LAUTECH Journal of Engineering and Technology 18 (1) 2024: 38-46

EVOLUTION OF A MODEL TO DETERMINE UNSECURED TRANSACTIONS

[1*]Ozoh P., [2]Olawuyi E., [3]Olayiwola M., [4]Ibrahim M. , [5]Kolawole M., [6]Olubusayo O., [7]Adigun A. and [8]Ogundoyin K.

[1,2,4,7,8]Department of ICT, Osun State University, Nigeria.
[3,5]Department of Mathematical Sciences, Osun State University, Nigeria.
[6]Department of Physics, Osun State University, Nigeria.

Corresponding Author, email: patrick.ozoh@uniosun.edu.ng; olayiwola.oyedunsi@uniosun.edu.ng

ABSTRACT

The widespread presence of fraudulent transactions in financial institutions is of significance in banking operations. Examples of financial instruments that are utilized include credit cards, smart cards, swipe cards, etc. These cards provide important information and enable small costs to be incurred by customers. These small amounts are removed from customer accounts. Banks need to discover the correctness of transactions, thus the introduction of the evaluation of models to determine unsecured transactions. The focus of this research is to contribute to the field of the application of machine learning to banking operations by introducing tools for predicting unsecured transactions in the banking sector. The research objectives include the examination of different methods utilized in machine learning for investigating unsecured transactions about the physical stealing of credit cards and the illegal collection of details on credit cards. To accomplish the aims of this research, information gathering is done using Kaggle. Kaggle is obtainable online. The major focus of this research is to examine cardholders' spending patterns. The method includes using a multilayer perceptron (MLP). This is utilized with training of 70% and testing of 30% subsets. The evaluation of the model is done using a confusion matrix technique. This research is implemented using the Python programming language. The model produces accuracy rates of 93% and 99% respectively. This research can leverage achievements recorded to improve security concerns in financial institutions.

Keywords: *Machine learning, Online transactions, Credit card, Fraud, predictions, MLP, Genetic algorithm, evaluation*

INTRODUCTION

Recently, Information and Communication Technology has been widely applied to various sectors (Wiraguna et al., 2023). The mechanisms involved in ICT cannot be differentiated from developments in technology, E-commerce, and innovation. This study investigated the legal aspect of e-commerce for the support of economic development. Li et al. (2023) investigates challenges in identifying assessment groups in research data that are applied to receive insights into customer pattern recognition and to evolve a marketing strategy. The paper applies k-anonymity and evolves a method known as k-anonymization (k-MM), developed to protect the method.

According to Taupit & Azizan (2023), E-commerce can be referred to as the sales and buying of yield from work, using an electronic network. These internet-based transactions are utilized by organizations to incur some costs on customers for goods and services that are rendered. The increasing volume of Internet transactions has made the number of Internet frauds to be bigger. The technique utilized in this study is the Waterfall method, which exists in the Software Development Life Cycle. Zhao et al. (2023) determined to upgrade

the knowledge of the total difference between exposed events and reverse models. The paper investigates the simulation model measuring the total model difference. This study was modeled using optimal control and connected to the proposed simulation model. The reliability of the proposed simulation model is established using empirical data. The review of the change from interbank offered rates (IBORs) to the risk-free rates (RFRs), was proposed for the act of determining IBOR being discontinued. As a result of the transition, the differences between the forward-looking rates and the backward-looking rates are determined. Majorly, the evolution of pricing models is formulated (Russo & Fabozzi, 2023).

Rantanen (2023) investigates the techniques used for modeling loss based on unsecured consumer loans. This paper also explores the process by which these techniques are presently being applied. An exhaustive literature review discovered that the most used modeling techniques are the regression-based models. The recommendations from the research can be summarized as recommendations received for companies seeking to improve their daily organizational functions. Depending on the application of the evolutionary concept to the world's human perspective, Søgaard et al. (2024) presented results from an existing database. 14 groups are classified as either global, technology, or structural systems. An examination of the study shows that 12 concepts (86%) are in an advanced stage of occurrence with a high risk of occurrence. Pan et al. (2024) propose a unique method for selecting dangerous activity on the blockchain. By incorporating profane transaction data characteristics, the system can discover different groups of threats and give an insight into the various characteristics. Large-scale attacks are increasingly becoming a great threat to systems (Li et al.,2024). Adversely, vigorously executing the system is a daunting task. It is important to uncover leakages in complex systems. As it is, the time process for leakage discharge can be inaccurate. The output from the study indicates that the proposed technique is more reliable than previous methods.

