fraud Detection

User behavior can exhibit subtle and obfuscating occurrences that cannot be clearly recognized as fraudulent transactions due to a lack of conclusive evidence. Algorithms that can handle large datasets with several variables can be developed thanks to machine learning, which also helps uncover hidden patterns between operator behavior and the probability of fraudulent activity. Because machine learning structures can analyze data quickly and provide automation for data handling, they are more sophisticated than traditional rule-based structures. For example, because they minimize the amount of verification procedures required, intelligent algorithms are great in behavior analytics [5].

Due to their need to recognize and report any suspicious online behavior, financial institutions are more involved in keeping an eye out for fraudulent activity. A situation involving the programming of a machine prototype on a dataset including transactions that had been carried out illegally is explained by a study by Villalobos. The rule-based model prototype aided in the discovery of the covert relationships that appeared in the transactions and illicit activity. These machine learning techniques decrease the workload for smaller financial institutions that carry out fraud surveillance. According to the article's recommended remedy, 99.6% of money laundering transactions were stopped, and reported transactions decreased from 30% to 1% [6-7]. Calculations, the groundwork of machine learning (ML), become more proficient with expanding data sizes. The machine learning model turns out to be progressively compelling and can recognize the likenesses and contrasts between different ways of behaving as how much data increments. The machine learning models frameworks become progressively successful at ordering data sets into the suitable classifications as it uncovers more differentiations among fraudulent and certifiable activities. As the size of the client database develops, machine learning models become progressively versatile. While machine learning calculations offer many benefits for financial associations looking to identify fraud, there are a few impediments that limit how generally they can be u [8]tilized. For example, one of the greatest drawbacks of machine learning is the colossal volume of data expected for the models to be exact. Although there should be enough of data points to pinpoint the valid causal relationships in smaller financial organizations, the data threshold is reasonable. Furthermore, behaviors, activities, and actions are what machine learning algorithms are based on. The fraud detection process could be imprecise if the model ignores obvious linkages, such as a card used in numerous accounts.

D. Global Trends in Credit Card Fraud and Fraud Prevention

Globally, the rising use of online and electronic payment methods is made possible by rapidly developing technologies, which has also led to an increase in fraud losses. According to Fig 1 [9] of the Nilson research, which displays global trends in fraud losses in billions of dollars from 2014 to 2028, credit card fraud is expected to reach $43.47 billion by 2028, or 6.4 cents per $100 of total volume. Additionally, the data reveals that fraud cost $33.45 billion globally in 2022, with nearly 41% of those losses occurring in the United States alone. Over the next ten years, an additional half a trillion dollars in losses are anticipated globally.

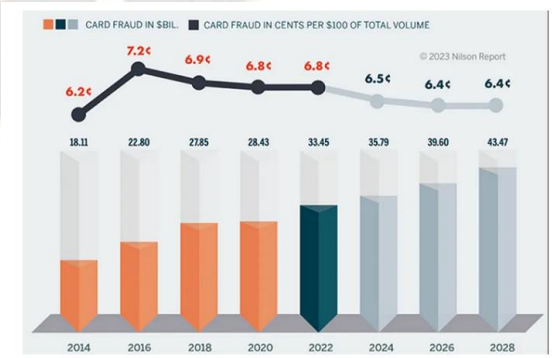

Fig 1 Worldwide Credit Card Fraud Trends (Nilson, 2023)

Hackers and fraudsters from all over the world are always taking advantage of holes and weaknesses in the security protocols that are now in place in a variety of businesses, with the financial sector being responsible for billions of dollars'

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 12 Issue: 2
***Article Received:** 25 November 2023 **Revised:** 12 December 2023 **Accepted:** 30 January 2024*

worth of fraud and lost income [9]. The expanded assumption among merchants and clients for Visa exchanges to happen quickly, both on the web and coming up, has prompted the improvement of constant irregularity detection frameworks that utilization factual, artificial intelligence (simulated intelligence), and machine learning (ML) models to screen exchanges progressively. This requirement also results from the shortcomings of the security protocols now in use to avoid fraud [10].

