

Intelligent Authentication Agent Model for Detection of ATM Card Fraud in Nigeria

D.N. NNAkwuzie¹, Obasi C.C.², C.I. Akobundu³, K.C Oketa⁴, Nneka Richard-Nnabu⁵, O.L Azegba⁶,
Afolabi Idris Yinka⁷, B.C. Ifebude⁸ and F. S. Bakpo⁹

^{1, 4-8}Department of Computer Science, Alex-Ekweme Federal University Ndufu Alike Ikwo, Ebonyi State
Email: okafordoris49@yahoo.com, christian.oketa@funai.edu.ng, nneka4him@gmail.com, okeazegba@gmail.com
afolabi.yinka@funai.edu.ng, barnabas.ifebude@funai.edu.ng

²Department of Computer Engineering, Edo State University Uzairue
Email: Chukwuemeka.obasi@edouniversity.edu.ng

³Department Computer Science, Evangel University, Akaeze, Ebonyi state, Nigeria
Email: chinyerekingleyfavor@evangeluniversity.edu.ng

⁹Department Computer Science University of Nigeria Nsukka
Email: francis.bakpo@unn.edu.ng

Abstract—ATM card fraud is on the increase with the recent cashless policy and change of naira notes in Nigeria. Transactions are mostly done through cards, making it vulnerable to fraud hence; it is a major issue of concern to the banking sectors as well as to the users, which hinders economic growth, and creates distrust to in the users. Several attempts have been made to investigate the issues and solutions proffered using different approaches such as conventional measures, machine learning, improved firewall security, probability, etc. However, detecting and preventing ATM card frauds has proven to be difficult in both data collection and fraud investigation for the following reasons: there is a shortage of knowledge concerning the access point where ATM fraud is committed and given its infrequency and parallel nature most conventional investigators lack the experience to detect it. The aim of this research work is to develop a high performance intelligent agent-based model for detection of ATM card frauds. Basically, the model involves 3-factor authentication process to determine if a user should be granted access or otherwise, deny access to a bank ATM, these includes; Password module, AI-Agent spending pattern detection and lastly Biometric modules authentication. Logistic regression, SVM and KNN machine learning methodology were applied to historical log files of the ATM cards usage patterns to devise an intelligent data mining classification model. This model was evaluated using confusion matrix with the dataset from the bank. Its effectiveness was ascertained as the accuracy of K Nearest Neighbor (KNN), Support Vector Machine (SVM) and Logistic regression are 93%, 89% and 96% and precision 73%, 98% and 80% respectively. This will help immensely in providing a way forward in curbing the high rate of ATM card fraud in banking sector.

Index Terms— ATM Fraud, Intelligent agent, SVM, KNN, Logistic Regression.

I. INTRODUCTION

Technology advancement has helped the rapid development of the internet, e-commerce and online payment to become major means of trading through the use of an automated teller machine (ATM) and associated cards.

The use of ATM cards brought about a lot of improvements and innovations within the banking industry by raising the financial standard as well as civilizing financial transactions among bank customers [1]. Several reasons trigger consumers, merchants, decision-makers, etc. to acquire ATM debit cards. It was believed that the use of the cashless system will reduce looting fraud to an extent and a lot more.

Although there are lots of benefits accrued to the use of ATM debit cards, it has also introduced some security challenges and fraud. ATM card fraud is one of the major problems in Nigerian banking and the world at large [2]. ATM card fraud is the unauthorized use of an ATM card to engage in unlawful practices with or without the awareness of the ACTM cardholder [3]. One of the key issues hindering the actualization of the cashless vision of Nigeria's economy is electronic banking fraud [4].

Recent developments in Nigeria in terms of commerce have made the ATM cards majorly sources of payment for goods and services, porous to fraudsters. There have been incessant reports of bank customers losing their money from their accounts without trace, this even led temporary shorting down of some banks in Nigeria.

In recent times, the data mining technique is a known technique used for combating crime due to its effectiveness. The data mining technique is a distinct routine that accepts data and generates results in the form of a pattern. The technique will be used to analyze the information and determine if the action on the account holder implies any fraudulent activities or not. It is important to know that data mining is a process that incorporates two elements; the database and machine learning. The former provides data management techniques, while the latter supplies data analysis techniques using an algorithm. The proposed system uses this approach to model the behavior of customers using the withdrawal amount, location, frequency as well as usual time of withdrawal to provide a solution to ATM fraud.