Hu et al. (2024) majorly investigate the real control system for a unique nonlinear multi-purpose system. This unique nonlinear system can develop its communication network system and can hold out against malicious threats. Furthermore, to manage the system, a control system is applied, that consists of increased efficiency, small cost, and high applicability. The reliability of the system was confirmed in the experiments carried out in the study. Babu (2024) provides a compressed strategy for investigating the mechanism of the cybersecurity system using improvised artificial intelligence. This stem considers the techniques involved in cybersecurity and classifies typical threats. The study presents the potential of innovative artificial intelligence, investigating actual malicious threats, proactive protection, and uninterrupted learning. The project is a guide to protecting the cyber system. Muksalmina et al. (2024) presents the attitudinal characteristics for the use of e-money in an environment. The research describes the significance of developing systems that make e-money attractive to users, at the same time giving insights to users and the financial sector looking to improve operations in the financial sector.

Kayikci & Khoshgoftaar (2024) presents a fundamentals report on blockchain and machine learning. The paper discusses their concise applications to challenges in the economy, healthcare, and security issues. Machine learning techniques are able can analyze huge quantities of information to achieve important insights, resulting in a data-driven decision-making process, and making stronger guaranteed measures for the removal of potential threats to the integrity of data

gathering. In as much as electronic financial transaction penetration is set to rise worldwide, knowing the variables that impact the acceptance of electronic money is important (Rekha, 2024). The study investigates the characteristics inherent in the use of electronic transfers in an environment. The takeaway from the study focuses on the important attributes of making use of the system and comfort in the use of the system. Also is accessing customer interest in the financial system, with variables such as speed and efficiency in making possible the use of the electronic platform.

An important study of computational techniques was undertaken by Ozoh et al. (2022). Each method possesses problems and impediments. Modeling information needs an elevated degree of reliability, In as much as any change from actual can result in inaccurate results. The methods applied in different studies include various challenges. In the study, the use of machine learning techniques can be applied to model non-linear problems, with some of the problems being ambiguous. Some of these methods can provide insight to expose secret information, as a result developing inaccurate models. A set of artificial learning methods, in as much as they possess different robustness and fragility, were used in Patrick et al. (2022). The research verifies various methods to select a reliable method. The proposed technique can make better preceding techniques for modeling information. The methods are computed by looking at the reliability of computed estimates by evaluating their respective errors. In the paper, in selecting a reliable method, the performance metrics are applied to the techniques to validate the reliability of the developed models.

METHODOLOGY

This section discusses the research methodology used to develop the proposed model. It gives the details of data collection, design model, implementation, and simulation.

Data collection

Data collection is a technique of collecting and assessing data in a measured manner that allows one to answer a set of questions and allows for result processing. Information gathering is processed using Kaggle. Kaggle is obtainable online. Kaggle is an internet-based platform in the data science community. It consists of powerful tools to assist in challenges in data science. In the data collected for this study, Table 1 is the dataset attribute. Table 2 is the site used for the study, Table 3 shows the costs incurred, and Table 4 shows the duration of the study.

Table 1: Sample dataset attribute

ID	Location	Expenditure	Date of birth	Family Type	Transaction Status	Data/time	Project Type	Phone	Email
0	500806	99501	10/0/1986	0	Successful	9/1/2020	10.4	101	0
1	500808	85001	12/10/2001	0	Successful	09/01/2020	22.5	109	1
2	500809	99505	10/05/1986	0	Successful	09/01/2020	15.1	204	1
3	500811	99504	5/12/1991	0	Successful	09/01/2020	15.1	204	1
4	500811	86162	07/10/1990	0	Successful	09/01/2020	3.1	301	1

Proposed algorithm

The proposed algorithm optimizes the dataset for the best feature selection because of its attribute to search for a solution. This attribute of the Genetic algorithm makes it stand out from all clustering algorithms. The proposed algorithm is in Figure 1. The pseudo-code is in Appendix 1.

