European Journal of Information Technologies and Computer Science
www.ej-compute.org

A Multi-Layer Perceptron Model for Classification of E-mail Fraud

Temitayo O. Oyegoke, Kehinde K. Akomolede, Adesola G. Aderounmu,
and Emmanuel R. Adagunodo

ABSTRACT

This study was developed an e-mail classification model to preempt fraudulent activities. The e-mail has such a predominant nature that makes it suitable for adoption by cyber-fraudsters. This research used a combination of two databases: CLAIR fraudulent and Spambase datasets for creating the training and testing dataset. The CLAIR dataset consists of raw e-mails from users' inbox which were pre-processed into structured form using Natural Language Processing (NLP) techniques. This dataset was then consolidated with the Spambase dataset as a single dataset. The study deployed the Multi-Layer Perceptron (MLP) architecture which used a back-propagation algorithm for training the fraud detection model. The model was simulated using 70% and 80% for training while 30% and 20% of datasets were used for testing respectively. The results of the performance of the models were compared using a number of evaluation metrics. The study concluded that using the MLP, an effective model for fraud detection among e-mail dataset was proposed.

Keywords: Spam mail, machine learning, advance fee fraud; fraud detection, neural network.

Published Online: December 7, 2021

ISSN: 2736-5492

DOI: 10.24018/ejcompute.2021.1.5.24

T. O. Oyegoke*
Department of Computer Science and Engineering, Obafemi Awolowo University, Ile-Ife, Osun State, Nigeria.
(e-mail: tooyegoke@oauife.edu.ng)

K. K. Akomolede
Department of Computer Science, The Federal Polytechnic, Ado Ekiti, Ekiti State, Nigeria.
(e-mail: akomskenny@gmail.com)

G. A. Aderounmu
Department of Computer Science and Engineering, Obafemi Awolowo University, Ile-Ife, Osun State, Nigeria.
(e-mail: gaderoun@oauife.edu.ng)

E. R. Adagunodo
Department of Computer Science and Engineering, Obafemi Awolowo University, Ile-Ife, Osun State, Nigeria.
(e-mail: eadagun@oauife.edu.ng)

Corresponding Author

I. Introduction

Reference [1] defined fraud as any action that is surreptitiously undertaken by one party called the perpetrator for the aim of obtaining an unfair advantage over another called the victim. Advance Fee Fraud (AFF) commonly known as 419, is a fraudulent process of enticing a victim with a bogus business proposal, with a promise to transfer large sums of money, usually in foreign exchange, purported to be part of the proceeds of certain contracts, to the addressee's bank account to be shared in some proportion between the parties [2], [3]. Among the different means by which fraudulent crimes are perpetuated, the email has remained the most popular among the techniques used to target innocent victims for fraud [4]. The rapid growth of the Internet which has led to a significant increase in the number of email users at the same time has also led to an increase in spam emails rate by fraudsters. A statistical report showed that 70% of the email in circulation during the second week of 2014 was spam or illegitimate emails [5].

Fraud detection protects customer and enterprise information, assets, accounts, and transactions through the real-time, near-real-time or batch analysis of activities by users and other defined entities [6]. Fraud detection uses background server-based processes that examine users' and other defined entities' access and behavior patterns, and typically compares this information with the profile of what is expected. Fraud detection is not intrusive to a user unless the user's activity is suspected, however it tries to detect and recognize fraudulent activities entering the systems in order to report them to the system manager.

Various research has adopted artificial intelligence and statistics in the development of fraud detection, many of which have a centralized control leading to difficulty in understanding the changing dynamics of fraudulent environments [7]. Artificial intelligence and statistical techniques applied to fraud detection includes but not limited to Naıve Bayes classifiers for email spam detection [8], [9] and network intrusion detection [10]; Support vector machines for email spam detection [11], phishing detection [12] and network intrusion detection [13] and Learning vector quantization [14], fuzzy association rules [15], agent-based IDS [16], and AdaBoost algorithm for network intrusion detection. Other techniques include Hidden Markov model [17], association rules [18], Dempster–Shafer theory [19] and Statistical methods [20].

