IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 6, June 2025

DOI: 10.17148/IJARCCE.2025.14601

# Fraud Detection and Prevention in Financial Transactions using Hybrid Machine Learning.

**Doris Chinedu Asogwa[1], Ebele Grace Onyedinma[2], Richard Orah Ojochegbe[3], Gloria Nkiru Anibogu[4], Emmanuel Chibuogu Asogwa[5]**

Department of Computer Science, Nnamdi Azikiwe University, Awka, Anambra State, Nigeria[1]

Department of Computer Science, Nnamdi Azikiwe University, Awka, Anambra State, Nigeria[2]

New Horizons Solution Centre, University Business Training Unit, Lagos, Nigeria[3]

Department of Computer Science, Nnamdi Azikiwe University, Awka, Anambra State, Nigeria[4]

Department of Computer Science, Nnamdi Azikiwe University, Awka, Nigeria[5]

**Abstract**: Recent analysis have identified a significant rise in transactional fraud, where bad actors seek to deceive individuals or firms into unauthorized financial actions. Traditional fraud detection systems frequently struggle to effectively identify such activities, leading to financial damages and security breaches. Addressing this problem requires employing sophisticated machine learning techniques specially designed to detect transactional fraud. This study presents a novel fraud detection method called "filter", aimed at uncovering misleading transactional behaviours. By employing tailored features to reveal fraudulent patterns and activities, our filter achieves a remarkable accuracy of over 99.01% in distinguishing fraudulent transactions from legitimate ones, while maintaining a low false positive rate. Our approach was evaluated with a dataset comprising 746 instances of fraudulent transactions and 4822 instances of legitimate transactions. The results underscore the superior performance of our filter compared to existing methods, particularly in accurately detecting fraudulent transactions. Moreover, our hybrid NB-ANN model achieves the highest accuracy of 99.01%, outperforming both Naïve Bayes (98.57%) and Artificial Neural Network (98.12%) techniques. This highlights the effectiveness of the hybrid method in boosting detection accuracy for transactional fraud. Implementing our filter and leveraging the hybrid NB-ANN model, organizations can greatly improve their ability to detect and prevent fraudulent activities thereby protecting their financial assets and maintaining customer trust.

**Keywords:** Machine learning, Predictive model, transaction fraud, dataset, hybrid NB-ANN, Filter, Legitimate.

## I. INTRODUCTION

Traditionally, banks offered solely in-person services until 1996, when Citibank and Wells Fargo Bank introduced the first Internet banking application in the United States [1]. This innovation led to a surge in online credit card usage, marking the beginning of a decade of rapid digital transformation. E-commerce, online payment systems, remote work, online banking, and social networking became commonplace [2]. Consequently, cybercriminals intensified their efforts to exploit online transactions, leading to increased instances of fraud [3].

Recent advancements in digital technology, particularly in cashless transactions, have revolutionized daily financial management. Many payment systems have shifted from physical to digital platforms, enhancing productivity and competitiveness [4][5]. Internet banking and online transactions have provided customers with convenient avenues to conduct financial activities remotely, predominantly through credit cards.

A credit card, as defined by [6], contains personal information and is issued by financial institutions to facilitate global purchases. Credit card fraud, the unauthorized use of another person's card for financial gain, poses significant financial risks [7][8]. The shift to online transactions has simplified fraudulent activities, as transactions can be completed without the physical presence of the card [9]. Furthermore, credit card introductions have influenced monetary policies and business strategies [10].

The Bank of Ghana reported a staggering increase in credit card fraud losses from GH¢ 1.26 million ($250,000) in 2019 to GH¢ 8.20 million ($1.46 million) in 2020 [11]. Digital transactions have witnessed the highest surge in fraud cases [11]. Perpetrators employ various tactics, including VPN connections and physical theft, making apprehension challenging [13]. Consequently, compliance and risk management services have turned to AI and machine learning for fraud detection [12].

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 6, June 2025
DOI: 10.17148/IJARCCE.2025.14601

Various machine learning models, including Decision Trees, Logistic Regression, and Random Forest, have been employed for fraud detection due to the classification and prediction nature of the problem [14].

In Nigeria, a similar trend is observed, with the rapid adoption of digital banking services and the consequent increase in cybercrime threats [15]. Nigerian banks are facing challenges in combating fraud, particularly in online transactions, necessitating the adoption of advanced technological solutions. Therefore, the hybridization of fraud detection techniques through machine learning models, as presented in this study, holds relevance for Nigerian financial institutions as they strive to enhance security measures and protect their customers' assets.