Table 2: Site for the study

Number	Site
1	S/West
2	S/East
3	S/South
4	N/Central
5	N/West
6	N/East

Table 3: Cost

Number	Cost	Level
1	0- 50,000	Bottom
2	51,000 -100,000	Middle
3	100,000 – 200,000	Top

Table 4: Duration

Number	Cost
1	06:01 – 18:00
2	18:01 – 24:00
3	24:01 – 06:00

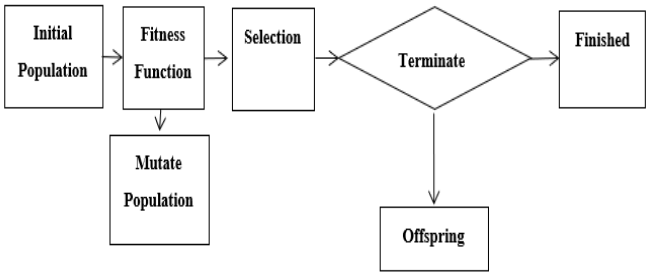

Figure 1: Generic algorithm architecture.

Unified modelling language

This method indicates the attribute of the study. The structure of this research is in Figure 2.

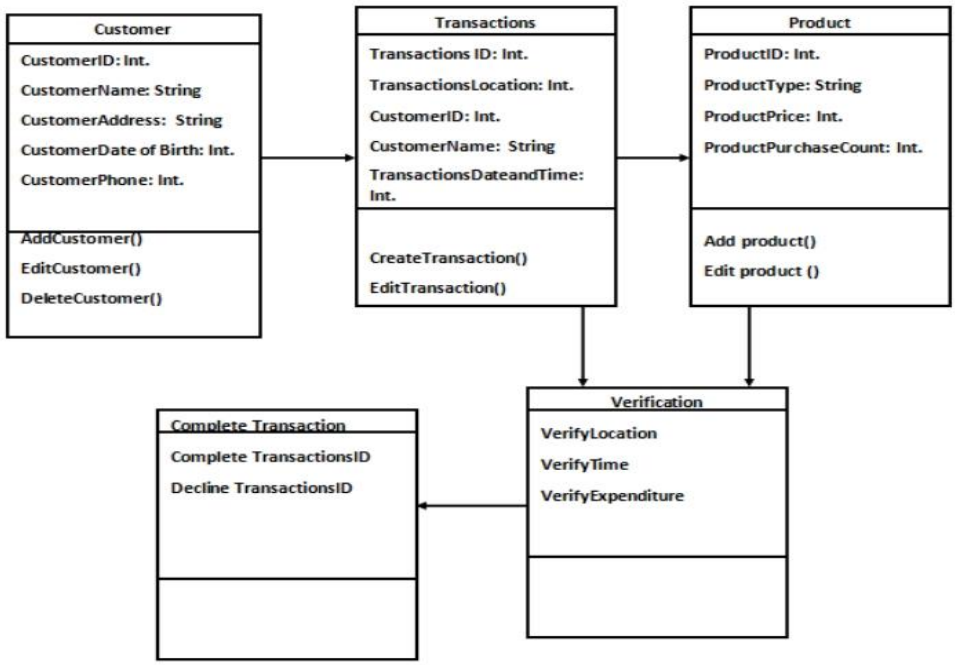

Figure 2: Class diagram.

Multilayer perceptron

The multilayer perceptron is used in this research. The MLP has a concealed layer and conveys outputs with more than two classes. The hidden layers contain adequate neurons to comprehend the information included and create two distinct groups. The multilayer perceptron architecture is in Figure 4. The pseudo-code is in Appendix 2.