Fraudulent activities are increasing at an alarming rate thereby causing great damage to the integrity of a majority of honest individuals. The email has also proven to be the most popular means of perpetuating fraudulent activities such as advance fee fraud and is responsible for financial losses of up to $6 billion in the last 2 years. The detection of fraudulent e-mails provides a means of safeguarding the interests of unsuspecting online users by providing a real-time and early detection of fraudulent e-mails on arrival. In this paper, an attempt was made to construct a dataset suitable for the detection of fraud from the content of e-mails following which a supervised machine learning algorithm was used to formulate the fraud detection model required for identifying fraudulent mails.

A review of related works revealed that a number of them were more concerned with the detection of spam mails and e-commerce fraud. A number of the reviewed papers are presented in the following paragraphs.

Reference [21] worked on the detection of spam mails using a combination of machine learning algorithms. Sample emails were collected from the UCI Database and pre-processed in order to remove redundant and root words. Relevant features were selected from the initially identified features using particle swarm optimization (PSO) and the Multi-Layer Perceptron (MLP). The features extracted were used to develop spam detection model using the support vector machines (SVM).

Reference [22] worked on the automatic classification of e-mails using genetic algorithm. The study adopted the use of 500 Sample e-mail data consisting of 300 spam mails and 200 ham mails from which keywords were extracted following the removal of articles and numerical texts. The email classification model was formulated using the genetic algorithm. The results of the study showed that the genetic algorithm was able to identify spam mails for the dataset using the data dictionary with an accuracy of 81%.

Reference [23] worked on the cost implication of credit card fraud. Data was collected from a database consisting of 80 million individual credit and debit card transactions containing 27 attribute data from which were selected 14 features based on the decision of the card processing company risk team. The results of the study showed that the Bayes minimum risk model was able to identify the risk associated with credit card fraud. The best solution showed a cost of 36,634 Euros owing for 23% savings of initial cost.

Reference [24] worked on the review of Fraud Detection Techniques: Credit Card. The study performed a review of techniques applied to credit card fraud detection highlighting the advantages and disadvantages of each technique. The results of the study showed that a number of novel techniques have been applied to the detection of credit card fraud among which were naïve Bayes, neural network, genetic algorithms and outlier detection methods.

II. METHODOLOGY

This section presents the materials and methods that were deployed for the development of the classification model required for the detection of fraudulent e-mails. The materials and methods used for data identification and collection, model formulation and simulation alongside performance evaluation were also presented.

A. Method of Data Identification and Collection

The dataset required for this study are fraudulent, spam and ham mails. However, the most commonly available dataset only contained records of spam and ham mails. As a result of this, an additional dataset was also identified which consisted of fraudulent e-mail (unstructured data). The e-mail contents of the dataset were preprocessed by applying tokenization for extracting e-mails contents, stop word removal for removing prepositions and articles, and stemming for converting word variations to their base words. This was done in order to extract the features within the mail for conversion into a structured dataset and combined with an existing dataset consisting of spam and ham mail records.

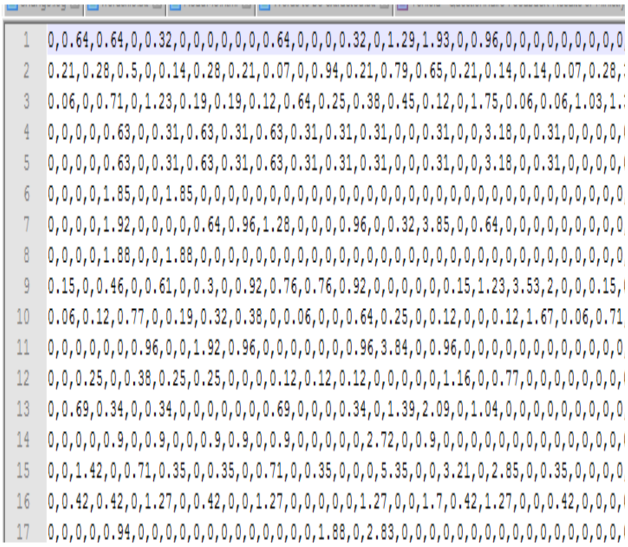
The Spambase dataset which consisted of records of spam and ham alone was collected from the UCI Machine Learning repository at https://archive.ics.uci.edu/ml/datasets/spambase. The dataset existed in structured form consisting of the features extracted alongside the class label classified spam and ham which were identified by the value 1 and 0 respectively. The dataset consisted of 4601 records composed of 1813 spam and 2788 ham mails identified using 57 continuous-valued features which identified the value of the term frequency-inverse document frequency (TF-IDF) of each feature which is in Fig. 1.