## II. RELATED WORKS

This section reviews the concept of relevant work on transaction fraud dataset prediction using a machine learning model. Logistic regression is a technique used to predict a binary outcome variable. Unlike some other methods, it doesn't assume that explanatory variables follow a normal distribution or are correlated [16]. This model is particularly useful when the outcome variable is qualitative and the explanatory variables can be numerical or categorical. Logistic regression has been widely employed in detecting financial bankruptcies, among other applications.

Decision trees, on the other hand, are non-linear classification techniques that recursively divide a sample into smaller subgroups based on explanatory variables [17]. At each node of the tree, the algorithm selects the variable that, according to a predetermined criterion, is most strongly correlated with the outcome variable. While decision trees are versatile and can handle both quantitative and qualitative data, they are prone to overfitting, especially when applied to the entire dataset. However, they find applications in various fields, such as filtering transaction fraud and predicting susceptibility to viruses in medicine.

Random forests, proposed by [18], introduce an additional level of randomness to bagging, a method for improving model accuracy. They employ various bootstrap samples of the data for each tree's construction and randomly select subsets of explanatory variables at each node for splitting [18]. While random forests can provide robust predictions and measure feature importance, they may exhibit bias towards attributes with numerous levels. They find applications in diverse areas such as bioinformatics, video segmentation, and image classification.

Credit card fraud encompasses various categories, including bankruptcy fraud, counterfeit fraud, application fraud, and behavioral fraud [20]. Different machine learning methods, including Logistic Regression, Naive Bayes, Random Forest, and others, have been utilized for fraud detection in various jurisdictions [21]. These methods often rely on feature importance techniques to select the most relevant features for the model. Hybrid models, such as those combining Ada Boost and majority voting strategies, have also been proposed for fraud detection [22]. Additionally, random forest models have been developed to identify behavioral characteristics of fraudulent transactions, but facing challenges such as imbalanced data [23].

In practical applications, machine learning algorithms like Random Forest, XGBoost, and Decision Tree have shown promising results in predicting credit card fraud, with high AUC values [24]. These algorithms enable financial institutions to model past transactions to identify fraudulent patterns and classify new transactions as genuine or fraudulent [25][26][27]. By leveraging these algorithms, financial institutions can detect fraudulent transactions efficiently and potentially prevent financial losses.

## III. MATERIALS AND METHODS

The approaches were based on the sample of the dataset collected from two data sources that were used in this research, utilizing the Fraud Transactions Dataset available on Kaggle. This dataset contains an extensive collection of transactions, each identified by a unique Transaction ID. It also includes information such as transaction amount, location, and customer ID. Additionally, the dataset includes a column labeled "Is Fraud?" which indicates whether a transaction is classified as fraudulent or not. By leveraging machine learning algorithms on this transaction data, we developed a fraud detection model capable of learning from historical patterns and predicting the likelihood of fraud in new transactions. The model considers various features, including transaction amount, location, and customer ID, to make accurate predictions. With a well-trained model, financial institutions can automate the detection of fraudulent transactions, enhancing their ability to identify and prevent fraudulent activities in real time. By analyzing patterns and anomalies present in the data, machine learning-based fraud detection systems significantly improve the security and trustworthiness of financial transactions.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 6, June 2025
DOI: 10.17148/IJARCCE.2025.14601

It is important to note that the effectiveness of the model relies on the quality and quantity of the training data, as well as the selection of the appropriate machine learning algorithm and its parameters.
Continuously refining the model and keeping it updated with new data allows businesses to stay ahead of fraudsters and safeguard themselves and their customers from financial losses. Based on this fact, the system was built with the available data set collected with other related literature reviews such as journals or articles.

Machine Learning Approach: Firstly, the sample datasets for transaction fraud detection were obtained. Before analysis, data pre-processing were performed, acknowledging that the dataset is annotated with binary labels (0 or 1), reflecting a supervised learning and binary classification framework. Python libraries were used for feature extraction to prepare the dataset for binary classification analysis. During the system development phase, the dataset were resampled, and divided into training and testing sets, utilizing scikit-learn and TensorFlow for analysis.

The model was developed using the Python programming language in conjunction with the Flash web framework to fulfill all outlined requirements. The proposed algorithm was employed for the classification model and the analysis of structured data.

