Malaysian Journal of Computing, 5 (2): 537-552, 2020
Copyright © UiTM Press
eISSN: 2600-8238 online

NEURO-FUZZY DATA MINING SYSTEM FOR IDENTIFYING E-COMMERCE RELATED THREATS

Saibu Aliyu Haruna[1], Akinyede Raphael Olufemi[2] and Boyinbode Olutayo Kehinde[3]

[1]*Department of Computer Science, The Federal University of Technology, Akure, Nigeria*
[2]*Department of Information Systems, The Federal University of Technology, Akure, Nigeria*
[3]*Department of Information Technology, The Federal University of Tech., Akure, Nigeria*
[1]saibualiy@gmail.com, [2]roakinyede@futa.edu.ng, [3]okboyinbode@futa.edu.ng

ABSTRACT

E-commerce is driven via Information Technology (IT), especially the web, and it mostly relies upon on innovative technologies that are facilitated by Electronic Data Interchange (EDI) and Electronic Payment over the web. Several researches have shown that e-commerce platforms are compromised by means of phishing and fraud attacks. This has necessitated the importance of trying to find innovative methodologies for protecting e-commerce systems and users from the said threats. This research integrates Case Based Reasoning Module (CBRM) and Adaptive Neuro-Fuzzy Inference System (ANFIS) to spot and categorise e-commerce websites transactions as legitimate or illegitimate by analysing and evaluating some attributes. This may provide an invulnerable platform for e-commerce users. The system which was implemented on MATLAB can be deployed on e-commerce systems and servers to watch e-commerce requests with the aim to identify legitimate and illegitimate websites and transactions. The result of the implementation indicates that the developed system is promising.

Keywords: *e-Commerce, Adaptive Neuro-Fuzzy Inference System (ANFIS), Electronic Data Interchange (EDI), Information Technology, K-nearest Neighbour (KNN) Algorithm*

Received for review: 01-04-2020; Accepted: 28-07-2020; Published: 08-09-2020

1. Introduction

E-commerce is widely considered as buying and selling of products and services over the web, and any transaction that is achieved fully through digital measures are often viewed as e-commerce (Kishor, 2013; Singh *et al*., 2016). Personal computers, laptops, mobile phones and the internet are considered as the infrastructures that aided the emergence of e-commerce and e-transaction (Akinyede & Akinyede, 2015). E-commerce has brought numerous benefits to technology-driven commerce which makes the process of shopping and selling faster and easier at any given time. Despite the convenience associated with shopping and selling on the web, e-commerce is bedevilled by some security threats like phishing and fraud which is largely because that it makes use of the web as its driving infrastructure. Securing customers' data is a major challenge hindering the growth of e-commerce (Bandara *et al.,* 2019). Fraudsters are constantly seeking ways to take advantage of online shoppers who commit novice errors. Common errors that make people susceptible to security threats include the following: shopping on websites that are not secure, giving out too many personal information and leaving computers vulnerable to viruses (Murphy, 2018). Phishing is a criminal mechanism employing both social engineering and technical subterfuge to steal purchasers' private identification data and financial account credentials for the malicious purpose (Niranjanamurthy & Dharmendra, 2013; Ramachandran & Chang, 2016). Usually, phishing attacks will direct the recipient to a website designed to mimic the target enterprise's actual visual identity with the aim of obtaining private personal data, which often result in the victim being unaware of the malicious event. Acquiring this type of private data is appealing to black hat hackers because it allows them to impersonate their victims and make fraudulent financial transactions. Victims often incur huge financial losses or have their entire identification stolen usually for criminal purposes (Chirag *et al*., 2012; Irvin-Erickson, 2019).

Despite numerous researches on enhancing the safety of e-commerce websites and transactions, yet various threats still exist with more experienced threats, namely phishing and financial fraud. Most e-commerce users are unaware that their web browsers can expose useful information about their transactions to hackers and fraudsters after visiting malicious/phishing websites. (Ramachandran & Chang, 2016).

With the increasing number of phishing websites which pose a threat to the overall security of e-commerce platforms, there is a need to develop a strong and effective solution to identify e-commerce phishing websites and fraudulent transactions. The study adopts a neuro-fuzzy based system and data mining techniques to assist in the evaluation and classification of e-commerce websites and transactions into "legitimate" and "illegitimate" The study also aims to employ knowledge mining which is processing data into information, which mostly involves identifying patterns within large data sets that are impossible for humans to discover manually (Lee & Yoon, 2017). Adaptive Neuro-Fuzzy Inference System (ANFIS) is one of the system that has the potential of acquiring knowledge from data that is inherently not accurate and maintain a high level of performance within the presence of doubt to supply solutions to problems (Arinkoola, 2016).

2. Related Work

Detecting phishing websites and fraud is a crucial step towards ensuring security in e-commerce platforms. Several approaches are adopted to unravel these problems. This section reviews different studies on phishing and fraud detection schemes.

A fraud detection system for e-commerce transactions was developed by employing a prudential multiple consensus model (Carta *et al*., 2019). This was achieved using data intelligence technique based on a prudential multiple consensus model which integrates the effectiveness of some modern classification algorithms by using a two-fold criterion, probabilistic and majority based selection. The aim was to maximise the effectiveness of the model in detecting fraudulent transactions regardless of any data imbalance. This model was validated with a set of experiments on a large real-world dataset characterised by a high degree of data imbalance and results confirmed that the proposed model performed best compared to other existing classification algorithms. However, the model could not be evaluated and also characterised by a high degree of data imbalance.

One approach is the phishing webpage detection for secure online transactions (Fowdur & Khader, 2018) that was designed to detect phishing websites used for e-commerce transactions. In this approach, three layers of criteria are used: Google page rank, IP address in URL and quality of webpage content. The phishing website detection system consists of three modules; data collection module which finds phishing and genuine e-commerce websites for analysis, fuzzy rule base containing fuzzy rules to assist the inference ripper engine make logical conclusion about the genuineness of a webpage or website, and classification module that uses symbolic logic to classify websites according to associated risk factors. In the second approach, a prototype intelligent Intrusion Detection System (IIDS) for e-banking was developed using the effectiveness of fuzzy logic and data mining techniques (Khraisat *et al*., 2019). The system was designed using fuzzy logic to provide more information for risk managers to efficiently manage and detect website phishing associated risks by combining historical data and expert input. Fuzzy logic and data mining algorithms which include; C4.5, RIPPER, PART, PRISM and CBA were used to assess e-banking phishing websites risk using twenty-seven (27) factors. Linguistic variables were used to represent key phishing characteristic indicators associated with e-banking phishing website probability. The system was implemented using WEKA and MATLAB. Two publicly available data sets were used to test the implemented system.