In this work, the model is designed to have three-factor authentication process first authentication process is the password authentication process, which will be implemented using encryption techniques, the second authentication process is the AI-agent module will be implemented using machine learning algorithm; the Logistic regression, SVM and KNN to learn ATM card users' behavior, lastly is the Biometric module while will be implemented to recognize and classify faces of the ATM card users. This model is a high-security performance model as; the password authentication ensures that the transaction is performed by the card user while the machine learning algorithm will have the capability of learning individual ATM user pattern and will trigger fraud when the patterns charges. Lastly, Biometric authentication will reduce ATM skimming and other similar attacks that relates to magnetic stripes.

II. LITERATURE REVIEW

ATM Card fraud detection is a means of monitoring the customer's ATM activities to detect fraud. . Nigeria's banking system in recent times has adopted some technical measures to combat fraud which is eating deeply into the economy of the country, these include; Personal Identification Number (PIN), Card Verification Method (CVM) Biometrics, and operational based rules method. CVM and PIN are used in checking for secret code that is entered into the machine by the customer [9]. Ibrahim et.al believes the act of fraud can be identified and a sophisticated approach can be placed to discover and control the fraudulent activities in the bank [10]. MS. Parktiksha, in his work said that the advanced approach used in identifying an ATM card crime include neural networks, data mining AI, inherent algorithm, games theory, etc [11]. Debit card discovery requires a lot of investigation, complex algorithms, and training to detect and prevent fraud in debit cards [12]. In recent times, the Nigerian banking industry has developed several measures to detect and manage the incidence of fraud in the sectors these include; daily checking, daily account reconciliation, taking stock, and others, in every bank branch [8]. However, it is discouraging that despite all these measures, fraud within the banking industry has continued to be on the increase, the major reason being that technologies behind the combating is the same technology are used to commit the crime. ATM card fraud prevention, the outcome of ATM card fraud in Nigeria shows that, if not taken seriously decision the economy of the country will one day collapse. Finally, the good news is that the technology for preventing debit card fraud is also increasing as debit card fraud in the banking industry increases [13].

To reduce this problem of fraud, the Nigerian banking industry need to adopt intelligent agent-based software for fraud detection systems needs to be built. In researcher's model, it will be able to learn each customer's pattern of transactions to detect fraud in a real-time banking system. Any time there is a change in the pattern of transaction criminal cases will be suspected and the bank will be notified and information on the fraud activities will be reported to the customer. The pattern will be predicted using the following machine learning; logistic regression, SVM and KNN (AI-agent pattern). The password and Biometric authentication process will be also implant to ensure a highly operational system and will use encryption techniques and openCV in Python.

III. PROBLEM DEFINITION

There is a shortage of knowledge concerning the access point where ATM fraud is committed. Again the existing security mechanism where transactions are authenticated using a pin code is not secured since fraudsters can spy on the account holder's pin code and use the same to do online fund transfers from the victim's account.

IV. METHODOLOGY

The technique used to analyze the system is quantitative and qualitative analysis to show the strength of both. Again Logistic regression, SVM and KNN machine learning methodology were applied to historical log files of the ATM cards usage patterns to devise an intelligent data mining classification model to build a better model for ATM fraud detection

V. THE STRUCTURE OF THE PRESENT ATMs BANK SYSTEM IN NIGERIA

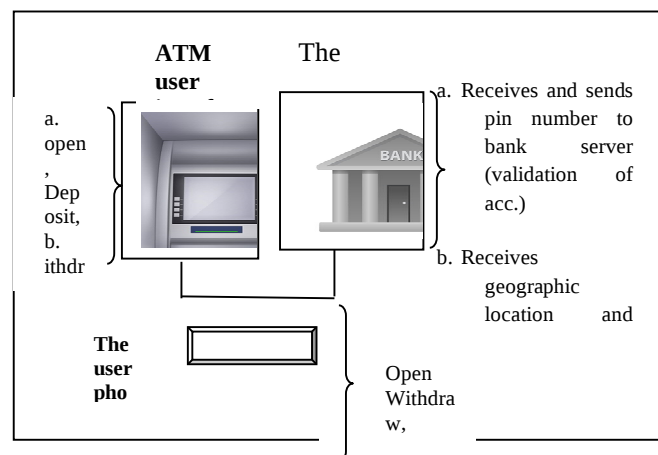


Fig 1: the bank ATM existing system

The figure1 shows the present interface of the ATM existing system, where one can open account, withdraw cash, deposit/cash transfer etc. The figure also shows some of the activities that can take place in the bank server during ATM bank transaction; receives and sends pin number to bank server, (validation of acc), receives geographic location cash transaction and other details, Initiates a fund transfer from bank account etc. the user phone is where detail of all bank transactions is sent.