E. Utilizing GANs for Financial Fraud Detection

An inventive method that makes use of deep learning capabilities to detect fraudulent activity in financial transactions is the use of Generative Adversarial Networks (GANs) for financial fraud detection. Here's how to go about it:

- **Data Generation:** Generator and discriminator neural networks, which make up GANs, are trained concurrently. While the discriminator learns to discern between real and phony data, the generator learns to create synthetic data that is similar to the real data. The generator can be trained to generate synthetic data in the context of financial transactions that looks like real transactions.
- **Anomaly Detection:** The generator's synthetic data can be compared to the actual data after the GAN has been trained. Inconsistencies between the simulated and actual data may point to possible irregularities or fraudulent transactions. This is due to the fact that fraudulent transactions frequently display patterns—such as unexpected sums, unusual timing, or unusual transaction sequences—that set them apart from legitimate transactions.
- **Feature Extraction:** Relevant features from financial data can also be extracted using GANs. The generator gains the ability to recognize the underlying patterns and structure of the data by being trained to generate synthetic data. The accuracy of conventional fraud detection algorithms or models can then be increased by utilizing these learned attributes.
- **Adversarial Training:** In order to enhance their performance, the generator and discriminator in a GAN can be taught in an adversarial fashion. Through the process of adversarial training, the model learns to adapt and produce synthetic data that is harder and harder for the discriminator to discern from genuine data, making it more resilient to new and growing forms of fraud.
- **Continuous Learning:** Since fraudsters are always coming up with new ways to avoid detection, financial fraud is a threat that is always changing. Systems that enable continuous learning, in which the model is updated with fresh data on a regular basis to keep ahead of evolving fraud tendencies, can incorporate GANs.

F. Research Objectives

- To evaluate the trained models' performance using measures such as area under the curve (AUC) based on a confusion matrix analysis, accuracy, precision, recall, and F1-score.
- To maximize false positives and false negatives while refining the model's efficacy in identifying fraudulent transactions.
- To recommend methods and preventive measures, such as enhanced transaction security and fraud detection systems, to lessen the likelihood of credit card fraud.

II. LITERATURE REVIEW

Zheng et al. [11] framed the new worldview for financial and monetary exploration utilizing GAI, enveloping the objectives of the review, the models, and the logical data. It also examines the perspective and underlying effects that this paradigm brings to the area. The study goes into further detail about each of the possible five scenarios, which include financial fraud detection, policy analysis, extreme scenario analysis, portfolio management, and economic and financial projection. This paper proposes a new research paradigm with GAI that can offer important insights for a thorough comprehension of innovation and transformation in this field.

Brühl [12] outlined of the standards of generative man-made intelligence (GAI), including machine learning, critical players in this creating field, future applications, and current monetary potential. It closes by looking at the present status of the impending European Artificial Intelligence Act (artificial intelligence Act), which will mark a huge defining moment in the making of a legitimate framework for solid man-made intelligence in Europe and then some.

Gupta [13] analyzed the connections between man-made intelligence, data examination, and other arising advancements to work on understanding of fraud anticipation. This study looks at the advantages of integrating data investigation and machine learning into artificial intelligence frameworks for fraud detection and avoidance across the area. To get data on the capability of artificial intelligence, data, and examination in fraud counteraction, the review approach contained an exhaustive investigation of the collection of current writing as well as various contextual investigations. An extensive variety of legislative, corporate, and scholastic sources are utilized to give the exploration concentrate on its worldwide reach. The distributions and headways from 2019 to 2023 are analyzed in this review. The exploration philosophy incorporated an intensive survey of the writing. Scholastic diaries, gathering procedures, and official distributions made up this evaluation. To survey the data, track down shared characteristics, and examine the benefits and drawbacks of computer based intelligence fraud insurance arrangements, a subjective investigation was completed. Contextual investigations made a more careful examination of true applications conceivable and worked on understanding of man-made intelligence driven fraud counteraction methods. Significant ends from the review were made on the various uses of artificial intelligence, data, and examination in halting fraudulent action. The benefits and weaknesses of different Artificial Intelligence (artificial

intelligence) methods were appeared through an examination of generative computer based intelligence for social designing, charge card investigation, and digital actual security for Internet of Things (IoT) networks. The study's conclusions suggest that analytics, data, and AI may change a system's defenses against fraud. The aforementioned findings highlight the importance of adaptable fraud prevention techniques. To stay ahead of evolving fraud schemes, you need to be collaborating constantly, using cutting-edge technologies, and conducting continuing investigations. The importance of upcoming problems and directions is emphasized in the paper's conclusion.