A secure environment for client-side e-commerce payment system using an encryption system (Akinyede *et al*., 2014) was developed to provide a secure means of protecting customers' personal and transaction data from fraud using encryption. The system was divided into three parts namely: merchant server-side scripting which handles customers' requests, customers-side scripting that makes a request to the online server and the host-side that deals with funds transfer. The security mechanism employed in this system is the symmetric cryptographic scheme supported by Advanced Encryption System (AES) encryption and decryption algorithm as a means of protecting transaction data and credentials in e-commerce transactions and this technique provided an efficient solution in protecting the transactions of consumers.

With the aim of protecting e-commerce systems from internet fraud (Phani & Mahaboob, 2013), a prototype application that detects fraudulent e-commerce transactions was developed. A genetic algorithm with multiple criteria is developed to detect fraud, namely payment card usage frequency, payment card usage location, overdraft on the payment card and payment card balance. A prototype application was built using JAVA and it was developed using the Graphical interface (GUI) to ensure user friendliness. Intensive performance evaluation of the prototype was also performed.

A neuro-fuzzy approach was employed to detect phishing websites and protect purchasers when performing online transactions (Aburrous *et al*., 2010). A hybrid neuro-fuzzy method was used to develop a phishing website detection model that offered an effective solution, using two-fold cross-validation. The results from this model suggest that the proposed Neuro-Fuzzy system that used five (5) inputs was powerful in detecting phishing websites with high accuracy in real-time. The proposed system made use of rules, user-behavior profile, phish-tank, pop-ups from emails). Two-Fold cross-validation was applied to carry out training of the proposed model and a set of 243 rules was generated. The researchers have proposed a Transductive Support Vector Machine (TSVM)-based system way of phishing page detection. The system was independent of the attack method and did not affect the users' behaviour. Though the system performs well but the result is only a preliminary investigation of detecting phishing web page using TSVM. As a result, much are expected to be done in improving its performance.

3. Methodology

ANFIS and CBRM will be adopted in developing the proposed Neuro-Fuzzy Data Mining System model. The model, which will be referred to as Adaptive Neuro-Fuzzy Inference System Design Model (ANFIS-DM), will intelligently identify and classify e-commerce websites and transactions into either legitimate or illegitimate entities by systematically evaluating features or attributes of e-commerce websites and transaction data to detect phishing websites and fraudulent transactions. A set of defined linguistic variables are modelled for correct interpretation of results using a scale as shown in Table 1 and Table 2 where the extracted features from e-commerce transactions are the formulated classification based on the possible outcomes or conditions of each parameter/attribute. Figure 1 depicts the procedure of the ANFIS-DM model and the crisp values of input parameters representing the model's attributes. The fuzzy set of parameters (attributes) is represented by 'X' which is defined as in equation 1.

$$X = \{x_1, x_2, x_3, \ldots x_n\} \tag{1}$$

where x_n represent the *n*th parameter or attribute of X and *n* is the total number of parameters in X (here n=15). For each of the parameter, a group of constraints are defined which makes it easy to scale properly. In each parameter, standard or acceptable range of values/labels is assigned as in Table 1. The linguistic value is classified as {Very Legitimate, Legitimate, Moderately Legitimate, Slightly Legitimate, Slightly Illegitimate, Moderately Illegitimate, Illegitimate, Very Illegitimate}.

Table 1. Linguistic Variables for ANFIS-DM.

Linguistic Values	**Range of Values**
Very Legitimate	(0.80) - (1.00)
Legitimate	(0.60) - (0.79)
Moderately Legitimate	(0.30) - (0.59)
Slightly Legitimate	(0.00) - (0.29)
Slightly Illegitimate	(-0.20) - (-0.01)
Moderately Illegitimate	(-0.40) - (-0.21)
Illegitimate	(-0.70) - (-0.41)
Very illegitimate	(-1.00) - (-0.71)

In this research, features/attributes of e-commerce transactions listed in Table 2 are extracted from an e-commerce transaction request which serves as an input/sample case to the ANFIS-DM. These extracted attributes are included into query of the case-based system for the closest similar case to the new sample case input into the ANFIS-DM. The closest similar case is retrieved by employing K-nearest Neighbour (KNN) algorithm using the Euclidean distance. The retrieved case serves as input to the fuzzy module where fuzzification and inference take place using generated rules to return a distinct final output/result; and the classification formulated based on the possible conditions of each parameter/attribute. Basically, the procedure of the proposed ANFIS-DM model is composed of six functional blocks (see Figure 1).

a. The input is a rule base containing a number of fuzzy if-then rules;
b. a database which defines the membership functions of the fuzzy sets used in the fuzzy rules;
c. CBR system is a decision-making paradigm that performs the inference operations on the rules;
d. A fuzzification module which transforms the fuzzy inputs into degrees of match with linguistic values;
e. ANFIS module serves as a basis for constructing a set of fuzzy if-then rules with appropriate membership functions to generate the stipulated input-output pairs.
f. A defuzzification module which transform the fuzzy results of the inference into a fuzzy output.

Usually, the rule base and the database are jointly referred to as the knowledge base.

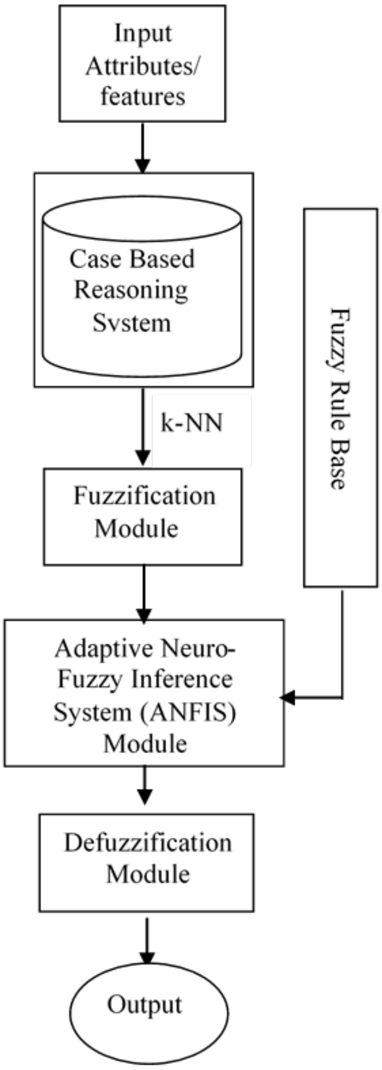

Input Attributes/ features
Case Based Reasoning Svstem
k-NN
Fuzzification Module
Fuzzy Rule Base
Adaptive Neuro-Fuzzy Inference System (ANFIS) Module
Defuzzification Module
Output

Figure 1. The procedure of proposed ANFIS-DM model.