1) **Multinomial Text Representation**
i) Tokenization
i) Convert text documents into tokens (words or n-grams).
ii) Feature Extraction
i) Represent each document as a bag-of-words or bag-of-n-grams.
ii) Calculate the frequency of each term in the document.
iii) Naive Bayes Classification
(i) Class Prior Probability (P(C))
a) Calculate the probability of each class based on training data.
(ii) Likelihood (P(Term|C))
a) Estimate the probability of each term given the class.
(iii) Posterior Probability
a) Use Bayes' theorem to calculate the probability of each class given the document.

2) The following equation shows how the multinomial Naive Bayes model calculates the probability of a text document $D$ belonging to class $C$:

3) P(C | D) = \frac{P(D | C) P(C)}{P(D)}

4) where:
i) $P(C|D)$ is the probability of class $C$ given document $D$
ii) $P(D|C)$ is the probability of document $D$ given class $C$
iii) $P(C)$ is the probability of class $C$
iv) $P(D)$ is the probability of document $D$

5) The multinomial Naive Bayes model assumes that the features are independent, given the class. This means that the probability of a word appearing in a document is independent of the probability of any other word appearing in the document, given the class.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 6, June 2025

DOI: 10.17148/IJARCCE.2025.14601

1. Feedforward Neural Network:

1) **Input Layer**

a) $X$ is the input data.

2) **Hidden Layers**

$Z^{[l]} = W^{[l]}A^{[l-1]} + b^{[l]}$, where $A^{[l-1]}$ is the activation from the previous layer.
$A^{[l]} = g^{[l]}(Z^{[l]})$, where $g^{[l]}$ is the activation function.

b)

3) **Output Layer**

$Y_{\text{pred}} = g^{[L]}(Z^{[L]})$, where $L$ is the number of layers. c)

**4) Loss Function (L)**

d) Define a suitable loss function, such as cross-entropy for classification tasks.

**5) Forward Pass**

e) Calculate predictions and intermediate values using the feedforward process.

**6) Backward Pass**

f) Calculate gradients using backpropagation.

**7) Parameter Update**

g) $W^{[l]} = W^{[l]} - \alpha \frac{\partial L}{\partial W^{[l]}}$, where $\alpha$ is the learning rate.

**1. Feature Stacking**

**1) Combine Predictions**

1) Concatenate or stack the predictions from the Naive Bayes and ANN models.
2) Create a new feature matrix.

**2. Meta-Model**

1) Input to Meta-Model

1) The stacked features from the previous step.

**2) Training**

1) Train a logistic regression model or another suitable meta-model.
2) Optimize the weights ($w_i$).

**1) Combine Predictions**

1) Use the trained meta-model to combine predictions from Naive Bayes and ANN.

$Y_{\text{ensemble}} = \sigma(w_1 \cdot \text{Naive Bayes Prediction} + w_2 \cdot \text{ANN Prediction} + b)$, where $\sigma$ is the logistic sigmoid function.

2) Define an overall objective function for the entire stack method.

3) It may involve the combination of the Naive Bayes likelihood, ANN loss, and meta-model loss.

## IV. SYSTEM DESIGN

A hybrid supervised learning model was used for training the algorithm with labels as to which class it belongs. Using the labeled data, the algorithm learns the relationship between the feature sets and the output, and hence it then classifies the unlabeled data from the learned relationship. Figure 1.0 shows the steps for fraud prediction and prevention.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 6, June 2025

