International Journal on Computational Engineering

Journal Homepage www.comien.org/index.php/comien

Modelling Credit Card Fraud Data using Machine Learning Algorithms

Tayo P. Ogundunmade [a,*], Adedayo A. Adepoju [a]

[a] *Department of Statistics, University of Ibadan, Oduduwa Road, Ibadan, 200005, Nigeria*

Corresponding author: [*]ogundunmadetayo@yahoo.com

***Abstract*— Credit card fraud refers to the unauthorized use of a credit card, often for illegitimate or illegal transactions. In recent years, it has emerged as a major concern, causing billions of dollars in losses annually, according to statistics. Moreover, the problem is becoming increasingly complex with the development of new fraud techniques. This alarming statistic underscores the urgent need for robust statistical analysis to understand, prevent, and combat fraudulent activities using credit card fraud data generated by European credit cardholders. Therefore, employing machine learning models with high accuracy ratings and optimal performance is essential for detecting credit card fraud. This study uses supervised machine learning techniques; decision trees (DT), Random Forests (RF), Artificial Neural Networks (ANN), Naïve Bayes (NB), and Logistic Regression (LR) to detect credit card fraud. The findings reveal that while identity theft, skimming, counterfeit cards, mail intercept fraud, and lost or stolen cards remain prevalent, there is a notable increase in other forms of fraud due to evolving techniques. Among the machine learning models evaluated, the Decision Tree method demonstrated the highest accuracy, outperforming the others.**

***Keywords*— Credit card; fraud; machine learning algorithms; accuracy; model performance.**

Manuscript received 15 Feb. 2024; revised 29 Mar. 2024; accepted 22 Apr. 2024. Date of publication 30 June. 2024.

I. INTRODUCTION

Since the inception of payment systems, there have consistently been individuals seeking unauthorized means to access financial information. In the contemporary era, this issue has escalated significantly, particularly with the ease of conducting online transactions by simply entering credit card details [1]. Fraud has existed for as long as humans have, and it can take countless forms [2]. Consequently, fraud detection has become increasingly critical and pressing for many establishments/organisations. Minimizing the misuse and preventing future occurrences of such fraudulent operations is achievable through implementing necessary preventive measures and thoroughly studying the behaviour underlying such practices. In situations where fraud cannot be stopped, it needs to be identified as soon as possible and dealt with appropriately. The rapid technological advancements of the contemporary era have spurred a high demand for more innovative payment methods.

In the past, cash and cheques were the predominant payment methods. However, with the rapid rise of credit cards, many individuals now consider using credit cards exclusively for online transactions as preferable to cash. The widespread accessibility and convenience of digital transactions owe much to the prevalence of credit cards. Credit card fraud, encompassing the unauthorized use of payment cards, including debit and credit cards, remains a significant concern in this context [3], [4].

When a lawful client uses his credit card to make a payment to an account that is run by an illegal or unauthorized person, credit card fraud can be authorized even when the account holder does not authorize the payment and a third party completes the transaction. There are various methods for a compromise to happen, and it usually happens without the cardholder's knowledge. It is often difficult to determine the exact point of penetration because a fraudster may retain the credentials to a compromised account for months before any theft occurs. Cardholders can report the loss or stolen cards quickly. The cardholder sometimes may be unaware of unauthorized use until payment alerts or statements are received. The credit card can be used for illicit purchases until the cardholder contacts the issuing bank and the bank blocks the account.

The majority of banks provide free, round-the-clock phone numbers for early reporting. Nevertheless, a fraudster may continue to use the card to make unauthorized purchases until it is revoked. Credit card fraud encompasses various forms and requires multifaceted approaches for prevention. Within large computing communities like machine learning and data science, automated solutions offer promising avenues for addressing this issue effectively. Class imbalance poses a unique challenge in addressing the issue, particularly from a learning perspective. The number of legal transactions greatly exceeds the number of fraudulent ones, and the patterns of transactions often change over time. Machine learning algorithms are used to examine all approved transactions and identify any that appear suspicious. Experts examine these records and communicate with the cardholders to verify if the transaction was authentic or fraudulent [5].