The CLAIR dataset which consisted of 3974 records of fraudulent e-mails was assessed from the Kaggle website at https://www.kaggle.com/rtatman/fraudulent-email-corpus. Each record consisted of the e-mail header and the body of the e-mail containing the text. Unlike the Spambase dataset, it existed in a raw and unstructured format that was required to be pre-processed using tokenization, stop-word removal and stemming in order to convert the e-mails into a structured format similar to the Spambase dataset. A sample e-mail record is presented in Fig. 2.

B. Extraction of Feature Set from Dataset

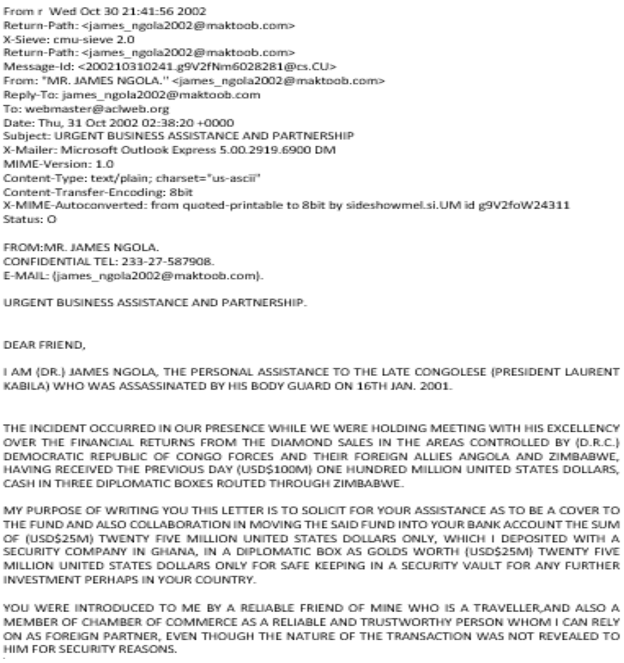
For the purpose of the development of the fraud detection model for emails, there was the need of converting the unstructured contents within the CLAIR emails collected into a structured format similar to the Spambase features. The features extracted are the words that were found within the body of the CLAIR email contents. The text preprocessing of emails required the use of NLP tools for the purpose of performing the different text preprocessing stages required for converting the unstructured data into a structured format.

The Tokenizer was used to divide sentences in each e-mail into the words found in the sentence. Stop word removal was performed in order to remove all punctuation marks, pronouns, prepositions and determinants present in the tokenized words. The stemmer was required for reducing variation of a particular word to a single word for the purpose of reducing all inflexed word forms such as activity, activate and active which can be reduced to activ or *families, familiar* which becomes *famili* and so on. The stemming algorithm used was the Porter's Stemmer. The features that were extracted from the CLAIR dataset were used to develop a Term-Document Frequency feature set which contained the identified words on each column with a set of values for each e-mail record.



```
0,0.64,0.64,0,0.32,0,0,0,0,0,0,0.64,0,0,0,0.32,0,1.29,1.93,0,0.96,0,0,0,0,0,0,0,0
0.21,0.28,0.5,0,0.14,0.28,0.21,0.07,0,0.94,0.21,0.79,0.65,0.21,0.14,0.14,0.07,0.28,
0.06,0,0.71,0,1.23,0.19,0.19,0.12,0.64,0.25,0.38,0.45,0.12,0,1.75,0.06,0.06,1.03,1.
0,0,0,0,0.63,0,0.31,0.63,0.31,0.63,0.31,0.31,0.31,0,0,0.31,0,0,3.18,0,0.31,0,0,0,0,
0,0,0,0,0.63,0,0.31,0.63,0.31,0.63,0.31,0.31,0.31,0,0,0.31,0,0,3.18,0,0.31,0,0,0,0,
0,0,0,0,1.85,0,0,1.85,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0
0,0,0,0,1.92,0,0,0,0,0.64,0.96,1.28,0,0,0,0.96,0,0.32,3.85,0,0.64,0,0,0,0,0,0,0,0,0
0,0,0,0,1.88,0,0,1.88,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0
0.15,0,0.46,0,0.61,0,0.3,0,0.92,0.76,0.76,0.92,0,0,0,0,0,0.15,1.23,3.53,2,0,0,0.15,
0.06,0.12,0.77,0,0.19,0.32,0.38,0,0.06,0,0,0.64,0.25,0,0.12,0,0,0.12,1.67,0.06,0.71
0,0,0,0,0,0,0.96,0,0,1.92,0.96,0,0,0,0,0,0,0.96,3.84,0,0.96,0,0,0,0,0,0,0,0,0,0,0,0
0,0,0.25,0,0.38,0.25,0.25,0,0,0,0.12,0.12,0.12,0,0,0,0,0,1.16,0,0.77,0,0,0,0,0,0,0,
0,0.69,0.34,0,0.34,0,0,0,0,0,0,0.69,0,0,0,0.34,0,1.39,2.09,0,1.04,0,0,0,0,0,0,0,0,0
0,0,0,0,0.9,0,0.9,0,0,0.9,0.9,0,0.9,0,0,0,0,0,2.72,0,0.9,0,0,0,0,0,0,0,0,0,0,0,0,
0,0,1.42,0,0.71,0.35,0,0.35,0,0.71,0,0.35,0,0,0,5.35,0,0,3.21,0,2.85,0,0.35,0,0,0,0
0,0.42,0.42,0,1.27,0,0.42,0,0,1.27,0,0,0,0,0,1.27,0,0,1.7,0.42,1.27,0,0,0.42,0,0,0,
0,0,0,0,0.94,0,0,0,0,0,0,0,0,0,0,0,0,0,1.88,0,2.83,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,
```