DOI: 10.17148/IJARCCE.2025.14601

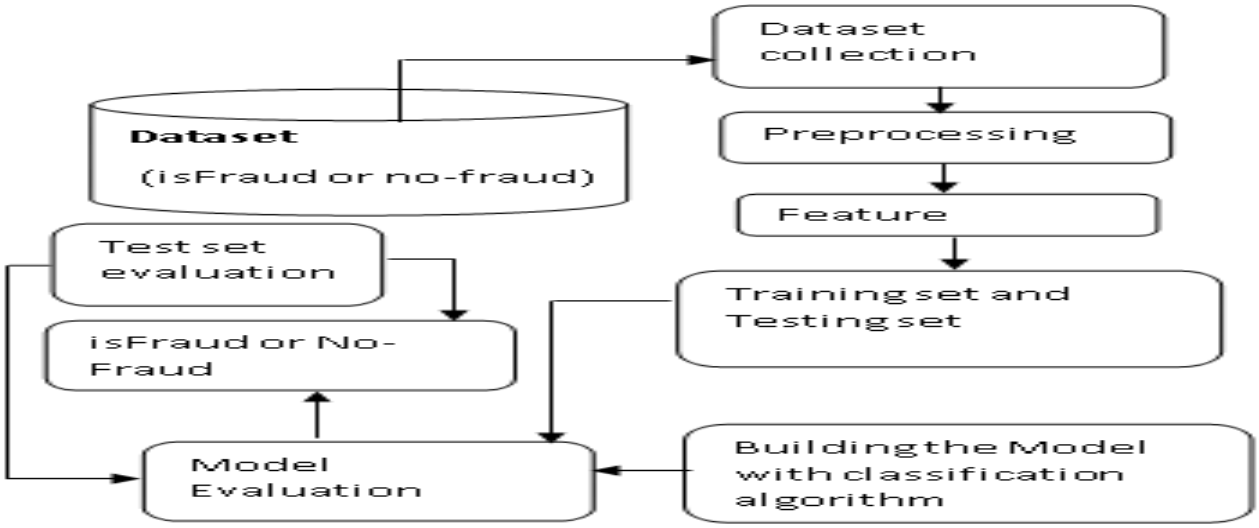



Figure 1.0: Steps for fraud prediction and prevention

**Pre-Processing**

In this step, complete geometric correction and filtering is done. The preprocessing uses the output of the classifier to take the required action to improve the performance.

**Dataset Description**

The dataset provides a wealth of features crucial for analyzing and detecting fraudulent activities. These features offer valuable insights for machine learning algorithms to discern patterns indicative of fraudulent behavior, effectively distinguishing them from genuine transactions.

It's worth highlighting that the selection of features for fraud detection may differ based on the dataset characteristics and organizational needs. Moreover, employing feature engineering techniques can enhance the model's accuracy by extracting pertinent information or crafting new features from existing ones.

Through the strategic utilization of these features and the deployment of sophisticated machine learning algorithms, organizations can construct resilient fraud detection systems. These systems not only accurately identify fraudulent transactions but also proactively prevent them, ensuring robust security measures against financial malfeasance.

**Experimental Set-Up**

The application was implemented using the open-source machine learning tool Jupyter Notebook, the Python Flask framework, and the Python programming language, supported by a Python IDE and a machine learning classification model. The subsequent subsection focuses deeper into the dataset's content, the preprocessing steps applied to the dataset, and the execution of binary class classification. The classification tasks were performed using Naïve Bayes (NB), Artificial Neural Network (ANN), and a hybrid of NB-ANN for an integrated transaction fraud prediction and prevention system. Additionally, the program's development extended to web development tools, incorporating Object-Oriented Analysis Design and Modeling (OOAD) principles.

**Classification of Outputs**

The output of the expected result is classified into different categories accordingly namely No Fraud or is Fraud.

## V. SYSTEM IMPLEMENTATION AND RESULTS

The main objectives of the system design are:

i) To use an unstructured dataset collected via online resources and clean it up for developing a hybrid Machine Learning Algorithm for detecting the two target values (isFraud or no-fraud).

ii) To label the dataset collected and categorize them into isFraud or No-Fraud using a feature set from the preprocessing Python library to avoid errors during the model training

iii) To Train the model for binary classification.

iv) To use hybrid-ML to evaluate the results in (iii) above.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 6, June 2025

DOI: 10.17148/IJARCCE.2025.14601

**A) Model Evaluation**

The experiment of the classification model was done in two folds, which are the sample of the dataset collected from which was used to perform prediction. The training set was used to build the model and then the test set for predicting the result with the unknown class label as well as to predict a new class label with their respective class. Below is a model evaluation of Naïve Bayes. Figure 2.0 shows the confusion matrix vs ROC curve for Naïve Bayes