In Nigeria, the frequency of credit card scams has sharply expanded along with the growth of internet commerce. An enormous amount of money is lost annually as a result of credit card fraud. In the UK, losses from unauthorized financial theft involving credit cards and Internet banking were £844.8 million in 2018. In 2018, credit card firms and banks prevented £1.66 billion in fraudulent transactions. This means that for every £3 attempted fraud, £2 is prevented. According to Bloomberg research, fraud losses on credit, debit, and prepaid cards issued globally in 2015 were $21.84 billion. Bloomberg estimates that this might expand at a 45 percent annual rate by 2020 [6]. It is interesting to note that card owners are least affected by credit card fraud because their obligation is only for the actual transactions. The interests of cardholders are safeguarded by current laws, rules, and insurance programs in the majority of nations. The businesses are most impacted, though, they typically lack the proof (such a digital signature) needed to refute the cardholders' allegations that their card information has been abused. All losses resulting from chargebacks, product delivery costs, card issuer fees, and administrative expenses are ultimately borne by merchants.

Repeated instances of fraud involving the same company can lead to client attrition, and cessation of payment acceptance by credit card issuer banks, thereby tarnishing the company's reputation and goodwill [7].

Credit cards have made payments much easier for people in the marketing industry. Now, transactions can be completed easily. Due to the widespread use of this technology, credit cards are now used for practically all transactions. This development has made a more secure method of handling these transactions necessary. Scammers have been using Technology to defraud individuals of their money, highlighting the urgent need for prevention measures. The transactions performed with a stolen or duplicated credit card are reported as fraudulent. If these fraudulent activities are not stopped or identified quickly, significant losses may ensue. As credit card use increases, a rising number of financial losses brought on by credit card fraud rises. Consequently, many stories described large losses in different countries. The only information required for an online fraud transaction is the card's data, which are made remotely. At the time of purchase, neither a physical signature nor a PIN nor a card imprint are necessary. Despite the implementation of preventive measures such as CHIP and PIN, online frauds, including mail orders and internet fraud, continue to escalate in both frequency and magnitude. These measures have only been able to effectively curb fraudulent operations involving simple theft, counterfeit cards, and non-receipt incidents (NRI). Approximately half of all credit card fraud-related losses in 2008 were attributable to online frauds, according to Visa reports on European nations. The fraudsters typically steal the entire available limit on the card as soon as they get it. Based on statistical data, individuals typically complete this task in four to five transactions. As a result, even though bank authorities stipulate that standard predictive modelling performance metrics are crucial for solving the fraud detection problem, a performance criterion that evaluates the amount of loss that can be prevented on cards whose transactions are found to be fraudulent takes precedence. Stated differently, there is greater value in detecting fraud on a larger available limit card than on a smaller available limit card.

The introduction of new banking technology has led to an increase in online transactions which has also raised the number of Cybercrimes, especially credit card fraud. This study aims to detect credit card fraud by comparing various machine learning techniques.

The detection of credit card fraud involves several challenges, some of which include figuring out which learning strategy (supervised or unsupervised), which algorithms (decision trees, logistic regression, etc.) to use, which features to use, and how to address the issue of class imbalance [3]. The credit card fraud databases are rare and highly biased; the best features (variables) for the models are chosen; the right metric is used to evaluate the efficacy of strategies on skewed credit card fraud data; and the profile of fraudulent conduct is dynamic, meaning that fraudulent transactions often resemble legitimate ones.

To improve the accuracy of credit card fraud detection, Carcillo et al. [4] suggested a novel approach that blends supervised and unsupervised learning algorithms. A subclass of machine learning known as "unsupervised learning" focuses on training models with unlabelled datasets to identify underlying patterns and data discrepancies. By combining the best features of both methods, this creative technique raises the accuracy and potency of credit card fraud detection systems. They evaluated the combined approach's performance against conventional supervised and unsupervised techniques. The accuracy of the model predictions across various thresholds was evaluated by the authors using evaluation measures, including area under the curve (AUC) values, F1 score, precision, recall, and receiver operating characteristic (ROC) curve analysis. The trials conducted by them showed that their combined technique achieved superior precision, recall, and F1 scores than either supervised or unsupervised methods alone [8]. When compared to independent unsupervised methods, the Autoencoder-based anomaly detection step successfully decreased false positives. By utilizing the supervised classifier model's generalization skills, the combined technique proved very successful in identifying fraudulent patterns that had not been seen before, according to the authors. This flexibility is essential for reducing new fraud risks. Compared to current techniques for identifying credit card fraud, their method has several benefits. Their approach overcomes the drawbacks of each particular method by mixing supervised and unsupervised

procedures, which leads to increased accuracy and fewer false positives.