Huang et al. [14] examined how artificial intelligence and finance are interacting, with a particular emphasis on the rise of intelligent investment advisers, or Robo-advisers (RAs). These RAs provide users with individualized asset management investment plans by using artificial intelligence algorithms and strong computer models. Interestingly, Wealthfront is cited as a well-known platform in this industry that provides automated investment management services with the goal of maximizing returns on investments. The study examines how users' adoption of intelligent advisers is influenced by their prior investment performance, taking into account things like recent investment performance and past defaults. It highlights the necessity of persistent use to effectively capitalize on the advantages of intelligent advisers, revealing that frequent alterations to their use may impede long-term investment objectives. In addition, the study highlights how important it is to have transparent, user-friendly interface designs, and customized financial services in order to build user confidence and improve the design of intelligent advisers.

Rane [15] explored the complex function and difficulties that these generative AI systems must overcome in the accounting and banking industries. ChatGPT simplifies client communications in the financial industry by providing tailored financial guidance, supporting investment plans, and enabling in-the-moment market research. It improves algorithmic trading, risk management, and fraud detection by processing large amounts of data quickly. These artificial intelligence models in bookkeeping lower working costs and human blunder via robotizing data section, order, and report creation. They likewise further develop scientific bookkeeping strategies and help with consistence exercises, which include sticking to evolving regulation. Nevertheless, there are challenges in coordinating ChatGPT and related artificial intelligence in bookkeeping and money. One-sided dynamic calculations, data protection, and security present moral problems. A major challenge still lies in making sure that AI systems adhere to industry norms and laws while protecting the privacy and integrity of sensitive financial data.

III. RESEARCH METHODOLOGY

The dataset, characteristics, analytical technique, research methodology, and procedures for conducting the study are all described in this part. The dataset is utilized to identify credit card fraud.

A. Research Method

An experimental design is a method for gathering information in a regulated setting in order to pinpoint and comprehend the causes of differences in variables. Depending on their research topic, aims, constraints, and resources, researchers can choose from a broad variety of design possibilities. The resources at your disposal, your goals, and your limits will all influence your final decision. Therefore, in order to execute a quality study, a number of procedures in the experimental design process must be followed, including data collection, preprocessing, dataset splitting, training and testing, model selection, and model evaluation.

Collecting Data

It's important to comprehend the problem at hand before using machine learning to solve it. Collect the information based on the problem statement. We have two options for dataset creation for machine learning: starting from scratch or using an existing one. Numerous systems facilitate the gathering of data to address issues related to machine learning.

Data Pre-processing:

Data collection is the next stage after data processing. The model does not produce the desired outcomes when the data are supplied to it raw or without any pre-processing. The accuracy of the model will rise if you employ methods that can communicate information in the best way possible. Use feature selection, feature extraction, and transferred learning to try to balance the dataset if it is uneven.

Testing and training:

After the dataset has been cleaned, divide it. The data can be separated using cross-validation, a train test ratio, or a train-test validation ratio. You can create an evaluation dataset and a training dataset for the model's training by separating the dataset. We keep the model from being overfit by doing this.

Selecting a Model:

After data processing, we must select a model that best fits the dataset and the current goal, such as clustering or classification. Choosing the right model is crucial to achieving the desired outcomes.

Evaluating the Model:

After the model has been trained, utilize the unused dataset to predict the results. If the predicted results are obtained using the prediction metrics, then the model is considered ready for data classification. Change the parameters, retrain the model, and repeat the procedure until the expected results are achieved.

B. Data Collection

Data Selection:

Datasets are an essential part of machine learning. The model's

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 12 Issue: 2
Article Received: *25 November 2023* ***Revised:*** *12 December 2023* ***Accepted:*** *30 January 2024*

effectiveness is dependent on the type of data that is given into it. It was not practical to generate a personal dataset because the project contains credit card information and credit card fraud detection. We used an available dataset of credit card transactions from Kaggle (Credit Card Transactions Fraud Detection Dataset, Nov 2023) for this study. A dataset instance's "is_fraud" feature tells you if it is a legitimate transaction or a scam. The is_fraud variable can have one of two values: 0 or 1. A transaction with a score of 0 is considered legal, but a score of 1 denotes fraud ((Credit Card Transactions Fraud Detection Dataset, Oct 2023).