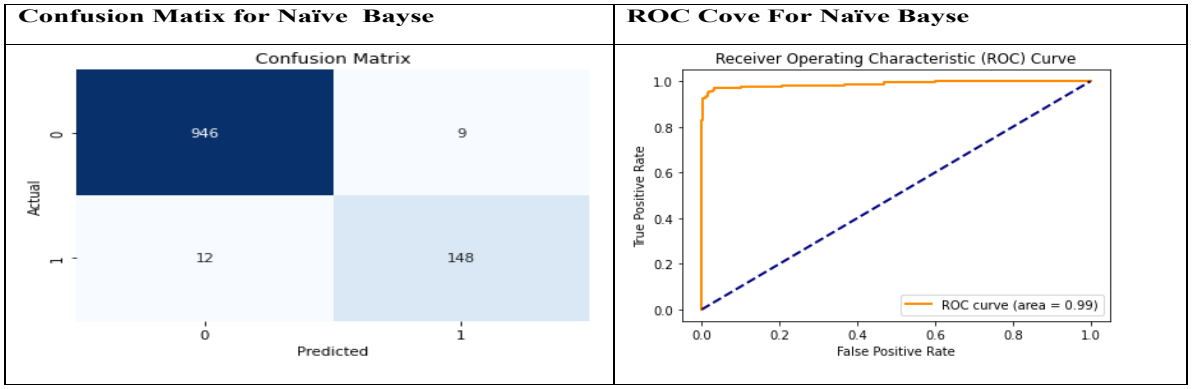



Figure 2:0 Confusion Matix for Naïve Bayse Vs ROC Cove For Naïve Bayse

**Confusion Matrix for Naïve Bayes**

1) The confusion matrix is created using the confusion_matrix function from scikit-learn.
2. The matrix is displayed as a heatmap using Seaborn and matplotlib. pyplot.
3. It provides insights into the classifier's performance by showing the counts of true positive, true negative, false positive, and false negative predictions.

**ROC Curve for Naïve Bayes**

1)The Receiver Operating Characteristic (ROC) curve is constructed to assess the classifier's performance across various threshold levels.
2. The ROC curve is plotted using the roc_curve and auc functions from scikit-learn.
3. The area under the ROC curve (AUC) is calculated, providing a single metric to evaluate the classifier's overall performance.
4. The plot visually represents the trade-off between the true positive rate and the false positive rate.

Both evaluations, the Confusion Matrix and the ROC Curve are essential tools for understanding the performance of a classifier. The Confusion Matrix offers a detailed breakdown of predictions, while the ROC Curve provides a graphical representation of the classifier's ability to discriminate between classes at different threshold levels. Together, they provide a comprehensive assessment of the Naïve Bayes classifier's accuracy performance.

Table 1.0: Details by categories of a classification model

| Class | | precision | recall | f1-score | support |
|---|---|---|---|---|---|
| **NoFraud** | **0** | 0.99 | 0.99 | 0.99 | 955 |
| **isFraud** | **1** | 0.94 | 0.93 | 0.93 | 160 |

Table 1.0 provides a summary of the classification model and demonstrates excellent precision, recall, and F1-score for the NoFraud class, with a precision of 0 and a recall of 0.99, indicating accurate predictions. In contrast, for the isFraud class, the model shows slightly lower performance but still achieves a reasonable balance between precision (1) and recall (0.94). The F1 score for isFrud is 0.93. The support values of 955 for NoFraud and 160 for isFraud provide insight into the distribution of instances in each class. Based on this concept, the model performs well in distinguishing between the NoFraud and isFraud classes, particularly excelling in accurately predicting instances of the NoFraud class.
Figure 2.1 shows the confusion matrix vs the ROC of the ANN.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 6, June 2025

DOI: 10.17148/IJARCCE.2025.14601

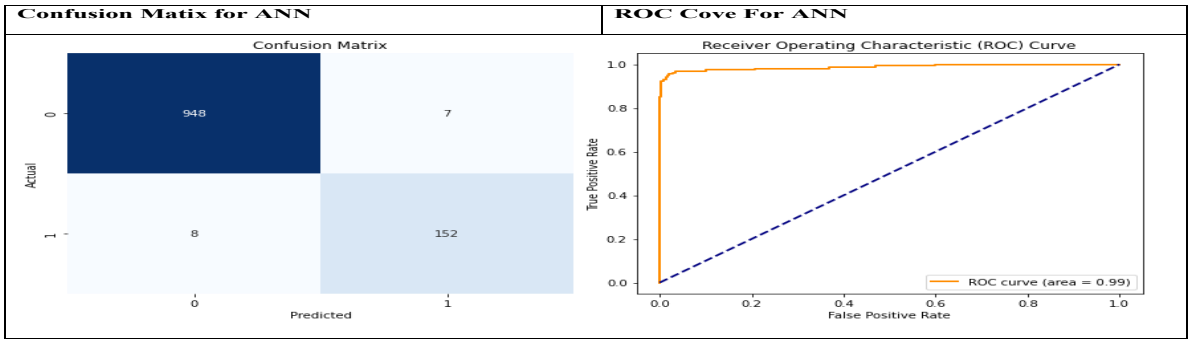



Figure 2:1 Artificial Neural Network (ANN) using both the Confusion Matrix and the ROC Curve