Design model

The classification is by using the algorithm shown in Figure 5.

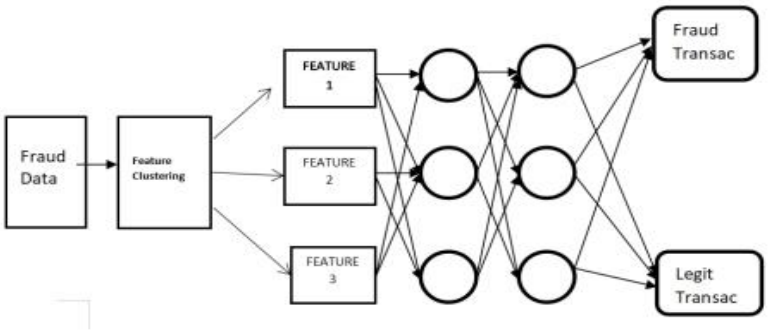

Figure 3: The proposed algorithm

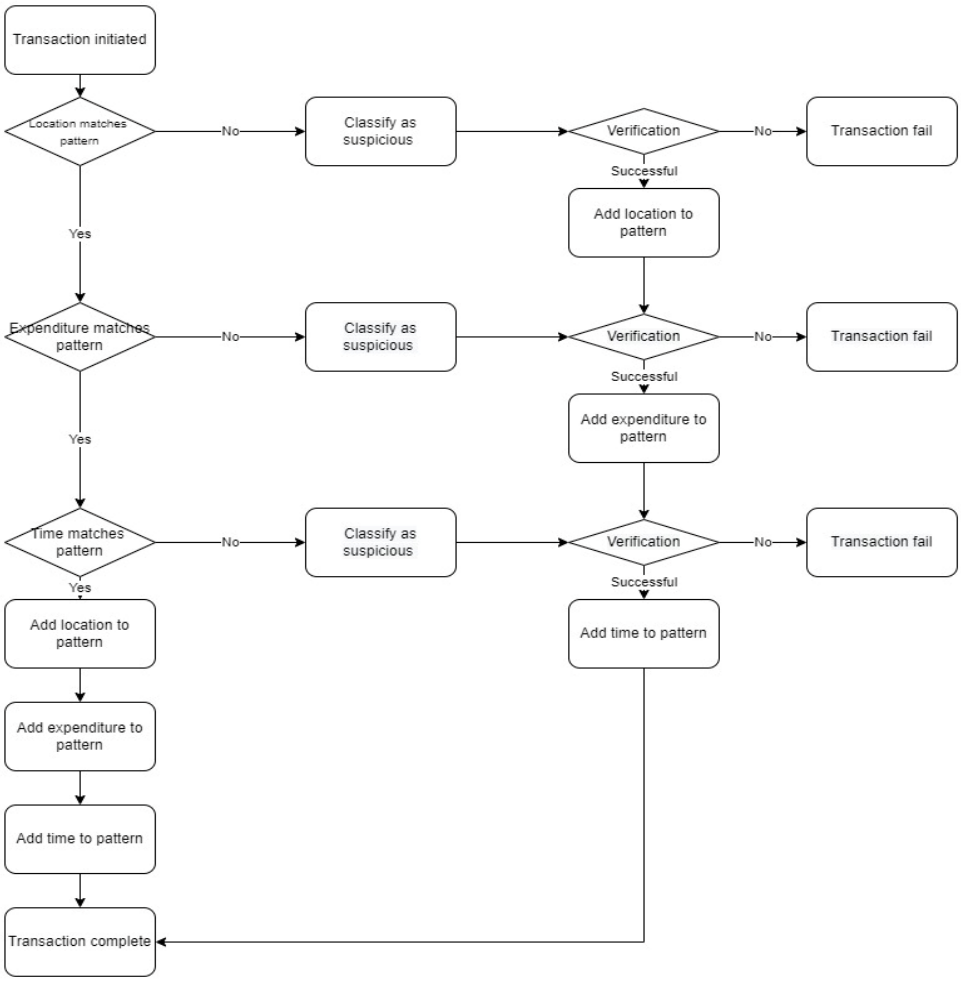

Figure 4: Algorithm for a Fraud Detection Model

The model checks the location, time, and expenditure as the parameters to decide a spurious transaction. The algorithm is in Figure 4.