Attribute:

This dataset contains credit card transactions from the Western United States. The dataset contains twenty-three features that give information about each transaction, such as the merchant's details, the consumer's details, the type of purchase, and if the transaction was fraudulent.

C. Data Preprocessing

Pre-processing the data was done to ensure that it was structured and of sufficient quality for training. This stage was divided into two subphases: feature engineering and missing value imputation. Every row was distinct and there were none that were repeated. The dataset also lacked null values and missing values. A few superfluous columns that depicted the customer's location were removed from the dataset. To make the data organized and easy to examine, the attribute "trans_date_trans_time," for instance, was changed to datetime format. Later, in order to extract more data from the dataset, this property is split into hour, day, month, and year categories. After that, the dataset was cleared of "trans_date_trans_time." After calculating the credit card holder's age using the dob attribute, the dob column was removed. We can determine the age range of credit card users who are the target of fraud with the help of this algorithm. In a similar vein, the dataset's variables "first," "last," and "merchant" were also removed.

D. Model Selection and Training - Proposed Algorithms

Based on the literature analysis, it has been demonstrated that the algorithms listed below are more precise and efficient than other methods at identifying credit card fraud.

Random Forest Classifier:

This method can be used for both regression and classification problems. Random forest algorithms are commonly utilized because of their versatility and ease of usage. There are numerous decision trees in the random forest, each of which is distinct from the others. Different trees are used to check different conditions or attributes. The average of all the predictions produced by the decision tree constitutes the final predictions of random forests.

K-Nearest Neighbor:

The KNN method learns significantly more slowly than other algorithms, making it a "lazy learner." KNN acquires knowledge by persistently focusing on data storage until it is supplied with input data whose class or label is meant to be predicted. The distance between the unknown tuple and the K-training set is predicted by the KNN classifier using a distance measure. Conceptually, KNN is straightforward and, thanks to its low input requirements, is yet capable of solving difficult problems.

Support Vector Machine (SVM):

A SVM finds the hyperplane that augments the edge between two classes to do grouping. The vectors or cases that characterize the hyperplane are known as the support vectors.

E. Model Evaluation

A confusion matrix and model predictive performance indicators were used to quantify and validate the model impacts in order to select the optimal model and evaluate its performance. The amounts of true positives (TP), false positives (FP), false negatives (FN), and true negatives (TN) are utilized to construct a disarray lattice. The performance measures used to assess the models were area under the curve (AUC), F1-score, exactness, accuracy, review.

True Positive (TP):

YES is the anticipated result according to the model, and YES is the actual value.

False Positive (FP):

Although the model anticipated YES, the real result was NO. Another name for it is Type-I error.

False Negative (FN):

Also known as Type-II error, this occurs when the model predicts a value of NO but the actual result is YES.

True Negative (TN):

Both the actual value and the model's forecast of NO were NO. This matrix can be used for a variety of computations, including accuracy and precision.

Accuracy:

It is a key performance evaluation statistic that measures the percentage of accurate predictions made by the classifier relative to all of its predictions. This is one way to portray it:

$$\text{Accuracy Score} = (TP + TN)/ (TP + FN + TN + FP) \quad \ldots (1)$$

Precision:

The ratio of true positive predictions to all positive predictions is referred to as positive predictive value:

$$\text{Precision Score} = \text{True Positives}/(\text{False Positives} + \text{True Positives}) \quad \ldots (2)$$

Recall:

In binary classification, recall—also referred to as sensitivity—calculates the classifier's true positive rate. The percentage of true positive predictions over all positive test occurrences is how it is calculated:

$$\text{Recall Score} = \text{True Positives} / (\text{False Negatives} + \text{True Positives}) \quad \ldots (3)$$

F1-Score:

The significance of true positive and true negative is examined by the F1-score, commonly referred to as the F-measure. It is