In their thorough analysis of statistical fraud detection strategies, Bolton and Hand [9] emphasized the value of statistical and data analytics methodologies in detecting fraudulent activity. They emphasized how important it is to study customer behaviour and transaction patterns in order to create efficient fraud detection systems.

Böhme and Moore [10] examined the economics of cybersecurity, illuminating the guiding ideas and available avenues for legislation to prevent credit card theft. Their research emphasized the negative economic effects of fraud and suggested frameworks for policy to reduce related risks. Artificial intelligence and machine learning have become powerful instruments for detecting fraud. Recurrent neural networks (RNNs) were investigated by Bose and Chen [11] as a potential tool for identifying credit card fraud. Their research demonstrated how well RNNs do sequential data analysis, which helps to improve fraud detection models.

Canova, et al [12] looked into the difficulties that unbalanced data presents when attempting to identify credit card fraud. To tackle the imbalance problem, they used convolutional neural networks (CNNs), which improved the fraud detection models' accuracy in situations where class distributions were skewed.

In the past decade, the internet has experienced exponential growth, leading to the widespread adoption and proliferation of services such as online bill payment, tap-and-pay, and e-commerce. Consequently, there has been a surge in fraudulent activity by criminals targeting credit card transactions. Tokenization and credit card data encryption are two of the many methods available to secure credit card transactions. While these techniques work well in most situations, they do not completely guard against fraudulent credit card transactions. A kind of artificial intelligence known as machine learning (ML) enables computers to gain better predicting skills without being specifically trained to do so by using data from past experiences [13]. Thus, it is essential to put in place a credit card fraud detection technique that works and can shield people from losing money. One of the primary challenges in utilizing machine learning techniques for credit card fraud detection is the inability to replicate the majority of published work. This issue stems from the extreme confidentiality surrounding credit card transactions. As a result, anonymised attributes are present in the datasets used to create machine-learning models for credit card fraud detection. Furthermore, credit card fraud detection is a tough process because of the continuously changing nature and patterns of fraudulent transactions. Therefore, the extremely skewed nature of credit card fraud datasets and current machine learning models for credit card fraud detection have poor detection accuracy. Consequently, it is critical to create machine learning models that function at their best and have a high accuracy rating for identifying credit card fraud.

The ML techniques utilized by Adepoju and his team on the skewed credit card fraud data included Logistic Regression, Support Vector Machine (SVM), Naive Bayes, and KNN (K-Nearest Neighbour). According to Adepoju et al. [14], the final model (SVM) Support Vector Machine scored 97.53%, whereas the other models including Logistic Regression scored 99.07%, Naïve Bayes scored 95.98%, and K-nearest Neighbour scored 96.91%.

Safa and Ganga examined the effectiveness of Naïve Bayes, K-nearest Neighbour, and Logistic Regression on a highly distorted credit card dataset. They implemented their findings in Python and used assessment to determine which approach performed the best. Their model's accuracy for Naïve Bayes is 83%, for Logistic Regression it is 97.69%, and for K-Nearest Neighbour it is 54.86% [15].

Saheed and colleagues' paper focuses on using genetic algorithms (GAs) as a feature selection method for credit card fraud detection [16]. The researchers employed ML techniques such as Naïve Bayes (NB), Random Forest (RF), and Support Vector Machine (SVM) for feature selection, which involves splitting the data into first-priority features and second-priority features. Based on Saheed et al. [16], the highest accuracy was achieved by Random Forest with 96.40%, followed by SVM with 96.3% and Naïve Bayes with 94.3%.

Three separate machine learning techniques are used in Itoo and his group's work: logistic regression, Naïve Bayes, and K-nearest neighbours. Itoo and his team used Python to implement their job, which included recording the work and comparative analysis. The accuracy of logistic regression is 91.2%, that of Naïve Bayes is 85.4%, while the accuracy of K-nearest neighbour comes in last with 66.9% [17].