Fig. 1. Spambase dataset.



From r Wed Oct 30 21:41:56 2002
Return-Path: <james_ngola2002@maktoob.com>
X-Sieve: cmu-sieve 2.0
Return-Path: <james_ngola2002@maktoob.com>
Message-Id: <200210310241.g9V2fNm6028281@cs.CU>
From: "MR. JAMES NGOLA." <james_ngola2002@maktoob.com>
Reply-To: james_ngola2002@maktoob.com
To: webmaster@aclweb.org
Date: Thu, 31 Oct 2002 02:38:20 +0000
Subject: URGENT BUSINESS ASSISTANCE AND PARTNERSHIP
X-Mailer: Microsoft Outlook Express 5.00.2919.6900 DM
MIME-Version: 1.0
Content-Type: text/plain; charset="us-ascii"
Content-Transfer-Encoding: 8bit
X-MIME-Autoconverted: from quoted-printable to 8bit by sideshowmel.si.UM id g9V2foW24311
Status: O

FROM:MR. JAMES NGOLA.
CONFIDENTIAL TEL: 233-27-587908.
E-MAIL: (james_ngola2002@maktoob.com).

URGENT BUSINESS ASSISTANCE AND PARTNERSHIP.

DEAR FRIEND,

I AM (DR.) JAMES NGOLA, THE PERSONAL ASSISTANCE TO THE LATE CONGOLESE (PRESIDENT LAURENT KABILA) WHO WAS ASSASSINATED BY HIS BODY GUARD ON 16TH JAN. 2001.

THE INCIDENT OCCURRED IN OUR PRESENCE WHILE WE WERE HOLDING MEETING WITH HIS EXCELLENCY OVER THE FINANCIAL RETURNS FROM THE DIAMOND SALES IN THE AREAS CONTROLLED BY (D.R.C.) DEMOCRATIC REPUBLIC OF CONGO FORCES AND THEIR FOREIGN ALLIES ANGOLA AND ZIMBABWE, HAVING RECEIVED THE PREVIOUS DAY (USD$100M) ONE HUNDRED MILLION UNITED STATES DOLLARS, CASH IN THREE DIPLOMATIC BOXES ROUTED THROUGH ZIMBABWE.

MY PURPOSE OF WRITING YOU THIS LETTER IS TO SOLICIT FOR YOUR ASSISTANCE AS TO BE A COVER TO THE FUND AND ALSO COLLABORATION IN MOVING THE SAID FUND INTO YOUR BANK ACCOUNT THE SUM OF (USD$25M) TWENTY FIVE MILLION UNITED STATES DOLLARS ONLY, WHICH I DEPOSITED WITH A SECURITY COMPANY IN GHANA, IN A DIPLOMATIC BOX AS GOLDS WORTH (USD$25M) TWENTY FIVE MILLION UNITED STATES DOLLARS ONLY FOR SAFE KEEPING IN A SECURITY VAULT FOR ANY FURTHER INVESTMENT PERHAPS IN YOUR COUNTRY.