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 12 Issue: 2
Article Received: *25 November 2023* ***Revised:*** *12 December 2023* ***Accepted:*** *30 January 2024*

the harmonic mean of the accuracy and recall performance measures that were previously computed:

F1 Score = (2* Precision Score * Recall Score)/ (Precision Score + Recall Score)….. (4)

AUC:

The model's accuracy in categorization is determined by looking at the area under the receiver operating characteristic (ROC) curve. A higher accuracy is indicated by a larger AUC. An AUC of one indicates that the classifier is very good. AUC values less than 0.5 indicate that the model underperforms random guessing, whereas values more than 0.5 and less than 1 indicate that the model outperforms random guessing

IV. RESULTS & DISCUSSION

A. Exploratory Data Analysis (EDA)

Data is investigated and evaluated using EDA to find information about its characteristics, distribution, trends, and potential problems

#	Column	Non-Null Count	Dtype
0	trans_date_trans_time	1296675 non-null	object
1	cc_num	1296675 non-null	int64
2	merchant	1296675 non-null	object
3	category	1296675 non-null	object
4	amt	1296675 non-null	float64
5	first	1296675 non-null	object
6	last	1296675 non-null	object
7	gender	1296675 non-null	object
8	street	1296675 non-null	object
9	city	1296675 non-null	object
10	state	1296675 non-null	object
11	zip	1296675 non-null	int64
12	lat	1296675 non-null	float64
13	long	1296675 non-null	float64
14	city_pop	1296675 non-null	int64
15	job	1296675 non-null	object
16	dob	1296675 non-null	object
17	trans_num	1296675 non-null	object
18	unix_time	1296675 non-null	int64
19	merch_lat	1296675 non-null	float64
20	merch_long	1296675 non-null	float64
21	is_fraud	1296675 non-null	int64

Fig 2 Dataset

Figure 1 displays the data types and the analysis of the categorical variables that were initially present in the dataset with a non-null count.

TABLE 1
DEMOGRAPHIC PROFILE

Variable	Class	Count (%)
Gender	Male	60%
	Female	40%
Age	<30	25%
	30-45	45%
	46-60	20%
	61-75	7%
	>75	3%
Shopping category	Health Fitness	15%
	Entertainment	10%
	Grocery	20%
	Personal care	25%
	Gas Transport	10%
	Travel	18%
	Others	2%

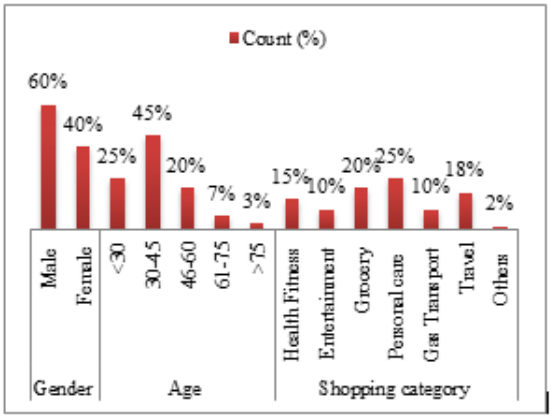

Fig 3: Graphical Presentation of Demographic Profile

The dispersion of respondents by age, orientation, and favored shopping classification is displayed in Table 1. 40% of respondents are female, while 60% of respondents are male. When it comes to age groupings, the bulk of respondents (45%) are between the ages of 30-45. Those under 30 come in second at 25%, those between the ages of 46 and 60 at 20%, those between the ages of 61 and 75 at 7%, and those above 75 at 3%. When it comes to shopping categories, personal care goods are preferred by the largest amount of respondents—25%. This is followed by grocery shopping (20%), travel (18%), health and fitness (15%), entertainment (10%), petrol and transportation (10%), and miscellaneous categories (2%). This information sheds light on the studied population's shopping habits and demographics.

A. *Comparison of the performance of the models.*

KNN, SVM, and Random Forest were utilized to build the training models.

Random Forest Classifier

In Table 2, the RF performance matrix is shown.