VI. DESCRIPTION OF PROPOSED SYSTEM

The Intelligent Agent-Based model proposed in this research work is a software system that is embedded with some intelligence from a machine learning algorithm known as logistic regression, SVM and K-NN with 3-factor authentication, the figure 2 is the proposed model for ATM card detection.

Input Interface: An input interface is a form that is used to receive the ATM card transaction information.

Output Interface: This interface would be used for informing the users of the status of the input details after the agent has been able to decide if the transaction was fraudulent or not.

Software Interface: The software interface is the implementation of a set of Application Interfaces that implement business logics such as a mining engine, database engine and rule engine, used by the system to achieve the respective objectives.

The method will be using logical conditions to compare known patterns of transactions to be able to detect fraud. Some conditions that may be tested include but not limited to the following: time interval of the transaction, the geographical distance and the amount involved in the transaction.

Database

This is a relational database system that store transaction logs for future references and for re-training of the intelligent agent to enhance learning.

Intelligent Agent

This is the most important component of the system, where the logics of the data mining is implemented. The intelligence would be implemented using three models, which includes the

i. Logistic Regression, which is a binary classification model that uses a logistic function, also known as the sigmoid function given in equation (i) to transform linear regression into Logit, is log of Odds.

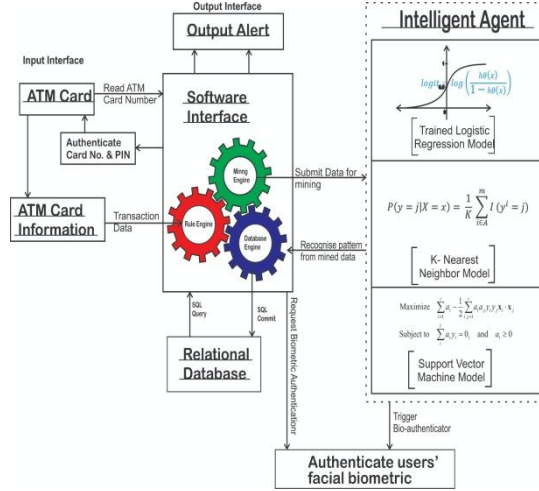


Fig 2: The Proposed model for detection of ATM card fraud detection

$$f(x) = \frac{1}{1 + e^{-x}} \quad (i)$$

ii. K-Nearest Neighbor, which finds the Euclidian distance between the mathematical values of each data point and the sample data, as is given in equation (ii) and also the probability of the point similar to the test data given as equation (iii)

$$d(x, y) = \sqrt{(x_{sample} - x_i)^2 + (y_{sample} - y_i)^2} \quad (ii)$$

$$P(y = j|X = x) = \frac{1}{K} \sum_{i \in A}^m I(y^i = j) \quad (vii) \quad (iii)$$

iii. Support Vector Machine models, and would be trained to classify differently, the transaction details into fraudulent and normal transactions.

The system involves a 3-stage or 3-factor authentication process to build a super security system and to determine if a user should be granted access or otherwise, deny access; the three main authentication modules, namely:

- ❖ Password authentication module
- ❖ AI-Agent spending pattern authentication module
- ❖ Biometric authentication module

The **Password authentication** is the 1st stage of the authentication process, will use encryption techniques, and will be implanted using Bcrypt hashing function to store and retrieve Personal Identification Number (PIN) in Python.

The **AI-agent Spending pattern module** will be implemented using a machine learning algorithm; Logistic regression, SVM and KNN to learn ATM card users' behavior based on the Frequency of ATM use, Amount involved, Distance between two usage points, and Time between two usage points. A dataset containing a similar usage pattern of an ATM card holder will be used to train the module to detect fraud in ATM transactions.