The present study is therefore based on the utilization of supervised machine learning methods, specifically decision trees (DT), Random Forests (RF), Artificial Neural Networks (ANN), Naïve Bayes (NB), and Logistic Regression (LR), for the detection of credit card fraud. Large datasets are used to train and evaluate machine learning systems. This work makes use of a credit card fraud dataset that was created by credit cardholders throughout Europe. These datasets frequently contain a variety of characteristics that may negatively affect the classifiers' performance during training. We implement a feature selection algorithm based on the Genetic Algorithm (GA) employing the RF approach in its fitness function to solve the problem of a high feature dimension space. Given its resilience to noisy data, capacity to handle numerous input variables, and ability to automatically address missing values, the Random Forest (RF) approach is integrated into the Genetic Algorithm (GA) fitness function.

II. Materials and Methods

A. Data for The Study

The dataset used originates from https://www.kaggle.com and consists of credit card fraud data generated by European credit cardholders. This extensive dataset comprises 172,792 rows and 30 feature columns, comprising various variables such as the country name, credit card fraud methods, year, incidence of fraud, and more.

B. Methods

1) Artificial Neural Network: Artificial Neural Networks (ANNs) are machine learning models that draw inspiration from the architecture of animal brains. ANNs are another type of supervised learning. They are made up of networked artificial neurons that mimic real neurons. Connected neurons send signals to neurons, and the weights

of these connections can be changed. Signals are layered and go through several changes from input to output. Processing samples with known inputs and outcomes and creating associations are all part of training. Through supervised learning, the network gains new skills by modifying weights in response to variations between target and expected outputs. In image recognition, neural networks, for example, autonomously identify objects such as cats by generating distinctive features from labelled instances without prior knowledge.

Neural networks learn tasks without explicit rules. To detect credit card fraud, artificial neural networks, or ANNs, are essential. To effectively train a network, they first gather and pre-process a transaction dataset and carefully split it into training, validation, and testing sets. During training, the input, hidden, and output layers of the neural network design constantly modify connection weights to distinguish between authentic and fraudulent transactions. A binary cross-entropy loss function is used to evaluate the correctness of the model, and validation and hyper-parameter tuning are used to further improve it. After the model is adjusted to perfection, it is used for real-world fraud detection and its adaptable features help it become more accurate over time. The ANN model performs a nonlinear functional mapping from the input observations (yt-1, yt-2,, yt-p) to the output value (yt). i.e.,

$$y_t = a_0 + \sum_{j=1}^{q} a_j f(w_{oj} + \sum_{i=1}^{p} w_q y_{t-i}) + \varepsilon_t \qquad (1)$$

Where, $a_j (j = 0, 1, 2, ..., q)$ is a bias on the j^{th} unit,
w_{ij} ($i = 0, 1, 2, ..., p$; j = 0, 1, 2, ... To detect, q) is the connection weights between layers of the model,
$f(...)$ is the transfer function of the hidden layer,
p is the number of input nodes and,
q is the number of hidden nodes.

The logistic sigmoid function, which is characterized by what? was the activity function that the hidden layer's neurons used.:

$$f(x) = \frac{1}{1+ e^{-x}} \qquad (2)$$

This function is a member of the sigmoid function class and has the advantages of being continuous, monotonically increasing and differentiable at all locations.

2) Logistic Regression: AWith regard to a categorical dependent variable, logistic regression serves as a predictive modelling technique. When identifying between fraudulent and legitimate transactions, a scenario where the outcome is binary, logistic regression proves to be useful in the identification of credit card fraud. Logistic regression algorithms can forecast the likelihood of fraud based on computed odds ratios by examining transaction specifics such as money, location, time, and previous data. These models, which make use of the logistic function, offer probabilities that help assess the possibility of fraudulent events. Credit card fraud detection systems can discover trends that point to fraudulent behaviour since logistic regression can adjust to non-linear connections between features, making it a strong option. The logistic model is hereby given as

$$P(Y = 1) = \frac{\exp X^T\theta}{1 + \exp X^T\theta} \qquad (3)$$

where,
$Y = 1$ if the respondent has a fraudulent transaction,
X is the vector of the independent variables,
θ are the unknown parameters to be estimated from the data.