YOU WERE INTRODUCED TO ME BY A RELIABLE FRIEND OF MINE WHO IS A TRAVELLER,AND ALSO A MEMBER OF CHAMBER OF COMMERCE AS A RELIABLE AND TRUSTWORTHY PERSON WHOM I CAN RELY ON AS FOREIGN PARTNER, EVEN THOUGH THE NATURE OF THE TRANSACTION WAS NOT REVEALED TO HIM FOR SECURITY REASONS.

Fig. 2. Example of e-mail from Clair dataset.

The values were represented by either a 0 or 1 such that whenever a word was found in a document then a feature value of 1 was assigned otherwise a 0 was assigned. For the Spambase dataset, each feature value was converted to a binary value by ensuring that whenever the value of the TF-IDF of each term was greater than 0 then a value of 1 was adopted and when 0 a value of 0 was adopted. Therefore, the two dataset (Spambase and CLAIR) consists of a feature set of words with values of either 0 or 1 whenever a word was not found or not found in an email with a target class label of either spam, ham, or fraud. By combining the Spambase and CLAIR dataset, a total of 8575 e-mail records were collected.

C. Data Preparation

In order to convert the preprocessed words in the emails collected from CLAIR dataset into a structured form, the words identified from both email samples (CLAIR and Spambase) were used to form a term-document matrix, D_{ij} which represents the occurrence of each term/word w_i within each e-mail d_j. In a term-document matrix D_{ij} the rows represent the presence (or absence) of a word w_i in a document D_j while each column identifies the words been identified in each document. There was the need to identify the features that are unique to CLAIR e-mails (A), identifying the features that are unique to Spambase e-mails (B) and identifying the features that are common to Spambase and CLAIR e-mails ($C=A\cap B$).

The features that were extracted from the CLAIR dataset were compared with those extracted from Spambase dataset in order to identify the features that were unique to CLAIR dataset (A), the features that were unique to Spambase dataset (B) and the features that were common to CLAIR and Spambase dataset $C=A\cap B$. The 3 set of features were combined into a single feature set $A\cup B\cup C$. Fig. 3 shows a Venn diagram which gives the description of how the features extracted from both datasets were combined.

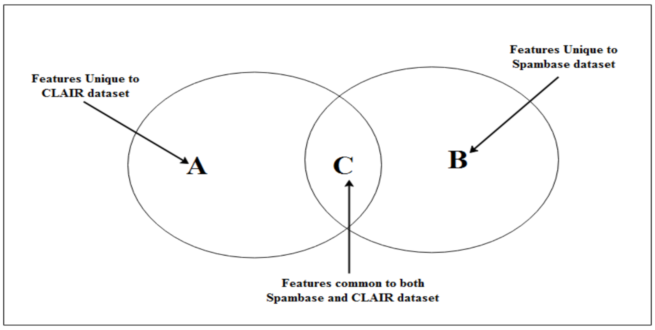


Fig. 3. Consolidation of Dataset Features.

It can be seen in Fig. 3 that the features present in the CLAIR dataset ($A\cup$C) features unique to CLAIR (A) and those common to Spambase dataset (C) while features present in Spambase ($B\cup C$) contains features unique to Spambase (B) and those common to CLAIR dataset (C). Following the identification of the features that were common and unique to both datasets, the features that were present and absent in the CLAIR dataset and in the Spambase dataset were identified by parsing the extracted features through each e-mail.

Following the process of creating the term-document matric for all 8575 email samples collected, the class of each email was used to map each row of the term-document matrix to either of Ham, Spam and Fraud. Therefore, the term-document matrix created in this study consisted of 8575 rows of documents D_j and n columns of features/words/terms extracted from emails. A function f which represents the fraud classification model was required for mapping the n words/features to the target class with the values Ham, Spam and Fraudulent according to the relationship shown in (1).

Defined as:

$$f(w_1, w_2, \dots \dots \dots, w_n\ ) = \begin{cases} Ham \\ Spam \\ Fraud \end{cases} \qquad (1)$$

D. Model Formulation and Simulation

The fraud detection model proposed in this study was formulated using the MLP architecture of the artificial neural network (ANN). The model required the use of neurons which were interconnected by weights which specify the

DOI: http://dx.doi.org/10.24018/ejcompute.2021.1.5.24

strength of connection between nodes. Also, neurons were arranged into layers such that neurons from previous layers were connected to other layers ahead using weights which connected them to activation functions which were used to produce normalized outputs. Therefore, the MLP model adopted for this study consisted of n input nodes which were used to capture data from the input features (words) to nodes in subsequent hidden layers using activation functions through to the output layer which consisted of 3 nodes for spam, ham and fraud outputs. The MLP made use of the back-propagation algorithm for determining the optimal weights required for creating optimal paths from the input nodes to the output node for the purpose of fraud detection.