**Confusion Matrix for ANN**

a). The Confusion Matrix provides an overview of the model's classification performance.
b). The matrix indicates the counts of true positive, true negative, false positive, and false negative predictions.
c) . High counts in the diagonal elements (true positives and true negatives) suggest accurate predictions.

**ROC Curve for ANN**

a). The ROC Curve evaluates the ANN's ability to discriminate between classes across various threshold levels.
b). The curve illustrates the trade-off between the true positive rate and the false positive rate.
b). The Area Under the Curve (AUC) summarizes the overall performance, with higher AUC values indicating better discrimination.
A high true positive rate and true negative rate, as depicted in the Confusion Matrix, suggest that the ANN is making accurate predictions for both positive and negative instances. The ROC Curve provides additional insights into the model's discriminatory power, with a higher AUC indicating superior performance in distinguishing between classes.
The model was analyzed on both the Confusion Matrix and ROC Curve, one can gain a comprehensive understanding of the ANN's classification performance, balancing accuracy, and discriminatory capability.

Table 2.0: Details by categories of a classification model

| Class | | precision | recall | f1-score | support |
|---|---|---|---|---|---|
| **NoFraud** | **0** | 0.99 | 0.99 | 0.99 | 955 |
| **isFraud** | **1** | 0.96 | 0.95 | 0.95 | 160 |

Table 2.0 provides the classification report for the ANN model indicating exceptional performance, particularly for the NoFraud class, with precision, recall, and f1-score all at 0.99. The isFraud class exhibits slightly lower but still impressive metrics, including precision (1), recall (0.96), and f1-score (0.95). The support values of 955 for NoFraud and 160 for isFraud provide insights into the distribution of instances in each class. Therefore, the ANN model demonstrates robust classification capabilities, especially in accurately predicting instances of the NoFraud class