3) Decision Tree: For both categorical and quantitative explanatory factors, decision tree models can be utilized. Finding non-linear correlations between independent and dependent variables is a fantastic use of decision trees. Divide the dataset into manageable chunks according to the decision tree model's guiding principle. Values of all information points associated with the issue articulation are plotted on subsets of the dataset. A decision tree with decision and leaf hubs is produced when the data is divided using this approach. In circumstances where enough change in the dataset does not machine learning experts choose this model. In this scenario, decision nodes stand in for tests of characteristics such as transaction amounts and locations, while branches denote results that direct the detection of possible fraudulent activity. At the end nodes, specific classifications that differentiate between safe and questionable transactions are provided. Decision trees may learn from past data on their own, adjust to new fraud trends, and make evaluations in real time thanks to this methodical technique. Moreover, decision trees improve fraud detection systems' interpretability and transparency.

4) Naïve Bayes: The idea of belief revision is that, whenever new information becomes available, it may require updating of prior beliefs. Bayes' theorem expresses how a subjective degree of belief should rationally change to account for the availability of related evidence. The main objective is to demonstrate how Bayes theorem can be used to identify falsified credit card transactions given a set of training data. The objective of utilizing the Bayes rule is based on its ability to accurately predict the value of a selected discrete class variable given a set of attributes.

5) Support Vector Machines (SVM): Support Vector Machine (SVM) is a machine learning approach that can be applied to problems involving regression or classification. It is usually applied to categorization difficulties, though. Each data item is plotted as a point in n-dimensional space (where n is the number of features you have) using the SVM algorithm. The value of each feature is represented by a specific position. Next, we carry out the classification process by identifying the hyper-plane that effectively separates the two classes.

Finding a hyperplane in an N-dimensional space (N - the number of features) that clearly classifies the data points is the aim of the SVM algorithm. SVM's primary objective is to classify the datasets in order to identify the largest marginal hyperplane. This can be accomplished in two steps:

i. SVM will first iteratively create hyper-planes that best separate the classes.
ii. After that, it will select the appropriate hyperplane to divide the classes.

6) K-Neighbouring Network (KNN): The KNN algorithm, sometimes referred to as KNN, is a supervised learning classifier that is non-parametric and relies on proximity to classify or predict how a single data point will be grouped. Although, it can be applied to classification or

regression issues, it is usually employed as a classification algorithm, based on the idea that comparable points can be located next to each other.

The KNN algorithm's objective is to locate a query point's closest neighbours so that a class label can be applied to it. KNN needs a few things in order to accomplish this:

$$d(x, y) = \sqrt{\sum_1^n (y_i - x_i)^2} \quad (4)$$

(i) Determine your distance metrics: Euclidean distance
(ii) Compute KNN: defining k

In the KNN method, the number of neighbours that will be examined to ascertain the categorization of a particular query point is defined by the k parameter. The instance will be placed in the same class as its single nearest neighbour, for instance, if k=1.

C. Performance Measures

The performance measure used for comparing the methods is the accuracy. The mathematical expression is given as:

$$Accuracy = \frac{TN + TP}{TN + TP + FN + FP} \quad (6)$$

In this case, TP stands for True Positive, FN for False Negative, and FP for False Positive. These are derived from the confusion matrix and utilized in the computation of the performance standards. A performance metric used in machine learning classification issues is the confusion matrix. The true positive, true negative, false positive, and false negative are displayed in a 2 by 2 table. The confusion matrix table uses the square of the number of classes as the size when addressing multi-class categorization.

III. Result and Discussion

This section shows the result of the analyses. It includes the exploratory data analysis of the data and the model performance for the machine learning models considered in this work.

A. Performance Measures

This section presents an analysis of the distribution of percentage occurrences of credit card frauds across major countries worldwide. Additionally, it explores the frequency of various types of credit card frauds and examines the amount lost to and recovered from credit card fraud in Nigeria. Table 1 shows the countries arranged according to the percentage occurrences of credit card fraud in the world.

TABLE I
COUNTRIES WITH THE PERCENTAGE OCCURRENCE OF CREDIT CARD FRAUD

Countries	Percentage Occurrence
Mexico	44%
United States	42%
India	37%
The UAE	36%
China	36%
United Kingdom	34%
Brazil	33%
Australia	31%
Singapore	26%
South Africa	25%
Canada	25%
Italy	24%
France	20%
Indonesia	18%
Germany	13%
The Netherlands	12%
Sweden	12%

TABLE II
DISTRIBUTION OF CREDIT CARD FRAUD OCCURRENCES BY THE METHODS USED FOR PERPETRATING THE FRAUD.