The formulation of the back-propagation algorithm involved the use of the Google drive for the movement of e-mail data in and out of the process and the TensorFlow back-end for modeling while 2 MLP models with 2 layers were constructed. The multi-layer perceptron model generated in this study was modeled using input neurons equal to the number of attributes 'n' in the dataset which represents the initially extracted 12831 words/features from the e-mails with values 0 or 1. The input neurons (i1 to iN) were attached to the first layer using n weights connecting previous nodes to nodes in subsequent layers. In all there were 20 neurons (r1 to r20) in the first layer connected to 10 neurons (r1 to r10) which were also in turn connected 3 neurons (Ham, Spam and Fraud) for the multi-layer perceptron network as shown in Fig. 4.

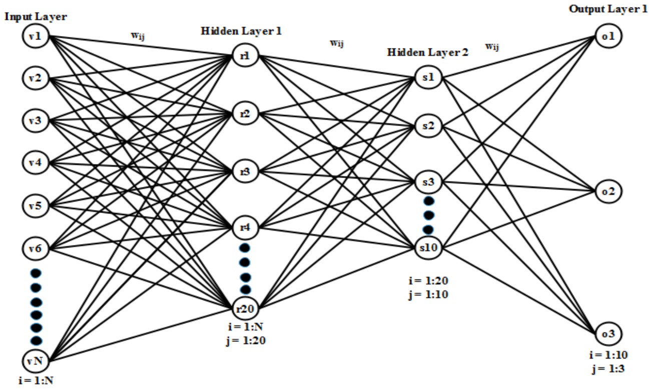


Fig. 4. Multi-Layer Perceptron Network for Fraud Detection.

Therefore, 20 sets of n weights were attached to n inputs to the neurons in the first layer of the MLP network. The second hidden layer had 10 neurons (s1 to s10) to which were attached 10 sets of 20 weights coming from the output of the first hidden layer. Finally, the output of the second hidden layer were attached to three output layers (o1 to o3) namely: Fraud, Ham and Spam. The outputs produced by each neuron in the hidden layer and outer later were created using activation functions, namely: the rectifier linear unit (ReLU) and the Softmax function also called the normalized exponential function. The values provided at the output nodes were used to determine the error in prediction which is required by the back propagation for weight update.

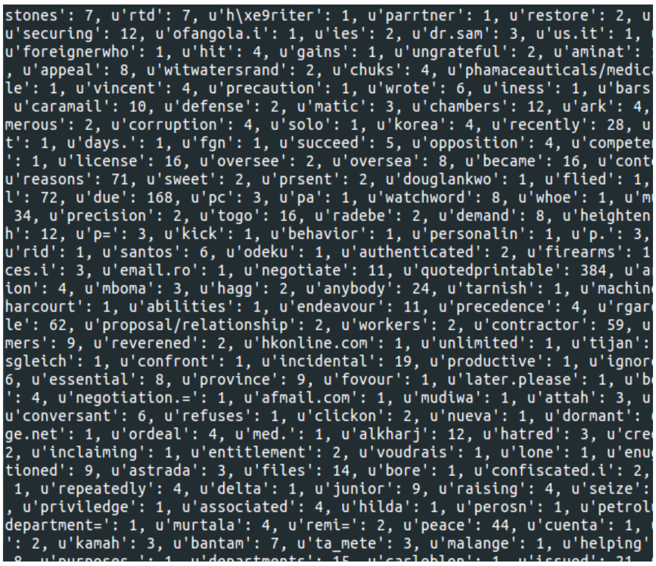
The mean square error of the prediction made in comparison to the actual values recorded in the validation dataset was used to adjust the values of the weights based on the results of the back propagation algorithm. The weights attached to the nodes from the output layer through the hidden layers to the input layers were adjusted following which another iteration is performed in order to perform further adjustments to the model. This process continued until there was no more error detected between the predicted and actual values of the validation dataset such that there was no need to adjust the weights. The final model developed following this procedure was then validated in order to determine the performance of the fraud detection model.