TABLE 2
PERFORMANCE MATRICES OF RANDOM FOREST

Metrics	**Results**
Accuracy	0.9963
Precision	0.9953
Recall	0.9945
F1-Score	0.9980

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 12 Issue: 2
Article Received: *25 November 2023* ***Revised:*** *12 December 2023* ***Accepted:*** *30 January 2024*

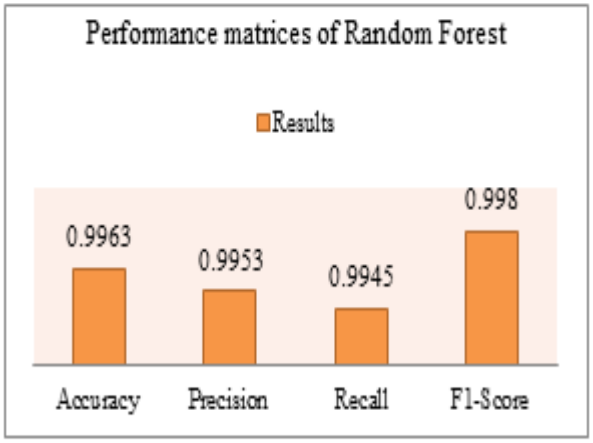

Fig 4: Performance matrices of Random Forest

Table 2 illustrates the outstanding performance of the Random Forest Classifier. With a 99.63% accuracy rate, the model was able to predict the class labels for most occurrences with precision. With relatively few false positives, the precision of 99.53% indicates that the model's positive predictions were extremely accurate. With a 99.45% recall rate, the model was able to correctly identify most of the positive cases. With a F1-score of 99.80%, which strikes a harmony among review and accuracy, the Random Forest Classifier performed splendidly in general in this order challenge.

K- nearest neighbor

In Table 3, the KNN performance matrix is provided.

TABLE 3
PERFORMANCE MATRICES OF KNN

Metrics	Results
Accuracy	0.9945
Precision	0.9987
Recall	0.9989
F1-Score	0.9980

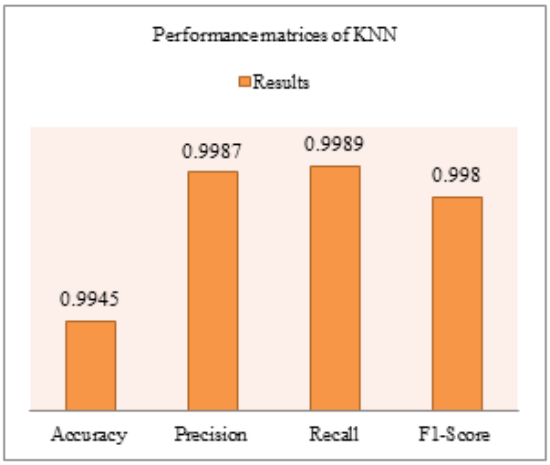

Fig 5: Performance matrices of KNN

As Table 4 illustrates, the K-Nearest Neighbors (KNN) model performed admirably. With an accuracy of 99.45%, the model was able to predict the class labels for most occurrences with precision. With extremely few false positives, the model's positive predictions appear to have been quite accurate, as indicated by its precision of 99.87%. With a 99.89% recall rate, the model was able to correctly identify the great majority of positive cases. The F1-score, which measures how well recall and precision are balanced, was 99.80%, indicating that the KNN model performed well overall in this classification job.

Support Vector Machine

The performance matrix of SVM is given in Table 4.

TABLE 4
PERFORMANCE MATRICES OF SVM

Metrics	Results
Accuracy	0.9958
Precision	0.9974
Recall	1
F1-Score	0.9961

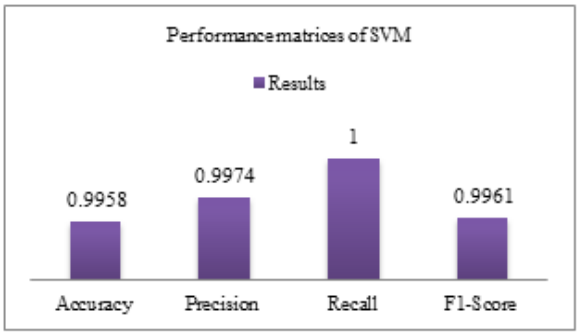

Fig 6 Performance matrices of SVM

Table 4 illustrates the robust performance of the Support Vector Machine (SVM) model. The model successfully predicted the class labels for the majority of occurrences, as seen by its 99.58% accuracy rate. With relatively few false positives, the precision of 99.74% indicates that the model's positive predictions were very accurate. The model successfully recognized every positive event, as indicated by the 100% recall score. The SVM model performed very well overall in this classification job, as evidenced by the F1-score of 99.61%, which balances precision and recall.