Training the model

The model is trained with 25 groups of batch size 32. During the training, the model tries to learn and recognize patterns for better accuracy. Figure 5 shows model training.

The accuracy, exactness, and responsiveness of results are evaluated by utilizing Equation (1), Equation (2), Equation (3), and Equation (4).

```
Epoch 1/10
164/164 [==============================] - 84s 188ms/step - loss: 0.6608 - accuracy: 0.7575 - val_loss: 0.6913 - val_accuracy: 0.6250
Epoch 2/10
164/164 [==============================] - 29s 173ms/step - loss: 0.3864 - accuracy: 0.7961 - val_loss: 0.6163 - val_accuracy: 0.6266
Epoch 3/10
164/164 [==============================] - 29s 173ms/step - loss: 0.3393 - accuracy: 0.8090 - val_loss: 0.7877 - val_accuracy: 0.6250
Epoch 4/10
164/164 [==============================] - 29s 173ms/step - loss: 0.3556 - accuracy: 0.8238 - val_loss: 0.7722 - val_accuracy: 0.6250
Epoch 5/10
164/164 [==============================] - 29s 173ms/step - loss: 0.3321 - accuracy: 0.8346 - val_loss: 0.5687 - val_accuracy: 0.6747
Epoch 6/10
164/164 [==============================] - 29s 173ms/step - loss: 0.2881 - accuracy: 0.8544 - val_loss: 0.6175 - val_accuracy: 0.7019
Epoch 7/10
164/164 [==============================] - 29s 174ms/step - loss: 0.2770 - accuracy: 0.8807 - val_loss: 0.6563 - val_accuracy: 0.6458
Epoch 8/10
164/164 [==============================] - 29s 173ms/step - loss: 0.2822 - accuracy: 0.8607 - val_loss: 0.5800 - val_accuracy: 0.6747
Epoch 9/10
164/164 [==============================] - 29s 173ms/step - loss: 0.2760 - accuracy: 0.8595 - val_loss: 0.5644 - val_accuracy: 0.6939
Epoch 10/10
164/164 [==============================] - 29s 174ms/step - loss: 0.2254 - accuracy: 0.8960 - val_loss: 0.7381 - val_accuracy: 0.6042
<keras.callbacks.History at 0x7fa9602d6a10>
```

Figure 5: Model training.

$$\text{Accuracy} = \frac{TP+TN}{TP+FPFN+TN} \quad (1)$$

Precision shows how much the forecasts are right.

$$\text{Precision} = \frac{TP+FP}{TF+FN} \quad (2)$$

The small part of true positives that are accurately identified as positives is measured by Equation 3.

$$\text{Sensitivity} = \frac{TP}{TP+FN!} \quad (3)$$

The F1 Score measures the test.

$$\text{F1 Score} = \frac{2TP}{(2TP+FP+FN)} \quad (4)$$

RESULTS AND DISCUSSION

This section presents the results of the implementation conducted in this study. This section also provides a detailed discussion of the results and findings of the proposed model. The result justifies the performance of this study in line with the aim and objectives of the study. A confusion matrix is a method used to outline the reliability of a classification method. Other methods can be used for this process, such as when there are unequal observations within a dataset. Estimating the confusion matrix will provide an accurate method for estimating the classification technique. A Two x Two confusion matrix is employed in this study because the number of correct and incorrect predictions are outlined with count values and broken down by each class. The confusion matrix consists of four characteristics (numbers) that define the measurement metrics of the classifier. These four numbers are TP (True Positive), TN (True Negative), FP (False Positive), and FN (False Negative). Figure 6 shows the confusion matrix of the Multilayer Perceptron. The result indicates that TP= 56792; FP= 72; FN= 5; TN= 93. The confusion matrix of the MLP + Genetic Algorithm is defined in Figure 7. The confusion matrix shows TP= 56822; FP= 42; FN= 3; TN=95.