Table 2. Operational Variables of the ANFIS-DM.

Variable Name	**Possible options**	**Value**	**Classification**
Port Number	Using HTTP, HTTPS port 80, port 443	1	Legitimate
	Using ports besides port 80, port 443	-1	Illegitimate
IP Address in URL	No IP address in URL e.g. www.jumia.com	1	Legitimate
	IP Address in URL http://128.98.3.123/paypal.net	-1	Illegitimate
URL Length	URL length is less than 54 valid characters	1	Legitimate
	URL length is Greater than 54 valid characters	-1	Illegitimate
Presence of Symbols in URL	URL do not contain (-) or (@) symbol	1	Legitimate
	URL contains (-) or (@) symbol	-1	Illegitimate
Domain Registration Validity	Domain expiration period is greater than a year	1	Legitimate
	Domain expiration period is less than a year	-1	Illegitimate

Favicon	Favicon not loaded from external domain	1	Legitimate
	Favicon is loaded from external domain	-1	Illegitimate
Server Form Handler (SFH)	(SFH) do not refer to a different domain	1	Legitimate
	(SFH) is blank and redirect to a different domain	-1	Illegitimate
Submitting data to email	Not using mail() or mailto function to submit user information	1	Legitimate
	Using mail() or mailto function to submit user information	-1	Illegitimate
Status of Customers Location	Unhide customers location by using an open IP proxy	1	Legitimate
	Hide customers location by using an open IP proxy	-1	Illegitimate
IP Location of Customer and Issue Bank	Customer's IP and the card-issuing bank are from the same country	1	Legitimate
	Customer's IP and the card-issuing bank are not from the same country	-1	Illegitimate
High Risk Country IP Address	Customer's IP is not from a high-risk country	1	Legitimate
	No, customer's IP is from a high-risk country	-1	Illegitimate
Matching of billing and delivery detail	Delivery and billing countries match	1	Legitimate
	Delivery and billing countries do not match	-1	Illegitimate
Domain of contact e-mails	E-mail is not from a free domain, such as Hotmail or Gmail	1	Legitimate
	E-mail is from a free domain, such as Hotmail or Gmail	-1	Illegitimate
Transaction exceeds maximum limit	Transaction do not exceed maximum allowable billing or withdrawal per individual or company	1	Legitimate
	Transaction exceeds maximum allowable billing or withdrawal per individual or company	-1	Illegitimate
Number of transactions	One or two transactions from same credit card number	1	Legitimate
	More than two transactions from same credit card number	-1	Illegitimate

3.1 Case Base Reasoning (CBR) Module

CBR is a decision-making paradigm where new cases are solved relying on previously solved comparable instances (Yikun *et al*., 2019). CBR approach mimics how humans' reason and learn; hence it makes it a promising approach for building intelligent systems (Zhai *et al*., 2019). In this research, a database of solved cases is employed, and every case is described via a group of input attributes associated with a designated output. Extracting useful information from this database can help the CBR system in providing a reliable result on yet to be solved cases.

The CBR model was adopted to ensure the efficiency and reliability of the system. In this research, the CBR module focuses on two primary steps of the CBR cycle which involves retrieval and reuse of solutions from previous cases. Case retrieval is performed based upon similarity of the solved case to the new (unsolved) case. Here, the closest similar cases to the new case are retrieved by employing K-nearest Neighbour (KNN) algorithm using the Euclidean distance as shown in equation 2.

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \tag{2}$$

where *i= 1, 2, 3, …, n*

d(*x*,*y*) computes the distance between new (unsolved) case and retrieved case, x_i represents the value of each attribute for the new case, y_i represents the value of each attribute for the retrieved cases and *n* is the total number of attributes. The retrieved cases are provided as input to the ANFIS module for further processing which incorporates model training and rules generation.

3.2 Adaptive Neuro-Fuzzy Inference System Module

The ANFIS is a Takagi-Sugeno-Kang (TSK) type of fuzzy model proposed by Takagi-Sugeno-Kang (Takagi & Sugeno, 1985; Shafaei *et al*., 2017). It integrates both neural networks and fuzzy logic principles and it has the potential of capturing the benefits of both techniques into a single framework (Sampson *et al*., 2019). ANFIS is a data-driven technique representing a neural network approach for the solutions of function approximation problems. Data-driven approaches for the synthesis of the networks are typically based on clustering a training set of numerical samples of the unknown function to be approximated. On account of its introduction, its networks have been efficiently applied in classification tasks, rule-based process control, pattern recognition and similar problems. This fuzzy model generates fuzzy rules from an input/output data set.

ANFIS under consideration has a number of inputs and one output. The rule base contains the fuzzy IF-THEN rules of Takagi and Sugeno’s type as follows:

IF antecedent (x_i) THEN consequents (w), f(x_i) (Rule 1)

where x_i is the antecedent, w is the firing strength of the rule and f(x_i) is a crisp function in the consequent. The ANFIS structure usually consists of 5 layers. Figure 2 shows the architecture of the ANFIS module of the ANFIS-DM.

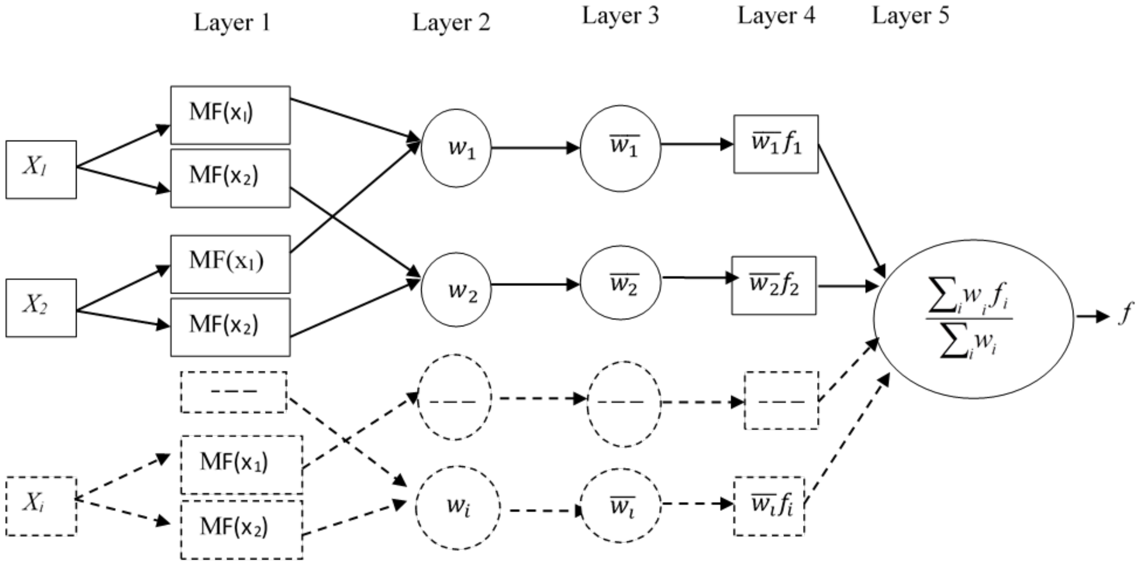


Figure 2. The architecture of ANFIS network.