The **Biometric module** will be implemented using OpenCV in Python to detect, recognize and classify faces of the ATM card users.

VII. THE APPLICATION INTERFACE OF THE AUTHENTICATION INTELLIGENT AGENT MODEL

The figure 3 shows the interface of data capturing, where the ATM user inputs their personal data. The information collected will be stored in the intelligent model database for use during ATM bank transaction to compared with new input to check if the person is a genuine user or not. Figure 4 shows an **Interface for card number input** and figure 5 shows interface of successful transaction.

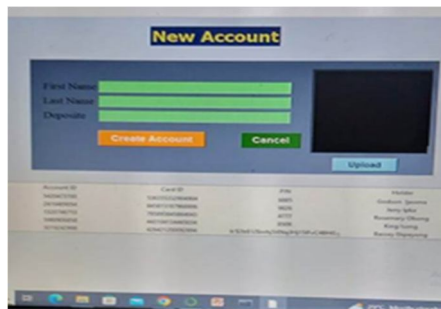


Fig 3 interface of creating New Account



Fig 4 Interface of entering an account number



Fig 5 : interface of a successful transaction

VIII. RESULT OF PERFORMANCE ANALYSIS

The intelligent agent was implemented using the Logistic Regression, SVM, and K-Nearest Neighbor Model and the tested for performance using the following performance matrices:

Receiver Operating Characteristic Curve (ROC), Confusion Matrix, System Accuracy and Area Under the ROC Curve (AUC)

Receiver Operating Characteristic Curve (ROC)

The ROC is a graphical representation of the performance of a classification model using the True Positive Rate (TP) and True Negative Rate (TN) values of the predictions.

Confusion Matrix

The visual representations of the overall performance of the classification models were done using confusion matrix. The matrix shows the percentage of success and failure a model could have. Hence it can be used to determine parameters used to perform classification model evaluation

System Accuracy (Acc)

The accuracy test of the system was performed to determine the percentage of correctness of the system in making correct predictions using the test dataset. The performance accuracy was determined using the formula below:

$$Acc = \frac{TP + TN}{TP + TN + FN + FP}$$

This implies that accuracy is given by the sum of the correctly predicted outcomes (TP and TN) divided by the total possible predictions (TN, TP, FN, and FP).

Area under the ROC Curve (AUC)

The total area under the ROC curve was plotted to determine for each model under investigation. This value is required to validate the performance accuracy of the classification model. This means that even when a system presents a high accuracy; it may still tend to perform poorly in trying to achieve correct classification.

Performance evaluation of the Model:

The **Performance** of KNN model : The KNN model was trained with the same dataset and subjected to the same test data. The result yielded the result shows, the AUC value of 0.98, precision of 0.73 and an accuracy value of 0.93.

AUC-ROC of the KNN Model generated with python code using the test dataset from the original dataset

Performance of SVM Model

The SVM model result shows AUC value of 0.92, precision of 0.98 and an accuracy value of 0.89.

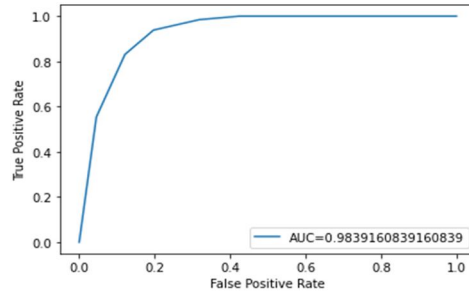


Fig 6: The Performance of KNN model

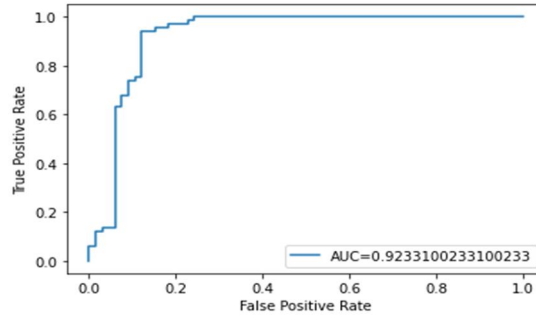


Fig8: Performance of SVM Model

AUC-ROC of the SVM Model generated with python code using the test dataset from the original dataset

Performance of Logistic Regression Model:

The Logistic Regression model result shows AUC value of 0.98, precision of 0.80 and an accuracy value of 0.96.