III. Results

This section presents the results of the identification of the initial features which were extracted from the sample e-mails used for this study using the Python Natural Language Toolkit (NLTK). It further presents the results of the back-propagation algorithm for the selection of optimal weights alongside the evaluation of the model performance.

A. *Results of Extraction of E-mail Features*

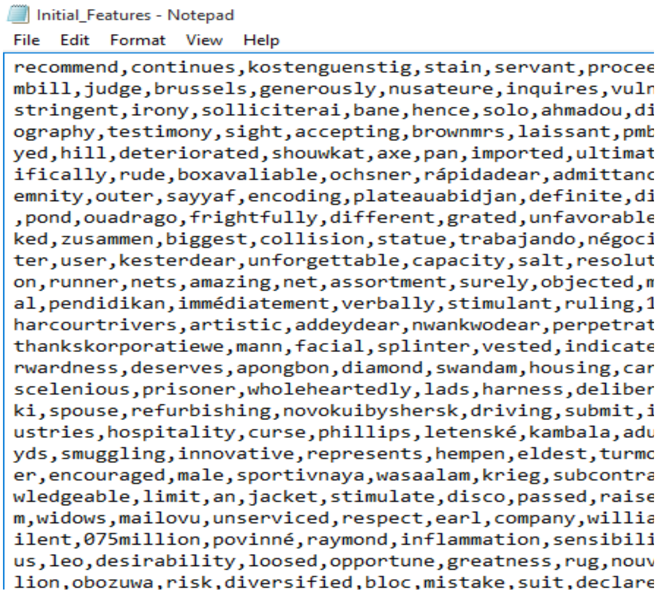
Following the process of tokenization, stop word removal and stemming, a total of 12831 features were extracted and parsed into each e-mail for generating a dataset containing the extracted features of the CLAIR dataset alongside their frequency as shown in Fig. 5. The information about the terms that were extracted from CLAIR e-mails alongside their frequencies were collected. The features were presented as columns and the value of 0 or 1 for each feature if the frequency of the term is 0 or greater than 0 was presented in the rows. This dataset following pre-processing were later consolidated with the Spambase dataset as a single dataset required for analysis was generated.



```
stones': 7, u'rtd': 7, u'h\xe9riter': 1, u'parrtner': 1, u'restore': 2, u
u'securing': 12, u'ofangola.i': 1, u'ies': 2, u'dr.sam': 3, u'us.it': 1,
u'foreignerwho': 1, u'hit': 4, u'gains': 1, u'ungrateful': 2, u'aminat':
, u'appeal': 8, u'witwatersrand': 2, u'chuks': 4, u'phamaceauticals/medic
le': 1, u'vincent': 4, u'precaution': 1, u'wrote': 6, u'iness': 1, u'bars
 u'caramail': 10, u'defense': 2, u'matic': 3, u'chambers': 12, u'ark': 4,
merous': 2, u'corruption': 4, u'solo': 1, u'korea': 4, u'recently': 28, u
t': 1, u'days.': 1, u'fgn': 1, u'succeed': 5, u'opposition': 4, u'compete
': 1, u'license': 16, u'oversee': 2, u'oversea': 8, u'became': 16, u'cont
u'reasons': 71, u'sweet': 2, u'prsent': 2, u'douglankwo': 1, u'flied': 1,
l': 72, u'due': 168, u'pc': 3, u'pa': 1, u'watchword': 8, u'whoe': 1, u'm
 34, u'precision': 2, u'togo': 16, u'radebe': 2, u'demand': 8, u'heighten
h': 12, u'p=': 3, u'kick': 1, u'behavior': 1, u'personalin': 1, u'p.': 3,
u'rid': 1, u'santos': 6, u'odeku': 1, u'authenticated': 2, u'firearms': 1
ces.i': 3, u'email.ro': 1, u'negotiate': 11, u'quotedprintable': 384, u'a
ion': 4, u'mboma': 3, u'hagg': 2, u'anybody': 24, u'tarnish': 1, u'machin
harcourt': 1, u'abilities': 1, u'endeavour': 11, u'precedence': 4, u'rgar
le': 62, u'proposal/relationship': 2, u'workers': 2, u'contractor': 59, u
mers': 9, u'reverened': 2, u'hkonline.com': 1, u'unlimited': 1, u'tijan':
sgleich': 1, u'confront': 1, u'incidental': 19, u'productive': 1, u'ignor
6, u'essential': 8, u'province': 9, u'fovour': 1, u'later.please': 1, u'b
': 4, u'negotiation.=': 1, u'afmail.com': 1, u'mudiwa': 1, u'attah': 3, u
u'conversant': 6, u'refuses': 1, u'clickon': 2, u'nueva': 1, u'dormant':
ge.net': 1, u'ordeal': 4, u'med.': 1, u'alkharj': 12, u'hatred': 3, u'cre
2, u'inclaiming': 1, u'entitlement': 2, u'voudrais': 1, u'lone': 1, u'enu
tioned': 9, u'astrada': 3, u'files': 14, u'bore': 1, u'confiscated.i': 2,
 1, u'repeatedly': 4, u'delta': 1, u'junior': 9, u'raising': 4, u'seize':
, u'priviledge': 1, u'associated': 4, u'hilda': 1, u'perosn': 1, u'petrol
department=': 1, u'murtala': 4, u'remi=': 2, u'peace': 44, u'cuenta': 1,
': 2, u'kamah': 3, u'bantam': 7, u'ta_mete': 3, u'malange': 1, u'helping'
```