In layer 1 (L1), Gaussian Membership (GM) function is used to map input values of each x_i node to its appropriate membership value (see equation 3). The Gaussian Membership Function is specified by two parameters c_i and σ_i, where c_i represents Membership Function's centre (threshold) and σ_i represents its width. These parameters are called the premise parameters and are used to adjust the shape of the membership function.

$$L1 = \mu(x_i) = e^{-\frac{1}{2}\left(\frac{x_i - c_i}{\sigma_i}\right)^2} \tag{3}$$

where *i= 1, 2, 3, ..., n*

In layer 2 (L2), each node in this layer calculates the firing strength of a rule via multiplication. Here, the calculation of the weight (w_i) or firing strength of each rule output is computed. In this layer, the input values are the membership functions and each node multiply inputs and gives an output which represents the firing strength of a rule. The output of this layer is given by equation (4).

$$L2 = w_i = \prod \mu(x_i).x_i \tag{4}$$

where *i= 1, 2, 3, ..., n*

In layer 3 (L3), the nodes calculate the ratio of the rule's firing strength to the sum of all the rules firing strength. The result is a normalized firing strength shown by equation (5).

$$L3 = \overline{w_i} = \frac{w_i}{\sum_{i=1}^{n} w_i} \tag{5}$$

where *i= 1, 2, 3, ..., n*

In layer 4 (L4), each node in this layer computes the contribution of each rule towards the overall output with equation (6).

$$L4 = \overline{w_i} f_i \tag{6}$$

where *i= 1, 2, 3, ..., n*

In layer 5 (L5), the single node in this layer computes the overall output as the summation of contribution from each rule. This simply implies that the output of the fuzzy inference system is calculated by summing all rule outputs using equation (7).

$$L5 = \sum_i \overline{w_i} f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} \tag{7}$$

4. Implementation

To successfully implement the model, the MATLAB programming tool was used. Several MATLAB commands was applied and stored in an M file (M file is a MATLAB code file). The implementation stages are listed as follows.

a. Building the Graphic User Interface (GUI), the GUI was designed using the MATLAB GUIDE command.
b. Retrieving the dataset from the database, the dataset named "e-commerce Phishing" was retrieved from the University of California machine learning repository (UCI). It is an Attribute Relation File Format (ARFF). It was converted to a Comma Separated Value (CSV) file and was later transferred to MySQL database called e-commerce. The database and tables in the e-commerce database is retrieved for pre-processing using MATLAB code fragments.
c. Data conversion and pre-processing, the retrieved tables are converted into matrix for further prepossessing using

    ```
    cc=traintable.Data; dd=testtable.Data; ee=checktable.Data; c=cell2mat(cc); d=cell2mat(dd); e=cell2mat(ee);
    ```

The ANFIS-DM model was developed using the ANFIS model development in three phases as follows:

i) Phase One: Generating Initial Rules

```
in_fis = genfis2(intrain,outtrain,radii);
```

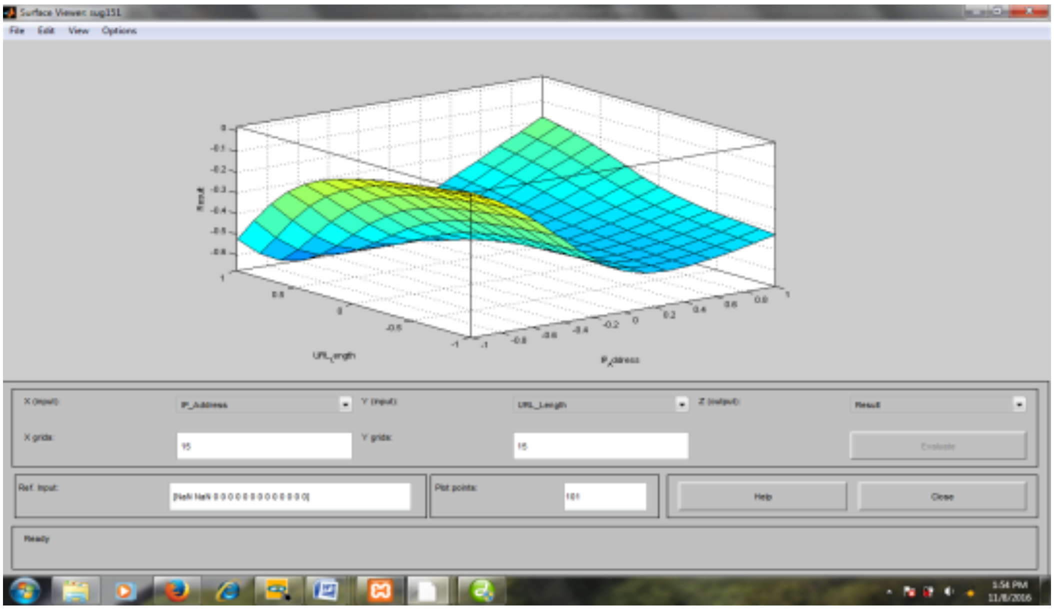
Note: genfis2 is a MATLAB command. genfis2 generates an ANFIS structure using subtractive clustering and requires separate sets of input and output data as input arguments. When there is only one output, genfis2 may be used to generate an initial FIS for model training. genfis2 accomplishes this by extracting a set of rules that models the data behavior or generate patterns from the initial data set.

ii) Phase Two: Generating Adaptive Model structures

```
[fis1,error,stepsize,fis,chkErr] = anfis(datatrain,in_fis, trnOpt, dispOpt,datatest);
```

iii) Phase Three: Assigning names to inputs and outputs. The results from the refined rules generated from the FIS as well as the input are assigned names for easy identification.