**Hybrid Algorithm:**

Figure 2.2 shows the confusion matrix vs the ROC curve of the NB-ANN

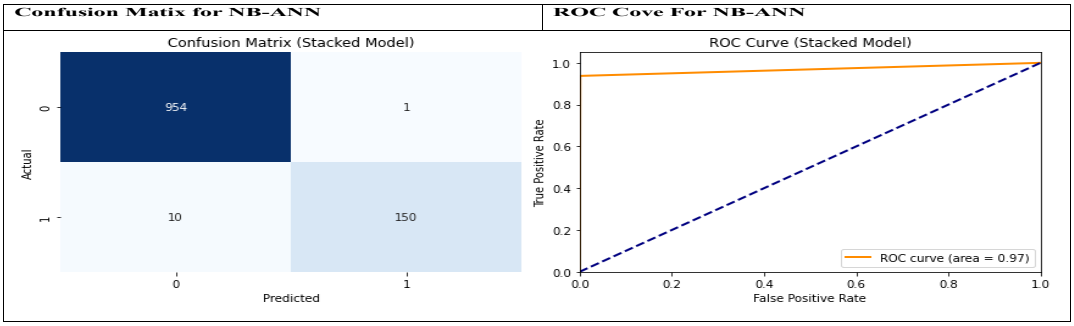



Figure 2.2 Hybrid NB-ANN

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 6, June 2025

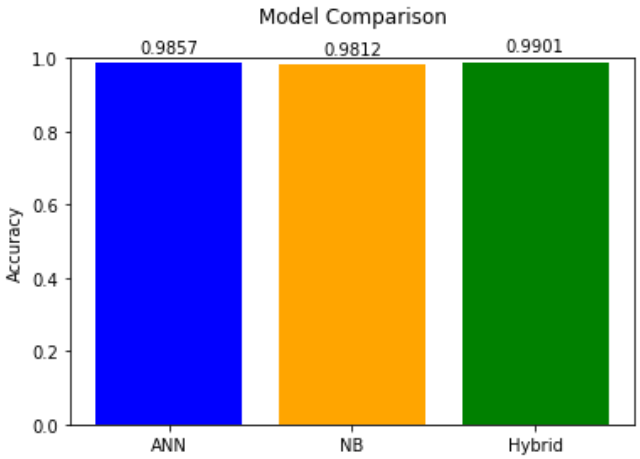
DOI: 10.17148/IJARCCE.2025.14601

The hybrid NB-ANN model demonstrates robust performance, as indicated by a high accuracy and a low loss. The Confusion Matrix illustrates accurate classification across NoFraud and isFraud categories. The ROC Curve further affirms the model's effectiveness, showcasing a strong area under the curve (AUC) and successful discrimination between true positive and false positive rates. Based on this concept, the hybrid NB-ANN model excels in isFraud detection and malware filtering.

Table 3.0: Details by categories of a classification model

| Class | | precision | recall | f1-score | support |
|---|---|---|---|---|---|
| **NoFraud** | **0** | 0.99 | 1.00 | 0.99 | 955 |
| **isFraud** | **1** | 0.96 | 0.94 | 0.96 | 160 |

Table 3.0 provides the Hybrid NB-ANN model achieves exceptional performance, with precision, recall, and F1-score metrics indicating highly accurate classification for both NoFraud and isFraud categories. The model exhibits a precision of 0.99 for NoFraud and 0.96 for isFraud, a recall of 1.00 for NoFraud and 0.94 for isFraud, and an overall F1-score of 0.99 for isFraud and 0.96 for isFraud. These metrics, combined with strong support values, highlight the model's effectiveness in transaction fraud detection and prevention. Hence figure 2.3 is a comparison graph of the models




The comparison graph Figure 2.5 reveals the accuracy performance of three models: Naïve Bayes (NB), Artificial Neural Network (ANN), and the Hybrid NB-ANN. The Hybrid NB-ANN model achieves the highest accuracy at 99.01%, outperforming both Naïve Bayes (98.57%) and Artificial Neural Network (98.12%). This underscores the effectiveness of the hybrid approach in achieving superior accuracy for transaction dataset detection and prevention

The Hybrid NB-ANN model achieved the highest accuracy at 99.01%, outperforming both Naïve Bayes (98.57%) and Artificial Neural Network (98.12%). This highlights the effectiveness of the hybrid approach in achieving superior accuracy for transaction fraud detection and prevention as shown in Table 3.0

## VI. CONCLUSION AND FUTURE WORK

The outcome of the study was derived from evaluating a transaction fraud dataset using models constructed from the training dataset, as demonstrated above which relies on a hybrid model. The research began by offering an overview of the Nofraud and Isfraud predictions. This model accurately classified a total of 5525 instances from the training set and achieved an accuracy rate of 99%. Consequently, the system demonstrated an effective learning process, successfully capturing all necessary sample data discussed in the research. Based on these findings, it is advised to employ the hybrid NB-ANN model for transactional fraud detection and classification. This approach helps protect users from being tricked into disclosing their personal credentials.

Researchers aiming to delve into fraud detection and classification or related fields should be encouraged to employ a variety of methods beyond just Naïve Bayes and ANN for optimal outcomes. Our key contribution involved introducing a correlation-based feature, customizing the system beyond merely implementing the two algorithm sets by default.

DOI: 10.17148/IJARCCE.2025.14601

This customization involved parameter updates using the Python programming language and libraries, where the algorithms underwent fine-tuning. This process demonstrated that the results achieved using the modified features surpassed those of the default algorithms, thus aiding in predicting and classifying unknown fraud cases through machine learning models.