Methods	Percentage
Lost or stolen card	45%
Identity theft	15%
Skimming (cloning)	14%
Counterfeit card	12%
Mail intercept fraud	6%
Other	8%

Table 2 details the various methods of credit card fraud employed globally. The table further outlines the global landscape of credit card fraud methodologies. The prevalence of the lost or stolen card method as a fraud drives its high global percentage because of the widespread impact of stolen or lost cards across diverse countries. This contributes to their dominant position in global fraud statistics. Although lost or stolen cards continue to be the most common method of credit card fraud, there is a troubling rise in identity theft, skimming, counterfeit cards, mail intercept fraud, and other emerging tactics as fraudsters adapt their strategies. Table 3 shows the amounts lost to and recovered from fraud annually over a nine-year period (2014-2023). Figure 1 shows the bar-plot for both the amount lost and the fund recovered.

TABLE III
AMOUNTS LOST TO FRAUD ANNUALLY

Years	Amount loss (in billions)	Fund recovered (In billions)
2014	3.9	2.1
2015	4.3	2.3
2016	3.4	1.8
2017	4.7	2.8
2018	4.2	1.5
2019	4.5	2.1
2020	5.1	1.1
2021	4.1	1.2
2022	3.9	1.8
2023	3.5	2.3

B. Machine Learning Algorithms Results

This section employs machine learning models to analyse credit card fraud data and assesses their predictive performance in identifying occurrences of fraud using the accuracy measure. Table 4 shows the accuracy values (in percent) of the six machine learning models considered.