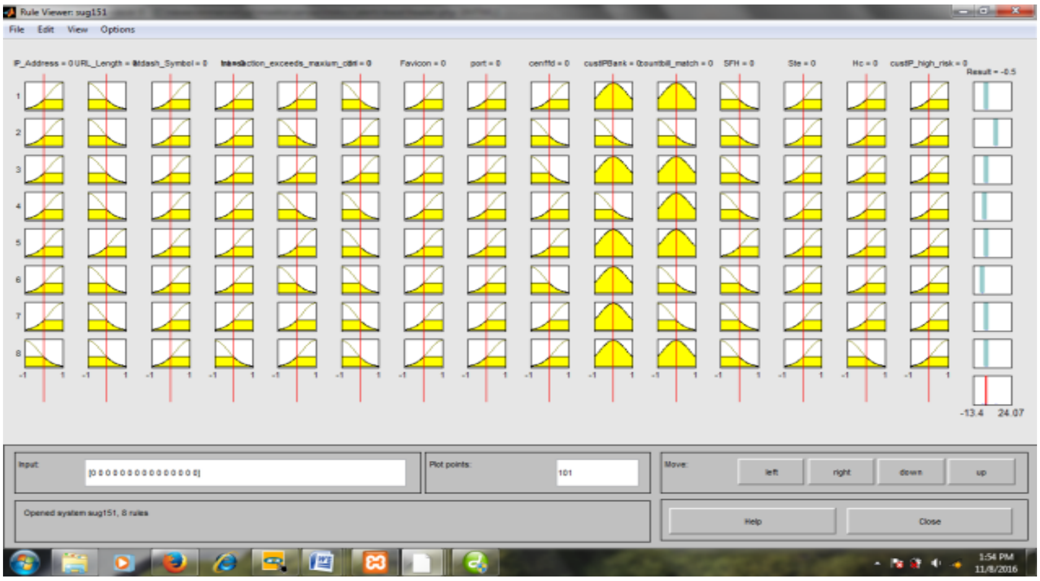
The Surface Viewer is a graphical interface that lets you examine the output surface of a fuzzy inference system for any one or two input variables. In Figure 3, the input variables URL length and IP address in URL was considered. The Surface Viewer is a read-only editor because it does not alter the fuzzy system or its associated fuzzy inference system structure in anyway.



X (input): IP_Address
Y (input): URL_Length
Z (output): Result
X grids: 15
Y grids: 15
Ref. Input: [NaN NaN 0 0 0 0 0 0 0 0 0 0 0 0 0]
Plot points: 101
Ready

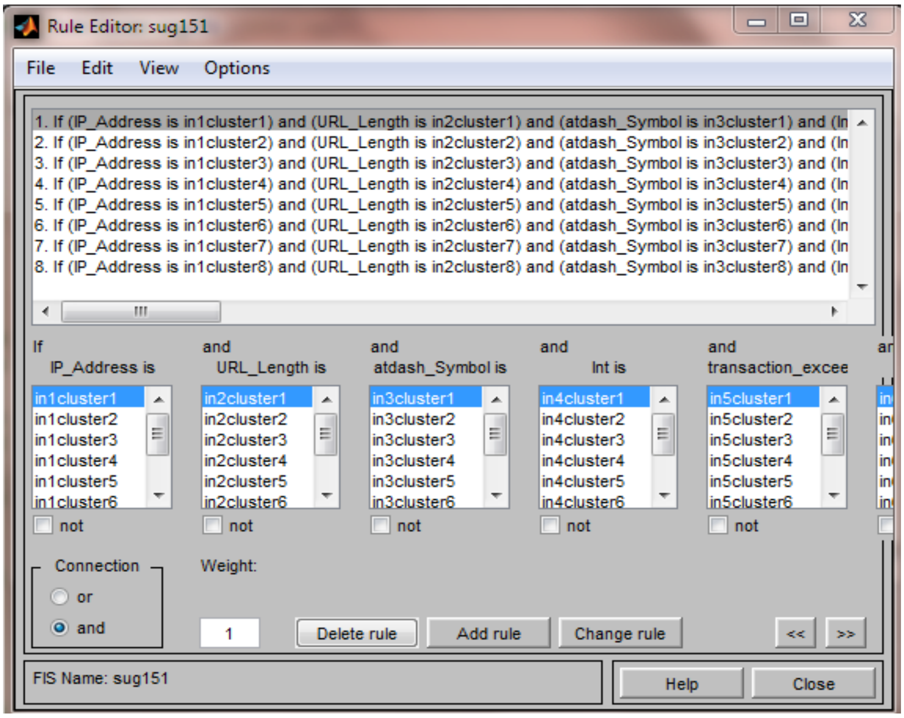
Figure 3. Surface View of the system.

The rule viewer is used to view the entire implication process of the Fuzzy Inference System from the beginning to the end. The line of indices can be moved around corresponding to the inputs. The system re-adjusts and computes new output as shown in Figure 4. Eight rules were generated to drive the inference mechanism for the ANFIS-DM as shown in Figure 5.



Input: [0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
Plot points: 101
Move:
Opened system sug151, 8 rules

Figure 4. Rule Viewer.



```
1. If (IP_Address is in1cluster1) and (URL_Length is in2cluster1) and (atdash_Symbol is in3cluster1) and (In
2. If (IP_Address is in1cluster2) and (URL_Length is in2cluster2) and (atdash_Symbol is in3cluster2) and (In
3. If (IP_Address is in1cluster3) and (URL_Length is in2cluster3) and (atdash_Symbol is in3cluster3) and (In
4. If (IP_Address is in1cluster4) and (URL_Length is in2cluster4) and (atdash_Symbol is in3cluster4) and (In
5. If (IP_Address is in1cluster5) and (URL_Length is in2cluster5) and (atdash_Symbol is in3cluster5) and (In
6. If (IP_Address is in1cluster6) and (URL_Length is in2cluster6) and (atdash_Symbol is in3cluster6) and (In
7. If (IP_Address is in1cluster7) and (URL_Length is in2cluster7) and (atdash_Symbol is in3cluster7) and (In
8. If (IP_Address is in1cluster8) and (URL_Length is in2cluster8) and (atdash_Symbol is in3cluster8) and (In
```

If IP_Address is	and URL_Length is	and atdash_Symbol is	and Int is	and transaction_excee	an
in1cluster1	in2cluster1	in3cluster1	in4cluster1	in5cluster1	in
in1cluster2	in2cluster2	in3cluster2	in4cluster2	in5cluster2	in
in1cluster3	in2cluster3	in3cluster3	in4cluster3	in5cluster3	in
in1cluster4	in2cluster4	in3cluster4	in4cluster4	in5cluster4	in
in1cluster5	in2cluster5	in3cluster5	in4cluster5	in5cluster5	in
in1cluster6	in2cluster6	in3cluster6	in4cluster6	in5cluster6	in
□ not	□ not	□ not	□ not	□ not	□