B. Receiver operating characteristic curve (ROC)

For the classifiers in this study, the Receiver Operating Characteristic (ROC) curve was considered an extra performance metric. One widely used performance indicator for evaluating the effectiveness of binary classifiers is the ROC curve. The responsiveness of the classifier is plotted against the false certain rate in the ROC curve. One can figure the false certain rate by deducting the particularity of the characterization model from one.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 12 Issue: 2
Article Received: *25 November 2023* ***Revised:*** *12 December 2023* ***Accepted:*** *30 January 2024*

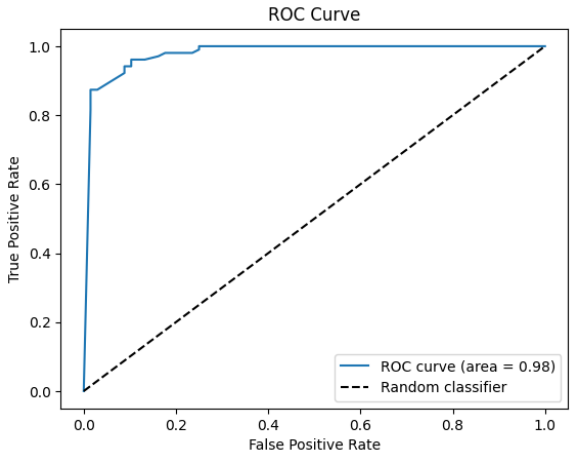

Fig 7 ROC of Random Forest model

Fig 7 shows the expectations of the random forest model's Receiver Operating Characteristic (ROC) curve and the compromise between true certain and false sure rates. An area under the curve (AUC) estimation is likewise remembered for the figure. The area under the curve for the Random Forest characterization model is 0.98. For the dataset used in this review, Random Forest and KNN perform better compared to SVM.

V. CONCLUSION

The use of machine learning methods, for example, Support Vector Machine, Random Forest, and K-Nearest Neighbor, to the detection of Visa fraud was analyzed in this work. Credit card transaction details were the dataset that was sourced from Kaggle and used for both the models' training and evaluation. Based on the results, it was evident that all three models had high recall, accuracy, precision, and F1-score. With an accuracy of 99.63%, the Random Forest model was just behind the SVM model, which had the best accuracy of 99.58%. The results showed that the Random Forest and K-Nearest Neighbor models performed better on the dataset than the Support Vector Machine model. To improve these algorithms' ability to identify credit card fraud, more tinkering and optimization will be necessary. Ensuing examinations might focus on researching bunch procedures or profound learning methodologies to improve the accuracy and adequacy of fraud detection frameworks. Including real-time data processing skills could further improve the algorithms' capacity to quickly identify fraudulent transactions

REFERENCES

[1] M. Alghofaili, "The Role of Artificial Intelligence and Machine Learning in Detecting Financial Fraud," Journal of Digital Forensics, Security and Law, vol. 15, no. 4, pp. 43-58, 2020.

[2] J. Villalobos and J. Silva, "Machine Learning in Fraud Detection," International Journal of Advanced Computer Science and Applications, vol. 8, no. 9, pp. 97-103, 2017.

[3] Y. Luo, Y. Xiao, L. Cheng, G. Peng, and D. Yao, "Deep learning-based anomaly detection in cyber-physical systems: Progress and opportunities," ACM Computing Surveys (CSUR), vol. 54, no. 5, pp. 1-36, 2021.

[4] E. Berlinski, J. Morales, and S. Sponem, "Artificial imaginaries: Generative AIs as an advanced form of capitalism," Critical Perspectives on Accounting, vol. 99, p. 102723, 2024.

[5] H. B. Shah, "Comparing Machine Learning Algorithms For Credit Card Fraud Detection.

[6] "Credit Card Transactions Fraud Detection Dataset," Accessed: Nov. 28, 2023. [Online]. Available: https://www.kaggle.com/datasets/kartik2112/fraud-detection.