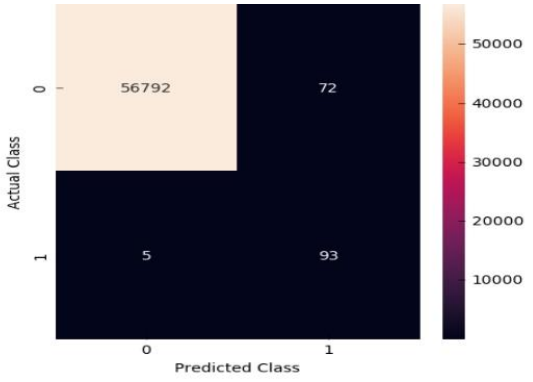

Figure 6: Fraud dataset.

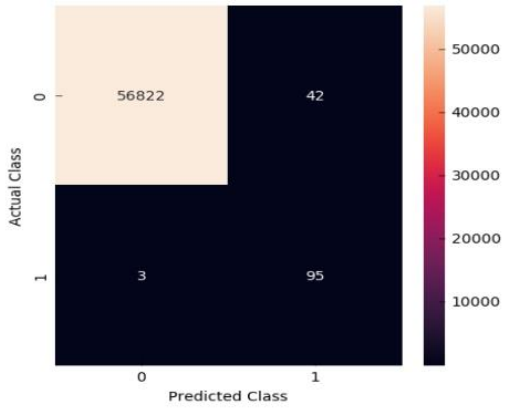

Figure 7: Confusion matrix.

Performance evaluation.

The performance of the model is given by its accuracy. The dataset was pre-processed and split into the training and testing set. The ideal combination of variables for a decent prediction model was determined on the training dataset (70%) and evaluated on the testing dataset (30%).

Table 5 shows performance evaluation in percentage against the metrics for Multilayer Perceptron and Multilayer Perceptron with Genetic Algorithm. The data was passed into the MLP and MLP with a Genetic Algorithm classifier and results were obtained with an accuracy of 93% and 99%.

Table 5: Performance evaluation

Performance measure %	MLP+GA	MLP
Accuracy	99.86	93.99
Sensitivity	99.99	96.99
Specificity	69.34	56.36
Precision	99.83	99.87
F1-Score	99.96	99.93

The findings of this study reveal that the multilayer perceptron (MLP) achieved an accuracy of 93.99% in modeling the dataset. The comparison with the results of Berhane et al. (2023) indicates that their method yielded an accuracy of 91.88%. The proposed model demonstrates superior performance compared to the previous approach.

CONCLUSIONS

In this study, Machine learning models, such as MLP and RF were used in the development of a credit card prediction system to help detect fraud. The study made use of a bank credit card dataset which was obtained from the Kaggle repository. The dataset was pre-processed and split into the training and testing set. An ideal combination of variables for a decent prediction model was determined on the training dataset (70%) and evaluated on the testing dataset (30%). The data was passed into the MLP with a Genetic Algorithm classifier and the results were obtained with an accuracy of 93% and 99% respectively.

This work is enormous and advantageous to the banking sector in predicting fraudulent transactions. The result shows that MLP with a Genetic Algorithm has better accuracy compared to MLP alone. However, future research work could explore further by incorporating real-life data with the model to test the accuracy in real-time.

REFERENCES

[1] Wiraguna, S. A., Santiago, F., & Redi, A. (2023). Legal Harmonization Of E-Commerce Transactions In Order To Support Indonesia's Economic Development. Journal of Social Research, 2(6).

[2] Li, S., Schneider, M. J., Yu, Y., & Gupta, S. (2023). Reidentification risk in panel data: Protecting for k-anonymity. Information Systems Research, 34(3), 1066-1088.

[3] Taupit, A. S. M., & Azizan, N. (2023). The Planning Process of the Online Transaction Fraud Detection Using Backlogging on an E-Commerce Website. Malaysian Journal of Science, Health & Technology (MJoSHT).