## REFERENCES

[1] K. Yak, D. Tudeal, Internet Banking Development as A Means of Providing Efficient Financial Services in South Sudan. 2 (2011) 139–148.
[2] Madan S., Sofat S., Bansal D. Tools and Techniques for Collection and Analysis of Internet-of-Things malware: A systematic state-of-art review. J. King Saud Univ. - Comput. Inf. Sci. (2021), p. xxxx, 10.1016/j.jksuci.2021.12.016
[3] Yann-a F.C. Streaming active learning strategies for real-life credit card fraud detection: Assessment and visualization. (2018)
[4] Nath V. ScienceDirect credit card fraud detection using machine learning algorithms Credit card fraud detection using machine learning algorithms. Procedia Comput. Sci., 165 (2020), pp. 631-641, 10.1016/j.procs.2020.01.057
[5] Pencarelli T. The digital revolution in the travel and tourism industry. Inf. Technol. Tourism (2019), Article 0123456789, 10.1007/s40558-019-00160-3
[6] S.B.E. Raj, A.A. Portia, A. Sg, Analysis on Credit Card Fraud Detection Methods. (2011) 152–156.
[7] Carcillo F., Borgne, Le Y., Caelen O., Kessaci Y., Oblé F. Combining unsupervised and supervised learning in credit card fraud detection. Inform. Sci., 557 (2021), pp. 317-331, 10.1016/j.ins.2019.05.042
[8] Xuan S., Wang S. Random forest for credit card fraud detection. (2018)
[9] Vlasselaer V., Van, Bravo C., Caelen O., Eliassi-rad T., Akoglu L., Snoeck M., Baesens B. APATE: A novel approach for automated credit card transaction fraud detection using network-based extensions. Decis. Support Syst., 75 (2015), pp. 38-48, 10.1016/j.dss.2015.04.013
[10] L.E. Faisal, T. Tayachi, S. Arabia, L.E. Faisal, O. Banking, The role of internet banking in society. 18 (13) (2021) 249–257.
[11] Dorphy, Hultquist H. 2017 Financial Institution Payments Fraud Mitigation Survey. Federal Reserve Bank of Minneapolis (2018)
[12] Kurshan E., Shen H., Yu H. Financial crime & fraud detection using graph computing: Application considerations & outlook. 2020 Second International Conference on Transdisciplinary AI (TransAI), IEEE (2020), pp. 125-130
[13] Alhassan A.R.K., Ridwan A. Identity expression—the case of 'sakawa' boys in ghana. Hum. Arenas (2021), Article 0123456789, 10.1007/s42087-021-00227-w
[14] Lebichot B., Borgne Y.A.L., He-Guelton L., Oblé F., Bontempi G. Deep-learning domain adaptation techniques for credit cards fraud detection. INNS Big Data and Deep Learning Conference, Springer, Cham (2019), pp. 78-88
[15] Lebichot B., Siblini G.M.P.W., Bontempi L.H.F.O.G. Incremental learning strategies for credit cards fraud detection. Int. J. Data Sci. Anal., 12 (2) (2021), pp. 165-174, 10.1007/s41060-021-00258-0
[16] Tabachnick B.G., Fidell L.S. Using Multivariate Statistics. Harper Collins, New York (1996)
[17] [Michael J.A., Gordon S.L. Data Mining Technique for Marketing, Sales and Customer Support. John Wiley & Sons INC, New York (1997), p. 445
[18] Breiman L. Random forests. Mach. Learn., 45 (1) (2001), pp. 5-32
[19] Liaw A., Wiener M. Classification and regression by randomForest. R News, 2 (3) (2002), pp. 18-22
[20] Citation O., Systems B. University of Huddersfield Repository Credit card fraud and detection techniques: a review. (2009)
[21] A. Aditi, A. Dubey, A. Mathur, P. Garg, Credit Card Fraud Detection Using Advanced Machine Learning Techniques. (2022) 56–60. http://dx.doi.org/10.1109/ccict56684.2022.00022.
[22] Randhawa K., Loo C.H.U.K., Member S. Credit card fraud detection using AdaBoost and majority voting. IEEE Access, 6 (2018), pp. 14277-14284, 10.1109/ACCESS.2018.2806420
[23] Guanjun L., Zhenchuan L., Lutao Z., Shuo W. Random forest for credit card fraud. IEEE Access (2018)
[24] Ayorinde K. Cornerstone: A Collection of Scholarly and Creative Works for Minnesota State University, Mankato a Methodology for Detecting Credit Card Fraud a METHODOLOGY for DETECTING CREDIT CARD FRAUD Kayode Ayorinde (Thesis Master's) Data Science Minnesota State University Mankato, MN (2021)
[25] Dal Pozzolo A., Caelen O., Le Borgne Y.A., Waterschoot S., Bontempi G. Learned lessons in credit card fraud detection from a practitioner perspective. Expert Syst. Appl., 41 (10) (2014), pp. 4915-4928
[26] Bontempi G. Reproducible machine learning for credit card fraud detection - practical machine learning for credit card fraud detection - practical handbook foreword. May (2021)
[27] A.D. Pozzolo, G. Boracchi, O. Caelen, C. Alippi, Credit Card Fraud Detection: A Realistic Modeling and a Novel Learning Strategy. 29(8) (2018) 3784–3797.