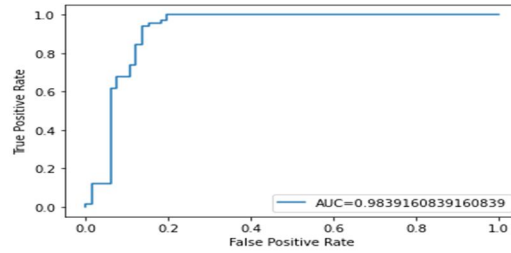


Fig 9: Performance of Logistic Regression Model

AUC-ROC of the Logistic Regression Model generated with python code using the test dataset from the original dataset.

TABLE I: SUMMARY OF THE RESULTS

Model	Accuracy	AUC	Precision
KNN	93%	98%	73%
SVM	89%	92%	98%
Logistic Regression	96%	98%	80%

Discussion of the Result of Performance Evaluation

From result summary table 1, the largest percentage accuracy is the logistic regression model, while the AUC values for the logistic regression and KNN is the same (98%). The precision value is for the SVM is the highest. Overall observation shows that the three models would do very well in the classification of ATM fraud detection.

IX. CONCLUSION

The issue of ATM fraud is on the increase, which is highly exasperating; hence the researchers developed an intelligent agent model using logistic regression, SVM and KNN machine learning with 3-factor authentication

process to detect fraudulent activities in the daily ATM transactions. The model is built in the form that it will learn each customer spending pattern, in order to classify transaction as fraud or not. To have a high operational intelligent agent-based model for detection of ATM card frauds, for any user to access the intelligent agent model must pass through the three authentication process and must be satisfied as a genuine card user before the user can access any bank transaction services, this makes the model a powerful security system.

REFERENCE

- [1] Ismail, S., Amin, H., Shayeri, S. F., & Hashim, N. (2014). Determinants of attitude towards credit card usage. *Journal Pengurusan*, 145-154.
- [2] CBN. (2013). Payment system vision 2020. Techniques: A review, *Bank and Bank system*, 4(2013).
- [3] Masoumeh, Z., Seeja, K.R and Alam, M.A., "Analysis of credit card fraud detection technique: based on certain design criteria." *International journal of computer applications* (0975-8888), vol. 52-no 3, August, 2012.
- [4] Oludaya, T., & Oluwatosia, A. (2016). Banking fraud is costing the Nigeria economy dear. . Retrieved from the conversation.com/banking_fraud_is_costing_the_Nigeria_economy_dearly_68889.
- [5] Egu.J. (2008). The role of information and communication and communication technology (ICT) In fraud Detection Nigeria Bank.
- [6] Ojigbede.F. "Fraud in Bank", a paper presented at the effective bank institute course organized by FITC, lagos, 2000
- [7] Chijioke Amaka, Agozie Eneh, Collins Udanor, Onyesolu M.O and Uchenna Nduka(2020), Determining the adoption of e- transaction authentication frameworks in Nigerian commercial banks. *International journal of Engineering and Technology* (IJET).
- [8] Anuar.N Sallehudia, H., Gani, A., & Zakari, O. (2008). Identifying false alarm for network intrusion detection system using hybrid data mining and decision tree. *Malaysian Journal for computer science*, 110-115.
- [9] Bolton, R., & Hand, D. (2002). *Unsupervised profiling methods for fraud detection*. London.
- [10] Ibrahim, A., & Barroon(2010), M. A. (2010). Fraud Detection and control on ATM Machines, An Algorithm for combating cash and fund transfer. *International Journal of physical science*, 5.
- [11] Ms. Partiksha.L., Meshram and Tarun Yenganti "Credit and ATM card fraud prevention using multiple cryptography algorithms" *International Journal of Advanced Research in Computer Science and Software Engineering* Volume 3, 2013
- [12] Ermatita and Indrajani Sutedja, "Detection of fraud for debit card transactions at automated teller machine in Indonesia using Neural network". *Journal of physics: conference services*. Volume 1196, conference 1 published by IOP science. COF SER.1196 012076. 2019
- [13] Sivakumar, and R. Balasubramanian, "Fraud Detection in credit card transaction classification, Risks and prevention techniques" *international journal of Computer science and information technologies* vol. 6(2), 1379-1386. ISSN: 0975-9646. 2010.