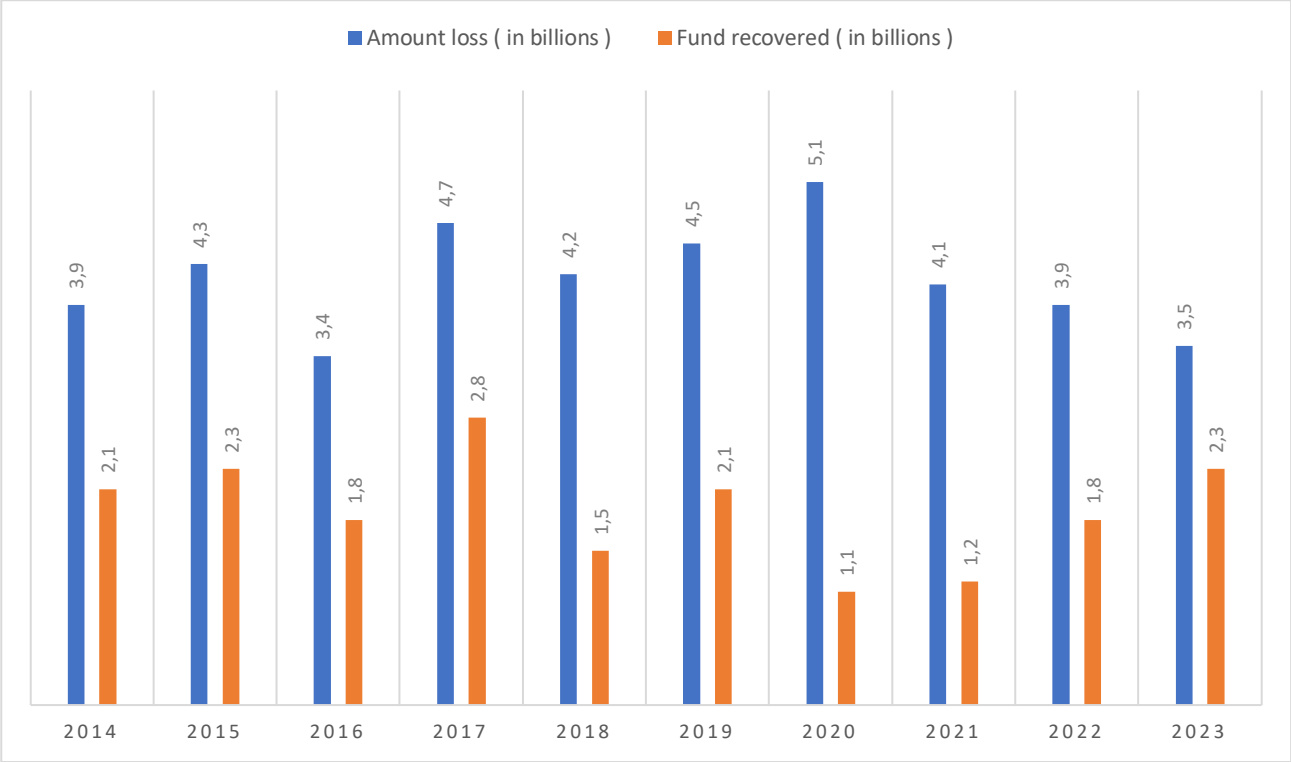

Fig. 1 Bar-plot for amount lost and Fund Recovered

TABLE IV
ALGORITHMS USED IN DETECTING CREDIT CARD FRAUD AND THEIR ACCURACY

Algorithm	Accuracy (%)
SVM	94.7
KNN	87
Logistic Regression	90
Naïve Bayes	94
Decision Tree	94.9
ANN	93.73

Table 4 shows that the Decision Tree algorithm outperformed other models with the highest accuracy value of 94.9% closely followed by the SVM while the KNN records the lowest performance.

IV. Conclusion

While lost or stolen cards remain the most common type of fraud, emerging fraudulent techniques such as identity theft, skimming, counterfeit cards, mail intercept fraud, and other forms are contributing to alarming increases in fraudulent activities. This trend highlights the evolving nature of fraud and the need for robust countermeasures to address emerging threats in the financial sector. This work uses machine learning models that operate optimally and have good accuracy ratings to detect credit card fraud. Using supervised machine learning, decision trees (DT), Random Forests (RF), Artificial Neural Networks (ANN), Naive Bayes (NB), and Logistic Regression (LR) are methods for identifying credit card fraud. The Decision Tree technique fared better than the other machine learning models, with the best accuracy value of 94.9%. This work therefore recommends the use of machine learning models for the detection and prediction of credit card fraud activities. For further work, deep learning models can be studied from other works [18-26].

References

[1] Chuprina, R. (2021, February 25). Credit Card Fraud Detection: Top ML Solutions in 2021.

[2] Lakshmi, M. G., Mettu, V., and Hameed, K. (2020). Credit Card Fraud Detection Using Random Forest. Journal of Information and Computational Science, 10(3). Retrieved from www/joics.org.

[3] Shakya, R. (2018). Application of Machine Learning Techniques in Credit Card Fraud Detection. University of Nevada, Las Vegas, Department of Computer Science.

[4] Carcillo, F., Le Borgne, Y. A., Caelen, O., Kessaci, Y., Oblé, F., & Bontempi, G. (2021). Combining unsupervised and supervised learning in credit card fraud detection. Information Sciences, 557(xxxx),317–331, doi:10.1016/j.ins.2019.05.042.

[5] Chung J, Lee K. Credit Card Fraud Detection: An Improved Strategy for High Recall Using KNN, LDA, and Linear Regression. Sensors (Basel). 2023 Sep 10;23(18):7788. doi: 10.3390/s23187788. PMID: 37765845; PMCID: PMC10535547

[6] Vaishnavi Nath Dornadula, S Geetha, Credit. Card Fraud Detection using Machine Learning Algorithms, Procedia Computer Science, Volume 165, 2019, Pages 631-641, ISSN 1877-0509, doi:10.1016/j.procs.2020.01.057.

[7] Andrea Dal Pozzolo, Giacomo Boracchi, Olivier Caelen, Gianluca Bontempi(2017). Credit Card Fraud Detection: A Realistic Modeling and a Novel Learning Strategy, September 2017, IEEE Transactions on Neural Networks and Learning Systems PP(99):1-14. doi: 10.1109/TNNLS.2017.2736643.

[8] Hosmer, D. W., and Lemeshow, S. (2000). Applied Logistic Regression (2nd ed.). New York: John Wiley and Sons., Menard, S. (2002). Applied logistic regression analysis (2nd ed.). Thousand Oaks, CA: Sage Publications.