[4] Zhao, J., Knoop, V. L., Sun, J., Ma, Z., & Wang, M. (2023). Unprotected Left-Turn Behavior Model Capturing Path Variations at

Intersections. IEEE Transactions on Intelligent Transportation Systems.

[5] Russo, V., & Fabozzi, F. J. (2023). The Transition from Interbank Offered Rates to Risk-Free Rates: Evolution in Pricing Models for Interest Rate Derivatives. The Journal of Fixed Income, 32(4), 45-59.

[6] Rantanen, J. (2023). Modeling loss given default on unsecured consumer loans: a case study of a Finnish financial institution.

[7] Søgaard Jørgensen, P., Jansen, R. E., Avila Ortega, D. I., Wang-Erlandsson, L., Donges, J. F., Österblom, H., ... & Crépin, A. S. (2024). Evolution of the polycrisis: Anthropocene traps that challenge global sustainability. Philosophical Transactions of the Royal Society B, 379(1893), 20220261.

[8] Pan, B., Stakhanova, N., & Zhu, Z. (2024). EtherShield: Time-interval Analysis for Detection of Malicious Behavior on Ethereum. ACM Transactions on Internet Technology, 21(1), 1-30.

[9] Li, Y., Zhu, J., Liu, Z., Tang, M., & Ren, S. (2024). Deep Learning Gradient Visualization-based Pre-silicon Side-channel Leakage Location. IEEE Transactions on Information Forensics and Security.

[10] Hu, X., Xiong, Y., Zhang, Z., & Li, C. (2024). Consensus of a novel heuristic nonlinear multi-agent system in DOS attack network environment via saturation impulse control mechanism. ISA transactions.

[11] Babu, C. S. (2024). Adaptive AI for Dynamic Cybersecurity Systems: Enhancing Protection in a Rapidly Evolving Digital Landscap. In Principles and Applications of Adaptive Artificial Intelligence (pp. 52-72). IGI Global.

[12] Muksalmina, M., Ahmadsyah, I., & Dianah, A. (2024). Understanding E-Money Preferences Among Students: A Case Study at FEBI UIN Ar-Raniry, Banda Aceh, Indonesia. Grimsa Journal of Business and Economics Studies, 1(1), 1-11.

[13] Kayikci, S., & Khoshgoftaar, T. M. (2024). Blockchain meets machine learning: a survey. Journal of Big Data, 11(1), 9.

[14] Rekha, M. (2024). Determining Intrusion Attacks Against Online Applications Using Cloud-Based Data Security. EAI Endorsed Transactions on Scalable Information Systems.

[15] Ozoh, P., Olayiwola, M. O., & Adigun, A. A. (2022). Prediction of electricity consumption based on complex computational method. Mathematical Sciences and Informatics Journal (MIJ), 3(1), 56-65.

[16] Patrick, O., Nwade, I., & Adigun, A. (2022). DEVELOPMENT OF AN INTELLIGENT SYSTEM FOR MODELING. The Islamic University Journal of Applied Sciences (JESC), 2022(12).

[17] Berhane, T., Melese, T., Walelign, A., & Mohammed, A. (2023). A Hybrid Convolutional Neural Network and Support Vector Machine-Based Credit Card Fraud Detection Model. Mathematical Problems in Engineering, 2023.

APPENDIX 1

Generic Algorithm Pseudocode.

```
START
Generate the initial population
Compute fitness
REPEAT
    Selection
    Crossover
    Mutation
    Compute fitness
UNTIL population has converged
STOP
```

Appendix 2

Multilayer Perceptron Algorithm

BEGIN

pd.read.csv(.csv) reads the dataset

resampling of data

StandardScaler() #scaling

Train_test_split()

Obtain the model

END

Author details

[1] University of North Dakota, Department of Mechanical Engineering

[2] Osun State University, Department of ICT

Competing interests

No competing interests were found.

Grant information

No funding was received for this study.