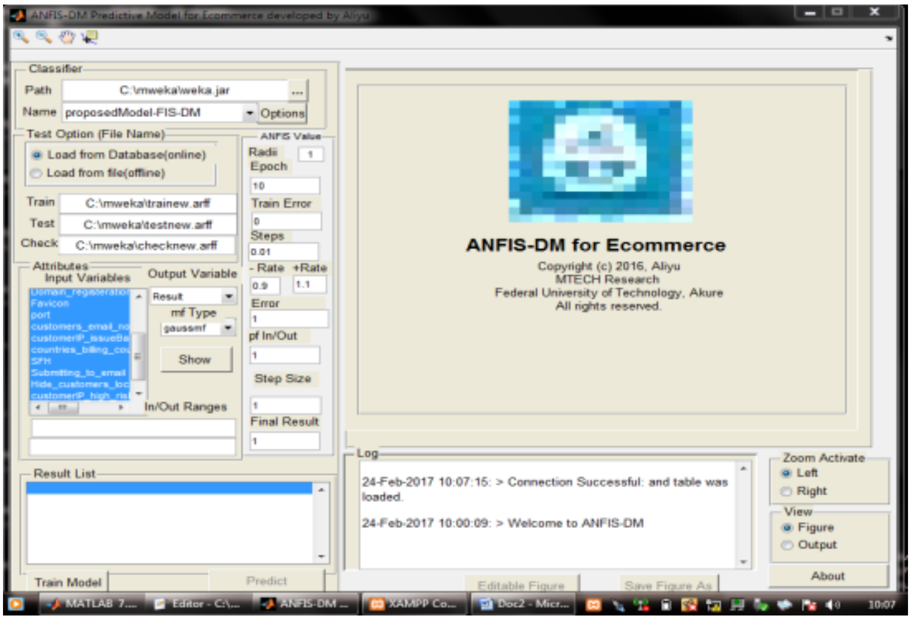
Connection
○ or
● and

Weight: 1

FIS Name: sug151

Figure 5. Generated fuzzy rules from ANFIS-DM.

The ANFIS-DM model is selected from the name drop down list and input attributes are successfully loaded from the database. These input attributes provide the model with the required attribute data for analysis and evaluation (see Figure 6). The computing clusters run the model and then the interface (in Figure 7) displays the parameters and information regarding result of the ANFIS-DM Model.



Path: C:\mweka\weka.jar

Name: proposedModel-FIS-DM

● Load from Database(online)
○ Load from file(offline)

Train: C:\mweka\trainew.arff

Test: C:\mweka\testnew.arff

Check: C:\mweka\checknew.arff

Input Variables: Domain_registeration, Favicon, port, customers_email_no, customerIP_issueBa, countries_billing_cou, SFH, Submitting_to_email, Hide_customers_loc, customerIP_high_ris

Output Variable: Result

mf Type: gaussmf

In/Out Ranges

Radii: 1

Epoch: 10

Train Error: 0

Steps: 0.01

- Rate: 0.9 +Rate: 1.1

Error: 1

pf In/Out: 1

Step Size: 1

Final Result: 1

ANFIS-DM for Ecommerce
Copyright (c) 2016, Aliyu
MTECH Research
Federal University of Technology, Akure
All rights reserved.

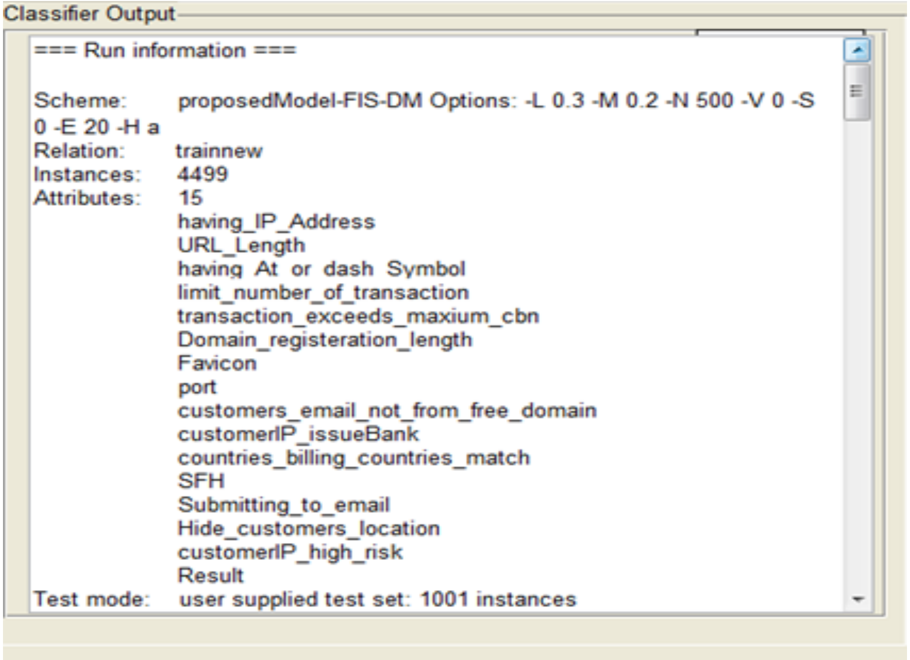
```
24-Feb-2017 10:07:15: > Connection Successful: and table was
loaded.

24-Feb-2017 10:00:09: > Welcome to ANFIS-DM
```

Zoom Activate: ● Left ○ Right

View: ● Figure ○ Output

Figure 6. Loading attributes/input variables for the ANFIS-DM.



```
=== Run information ===

Scheme:       proposedModel-FIS-DM Options: -L 0.3 -M 0.2 -N 500 -V 0 -S
0 -E 20 -H a
Relation:     trainnew
Instances:    4499
Attributes:   15
              having_IP_Address
              URL_Length
              having At or dash Symbol
              limit_number_of_transaction
              transaction_exceeds_maxium_cbn
              Domain_registeration_length
              Favicon
              port
              customers_email_not_from_free_domain
              customerIP_issueBank
              countries_billing_countries_match
              SFH
              Submitting_to_email
              Hide_customers_location
              customerIP_high_risk
              Result
Test mode:    user supplied test set: 1001 instances
```

Figure 7. Viewing ANFIS-DM Parameters and Results.

The interface in Figure 8 displays the result of the ANFIS-DM model result by showing plotted graphs of the desired output and the actual output for different instances of testing data. The screenshot in Figures 9a and 9b show the result for instances 1-33 and 860-892 of testing data. For every instance, the integer value on the left is the desired/expected output and the real value on the right is that of the actual output.