[7] "Credit Card Transactions Fraud Detection Dataset," Accessed: Oct. 16, 2023. [Online]. Available: https://www.kaggle.com/datasets/kartik2112/fraud-detection.

[8] P. Bhambri and A. Khang, "The Human-Machine Nexus With Art-Making Generative AIs," in Making Art With Generative AI Tools, IGI Global, 2024, pp. 73-85.

[9] Nilson, "Fraud Statistics," Nilson Report Issue 1254, Dec. 2023. [Online]. Available: https://nilsonreport.com/newsletters/1254.

[10] Arab Monetary Fund (AMF), "Study on the Role of Artificial Intelligence in Detecting and Preventing Financial Fraud.

[11] X. Zheng, J. Li, M. Lu, and F. Y. Wang, "New Paradigm for Economic and Financial Research With Generative AI: Impact and Perspective," IEEE Transactions on Computational Social Systems, 2024.

[12] V. Brühl, "Generative Artificial Intelligence–Foundations, Use Cases and Economic Potential," Intereconomics, vol. 59, no. 1, pp. 5-9, 2024

[13] P. Gupta, "Securing Tomorrow: The Intersection of AI, Data, and Analytics in Fraud Prevention," Asian Journal of Research in Computer Science, vol. 17, no. 3, pp. 75-92, 2024.

[14] Z. Huang, C. Che, H. Zheng, and C. Li, "Research on Generative Artificial Intelligence for Virtual Financial Robo-Advisor," Academic Journal of Science and Technology, vol. 10, no. 1, pp. 74-80, 2024.

[15] N. Rane, "Role and Challenges of ChatGPT and Similar Generative Artificial Intelligence in Finance and Accounting," SSRN 4603206, 2023. [Online]. Available: https://ssrn.com/abstract=4603206.

[16] G. Colvin, "The pandemic may be the greatest environment for business fraud in decades," 12 November 2020. [Online]. Available: https://fortune.com/2020/11/12/pandemic-corporate-fraudscams/. [Accessed 25 November 2023].

[17] A. Littman, The Fraud Triangle: Fraudulent Executives, Complicit Auditors and Intolerable Public Injury, CreateSpace Independent Publishing Platform, 2011.

[18] M. Ciobanu, "Why understanding your fraud false-positive rate is key to growing your business," 12 March 2020. [Online]. Available: https://thepaypers.com/thought-leader-insights/whyunderstanding-your-fraud-false-positive-rate-is-key-to-growing-your-business--1241130. [Accessed 25 November 2023].

[19] C. Mullen, "Card industry's fraud-fighting efforts pay off: Nilson Report," 5 January 2023. [Online]. Available: https://www.paymentsdive.com/news/card-industry-fraud-fighting-effortspay-off-nilson-report-credit-debit/639675/. [Accessed 20 November 2023].

[20] Pascual, Marchini, and Miller, " Identity Fraud: Securing the Connected Life.," 2017. [Online]. Available: https://javelinstrategy.com/research/2017-identity-fraud-securing-connected-life. [Accessed 23 November 2023].

[21] S. Sando, "Consumer Preference Drives Shift in Authentication.," 2021. [Online]. Available: https://www.javelinstrategy.com/coverage-area/consumer-preference-drives-shift-authentication. [Accessed 22 November 2023].

[22] Villalobos, Miguel Agustín, and Eliud Silva., "A Statistical and Machine Learning Model to Detect Money Laundering: An application," 2017. [Online]. Available: http://hddavii.eventos.cimat.mx/sites/hddavii/files/Miguel_Villalobos.pdf (accessed on 21 June 2018). [Accessed 23 November 2023].

[23] Awoyemi, et al, "Credit card fraud detection using machine learning techniques: A comparative analysis," 2017 international conference on computing networking and informatics (ICCNI). IEEE, 2017.

[24] Herland, Matthew, Khoshgoftaar, Bauder, "Big data fraud detection using multiple medicare data sources.," Journal of Big Data, vol. 5, no. 1, pp. 1-21, 2018.

[25] V. Ayyadevara, Pro machine learning algorithms, Berkeley, CA, USA: Apress, 2018