[9] Bolton, R. J., and Hand, D. J. (2002). "Statistical fraud detection: A review." Statistical Science, 17, 235-249.

[10] Böhme, R., and Moore, T. (2016). "The economics of cybersecurity: Principles and policy options." CESifo Economic Studies, 62

[11] Bose, I., and Chen, X. (2016). "Detecting payment card fraud using recurrent neural networks." Decision Support Systems, 87, 27-36.

[12] Canova, G., Lui, J., and Qin, B. (2016). "Credit card fraud detection for imbalanced data using convolutional neural networks." Expert Systems with Applications, 62, 90-102.

[13] Emmanuel, I., Sun, Y. and Wang, Z. A machine learning-based credit risk prediction engine system using a stacked classifier and a filter-based feature selection method. J Big Data 11, 23 (2024). doi:10.1186/s40537-024-00882-0.

[14] Adepoju, O., Wosowei, J., lawte, S., & Jaiman, H. (2019). Comparative evaluation of credit card fraud detection using machine learning techniques. 2019 Global Conference for Advancement in Technology (GCAT). doi:10.1109/gcat47503.2019.8978372

[15] Safa, M. U., & Ganga, R. M. (2019). Credit Card Fraud Detection Using Machine Learning. International Journal of Research in Engineering, Science and Management, 2(11).

[16] Saheed, Y. K., Hambali, M. A., Arowolo, M. O., & Olasupo, Y. A. (2020). Application of ga feature selection on Naive Bayes, random forest and SVM for credit card fraud detection. 2020 International Conference on Decision Aid Sciences and Application (DASA). doi:10.1109/dasa51403.2020.9317228

[17] Itoo, F., Meenakshi, & Singh, S. (2020). Comparison and analysis of logistic regression, Naïve Bayes and Knn Machine Learning Algorithms for credit card fraud detection. International Journal of Information Technology, 13(4), 1503–1511. doi:10.1007/s41870-020-00430-y.

[18] Ogundunmade TP, Abidoye M, Olunfunbi OM. Modelling Residential Housing Rent Price Using Machine Learning Models. Mod Econ Manag, 2023; 2: 14.

[19] T. P. Ogundunmade, A. O. Daniel, and A. M. Awwal, "Modelling Infant Mortality Rate using Time Series Models", Int. J. Data. Science., vol. 4, no. 2, pp. 107-115, Dec. 2023.

[20] Ogundunmade, T.P., Adepoju, A.A. (2023). Predicting the Nature of Terrorist Attacks in Nigeria Using Bayesian Neural Network Model. In: Awe, O.O., Vance, E.A. (eds) Sustainable Statistical and Data Science Methods and Practices. STEAM-H: Science, Technology, Engineering, Agriculture, Mathematics & Health. Springer, Cham. doi:10.1007/978-3-031-41352-0_14.

[21] Ayansola OA, Ogundunmade TP, Adedamola AO. Modelling Willingness to Pay of Electricity Supply Using Machine Learning Approach. Mod Econ Manag, 2022; 1: 9. doi:10.53964/mem.2022009

[22] Ogundunmade TP, Adepoju AA, Allam A. Stock price forecasting: Machine learning models with K-fold and repeated cross validation approaches. Mod Econ Manag, 2022; 1: 2. doi:10.53964/mem.2022001

[23] Ogundunmade TP, Adepoju AA. The performance of artificial neural network using heterogeneous transfer functions. Int J Data Sci, 2021; 2: 92-103. doi: 10.18517/ ijods.2.2.92-103.2021

[24] Adepoju AA, Ogundunmade TP. Economic Growth and its Determinant: A cross country Evidence. Statistical Trans New Ser, 2019; 20: 69-84. doi: 10.21307/ stattrans-2019-015

[25] Ogundunmade TP, Adepoju AA, Allam A. Predicting crude oil price in Nigeria with machine learning models. Mod Econ Manag, 2022; 1: 4. doi: 10.53964/mem.2022004

[26] Ogundunmade TP, Adepoju AA. Modelling Liquefied Petroleum Gas Prices in Nigeria Using Time Series Machine Learning Models. Mod Econ Manag, 2022; 1: 5. doi:10.53964/mem.2022005