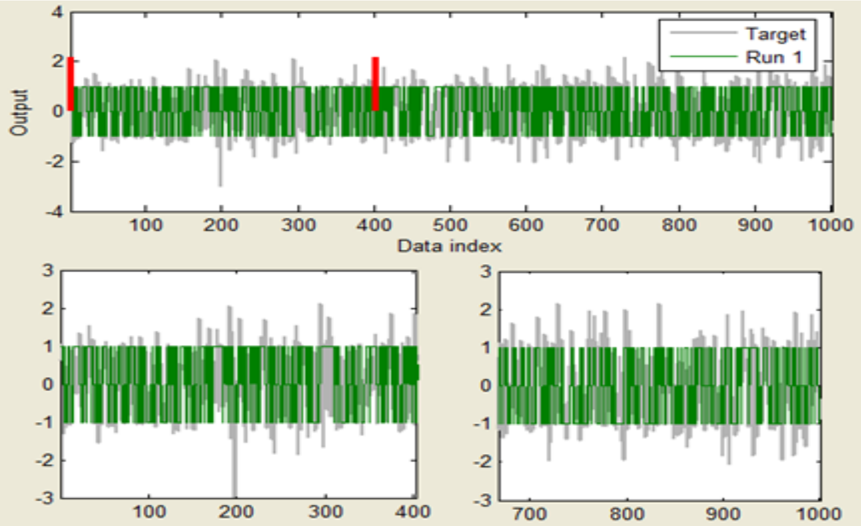
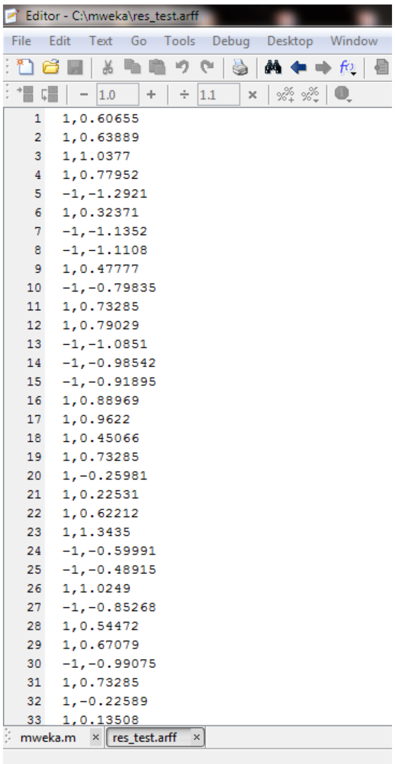
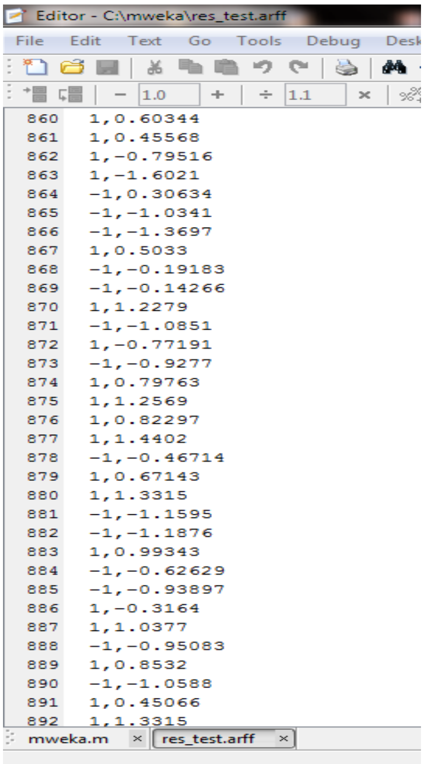


Figure 8. Viewing ANFIS-DM result.


Editor - C:\mweka\res_test.arff
File Edit Text Go Tools Debug Desktop Window
1.0
1.1
mweka.m
res_test.arff

```
1,0.60655
1,0.63889
1,1.0377
1,0.77952
-1,-1.2921
1,0.32371
-1,-1.1352
-1,-1.1108
1,0.47777
-1,-0.79835
1,0.73285
1,0.79029
-1,-1.0851
-1,-0.98542
-1,-0.91895
1,0.88969
1,0.9622
1,0.45066
1,0.73285
1,-0.25981
1,0.22531
1,0.62212
1,1.3435
-1,-0.59991
-1,-0.48915
1,1.0249
-1,-0.85268
1,0.54472
1,0.67079
-1,-0.99075
1,0.73285
1,-0.22589
1,0.13508
```

Figure 9a. ANFIS-DM Result on Testing.


Editor - C:\mweka\res_test.arff
File Edit Text Go Tools Debug Desk
1.0
1.1
mweka.m
res_test.arff

```
1,0.60344
1,0.45568
1,-0.79516
1,-1.6021
-1,0.30634
-1,-1.0341
-1,-1.3697
1,0.5033
-1,-0.19183
-1,-0.14266
1,1.2279
-1,-1.0851
1,-0.77191
-1,-0.9277
1,0.79763
1,1.2569
1,0.82297
1,1.4402
-1,-0.46714
1,0.67143
1,1.3315
-1,-1.1595
-1,-1.1876
1,0.99343
-1,-0.62629
-1,-0.93897
1,-0.3164
1,1.0377
-1,-0.95083
1,0.8532
-1,-1.0588
1,0.45066
1,1.3315
```

Figure 9b. ANFIS-DM Result on Testing.

5. Discussions

The ANFIS-DM was trained using 4499 data instances of known output and validated with 1001 instances of test data. Statistical tests offer a certain level of assurance about the validity and accuracy of a model. In this research, the performance of the proposed ANFIS-DM was computed using the Root Mean Square Error (RMSE). The performance of ANFIS-DM was evaluated and compared with other predictive models namely: Linear Regression and Artificial Neural Network (ANN). This was done to ascertain how accurate the ANFIS-DM model classifies e-commerce websites compared to Linear Regression and Artificial Neural Network. Also, the estimated time taken to completely build each of the models was also captured. The obtained result is presented in Table 3.

Table 3. Performance Evaluation Table

Model	RMSE on Training Data	RMSE on Testing Data	Start Time	Stop Time	Total Time taken to build model
Linear Regression	0.561	0.579	01:18:23	01:18:33	10 seconds
Artificial Neural Network	0.526	0.532	01:22:29	01:22:58	29 seconds
ANFIS-DM	0.501	0.510	01:27:44	01:28:31	47 seconds

The root mean squared error is the square root of the variance in the residuals and it indicates the absolute fitness of the model to the data. The model with the least RMSE value indicates the best fit model. Table 3 shows that Linear Regression had RMSE value of 0.561 on training data and 0.579 on testing data, while ANN had RMSE value of 0.526 on training data and 0.532 on testing data and ANFIS-DM had RMSE value of 0.501 on training data and 0.510 on testing data. From the result on Table 3, it was observed that ANFIS-DM had the least RMSE values of 0.501 on training data and 0.510 on testing data, this indicates that the proposed ANFIS-DM performed best compared to the other two models (ANN and Linear Regression). This also suggests that ANFIS-DM will offer the best classification with future data and instances than Linear Regression and ANN.

The time taken to build each model was considered in evaluating the performance of the two models and the proposed ANFIS-DM. It took Linear Regression 10 seconds to build its model, while ANN was built in 29 seconds and the proposed ANFIS-DM took 47 seconds in building its model. Analysing the time taken to build each of the models, it was observed that Linear Regression built its model in 10 seconds and that makes it the fastest running model and used less computing resources compared to ANN and ANFIS-DM. ANN was the next fastest model to run with an execution time of 29 seconds. ANFIS-DM built its model in 47 seconds, and this simply signifies that a lot of computing resources was used and also the slowest of the three models because it took a longer time in building its model compared to Linear Regression and ANN. The performance evaluation on Table 3 shows that ANFIS-DM model was the most accurate among the three models; however, it required more computing resources compared to Linear Regression and ANN.

This research was carried out with the goal of providing an end-user software model that help analyse and mine e-commerce website data to discover hidden patterns that are used in identifying and classifying e-commerce websites into either; legitimate or illegitimate/phishing e-commerce websites. ANFIS-DM provides the best fit solution to new instances based on existing instances. Results obtained from the system were satisfactory. The system can handle instances which may be very complex, since rules can be amended or added to adjust the decision mechanism of the software model.

6. Conclusion

It is noted that ANFIS-DM performed accurately in identifying and classifying the legitimacy and illegitimacy of e-commerce and transactions. A hybrid of CBR and ANFIS was used to develop a software model that assists e-commerce platforms and servers in the analysis of e-commerce data to identify and classify e-commerce websites and transactions into legitimacy or illegitimacy. It should, however, be noted that the system was not designed to prevent transactions or websites from security threat to e-commerce, it was intended only to identify and classify e-commerce websites and transactions as legitimate or illegitimate. Further research in designing the preventive measures is of most importance to the field.

References

Aburrous, M., Hossain, M. A., Dahal, K., & Thabtah, F. (2010). Intelligent Phishing Detection System for E-Banking using Fuzzy Data-Mining. Expert Systems with Applications, Elsevier, 3(7), 7913–7921.

Akinyede, R. O., Alese, B. K., & Adewale, O. S. (2014). Building a Secure Environment for Client-Side E-commerce Payment System Using Encryption System. Proceedings of the planet Congress on Engineering, 1(1), 86-90.

Akinyede, R. O., & Akinyede, A. S. (2015). A Prototype Design for Secure e-commerce Payment System Model. International Journal of Software and Web Science (IJSWS), 12(1), 17-23.

Arinkoola, A. O. (2016). Uncertainty Analysis in Simulation for Reservoir Management: Case Study from Niger Delta. A dissertation submitted to the department of petroleum engineering and the committee on graduate studies of African university of science and technology in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Bandara, R., Fernando, M., & Akter, S. (2019). Privacy Concerns in E-commerce; A Taxonomy and a Future Research Agenda. Electronic Markets, Online First 1-19.

Carta, S., Fenu, G., Recupero, D. R., & Saia, R. (2019). Fraud Detection for E-commerce Transactions by employing a Prudential Multiple Consensus model. Journal of Information Security and Applications, 46(1), 13-22.

Chirag, J., Jignesh, H., & Haresh, K. (2012). Protect the web Consumer's Identity against Attacks by Phishers. International Journal of Computer Application and Knowledge Technology, 1(2), 1014

Fowdur, T. P., & Khader, R. A. (2018). An Anti-Web Phishing Application for Analyzing the Security of Websites. Balkan Journal of Electrical & Computer Engineering, 6(3).

Irvin-Erickson, Y. (2019). Identity Theft and Fraud Victimization: What We Know about Identity Theft and Fraud Victims from Research- and Practice-Based Evidence. Center for Victim Research: Identity Theft and Fraud Victimization

Khraisat, A., Gondal, I., Vamplew, P., & Kamruzzaman, J. (2019). Survey of intrusion detection systems: techniques, datasets and challenges. Cybersecur (2), (20). https://doi.org/10.1186/s42400-019-0038-7

Kishor, N. (2013). E-commerce: Use of Its Common Application. Tactful Management Research Journal, 2(3), 1-6.

Lee, C. H., & Yoon, H. J. (2017). Medical big data: promise and challenges. Kidney research and clinical practice, 36(1), 3–11. https://doi.org/10.23876/j.krcp.2017.36.1.3

Murphy, R. (2018). Common Security Mistakes and How to Avoid Them. CyberShark. https://www.blackstratus.com/avoid-common-security-mistakes/

Niranjanamurthy, M., & Dharmendra, C. (2013). The study of E-Commerce Security Issues and Solutions. International Journal of Advanced Research in Computer and Communication Engineering, 2(7), 1-12.

Phani, A., & Mahaboob, B. (2013). Protecting E-Commerce Systems from Online Fraud. International Journal of Computer Trends and Technology, 4(10), 3549-3554.

Ramachandran, M., & Chang, V. (2016). Towards performance evaluation of cloud service providers for cloud data security. International Journal of Information Management, 36(4), 618-625

Sampson, S. U., Ozumba, S., & Ikpe, J. D. (2019). Adaptive Neuro-Fuzzy Inference System (ANFIS) for Forecasting and Predicting Industrial Electricity Consumption in Nigeria. Horizon Research Publishing, Advances in Energy and Power, 6(3), 23-36.

Shafaei, S. M., Loghavi, M., & Kamgar, S. (2017). Appraisal of Takagi-Sugeno-Kang sort of adaptive neuro-fuzzy inference system for draft force prediction of chisel plow implement. Journal Computers and Electronics in Agriculture archive, 142, 406-415, Elsevier Science Publishers.

Singh, N., Yadav, M., & Sahu, O. (2016). Consumer acceptance of apparel e-commerce– Ethiopia. Intellectual Economics, ISSN: 1822-8011, 10(1), 55-62.

Takagi, T., & Sugeno, M. (1985). Fuzzy identification of systems and its applications to modelling and control. IEEE Trans. Syst. Man. Cybern. Syst, 15(1), 116–132.

Yikun, S., Shijing Y., Kangning L., Kaicheng, H., & Qi, Y. (2019). Developing A Case-Based Reasoning Model for Safety Accident Pre-Control and Decision Making in the Construction Industry. Int. J. Environ. Res. Public Health, 16(9), 1511; https://doi.org/10.3390/ijerph16091511

Zhai, Z., Jose-Fernan, M. O., Castillejo, P., & Beltran, V. (2019). A Triangle Similarity Measure for Case Retrieval in CBR and its Application to an Agricultural Decision Support System. Multidisciplinary Digital Publishing Institute, Sensors (Basel), 19(21), 4605.