Fig. 5. Terms extracted from CLAIR Dataset using Tokenization.

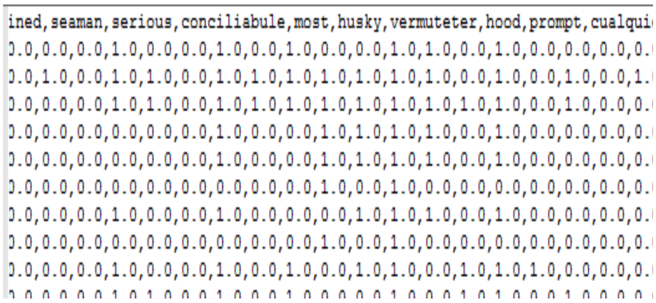
However, the Spambase dataset contained structured data consisting of the features in the Spambase e-mails alongside their respective TF-IDF (which was used to convert the Spambase into a binary dataset). The Spambase dataset contained 2788 spam mails and 1813 ham mails. The final list of features that were extracted from both datasets were 8575 and a screenshot of some of these features is presented in Fig. 6.



```
recommend,continues,kostenguenstig,stain,servant,procee
mbill,judge,brussels,generously,nusateure,inquires,vuln
stringent,irony,solliciterai,bane,hence,solo,ahmadou,di
ography,testimony,sight,accepting,brownmrs,laissant,pmb
yed,hill,deteriorated,shouwkat,axe,pan,imported,ultimat
ifically,rude,boxavaliable,ochsner,rápidadear,admittanc
emnity,outer,sayyaf,encoding,plateauabidjan,definite,di
,pond,ouadrago,frightfully,different,grated,unfavorable
ked,zusammen,biggest,collision,statue,trabajando,négoci
ter,user,kesterdear,unforgettable,capacity,salt,resolut
on,runner,nets,amazing,net,assortment,surely,objected,m
al,pendidikan,immédiatement,verbally,stimulant,ruling,1
harcourtrivers,artistic,addeydear,nwankwodear,perpetrat
thankskorporatiewe,mann,facial,splinter,vested,indicate
rwardness,deserves,apongbon,diamond,swandam,housing,car
scelenious,prisoner,wholeheartedly,lads,harness,deliber
ki,spouse,refurbishing,novokuibyshersk,driving,submit,i
ustries,hospitality,curse,phillips,letenské,kambala,adu
yds,smuggling,innovative,represents,hempen,eldest,turmc
er,encouraged,male,sportivnaya,wasaalam,krieg,subcontra
wledgeable,limit,an,jacket,stimulate,disco,passed,raise
m,widows,mailovu,unserviced,respect,earl,company,willia
ilent,075million,povinné,raymond,inflammation,sensibili
us,leo,desirability,loosed,opportune,greatness,rug,nouv
lion,obozuwa,risk,diversified,bloc,mistake,suit,declare
```

Fig. 6. Features extracted from Both Dataset.

Fig. 7 shows the final structured dataset that was extracted from both datasets. This contains the finally reconciled features required for the formulation of the fraud detection model. The results of feature extraction from the e-mails were followed by the results of the formulation of the back-propagation algorithm required for the detection of fraudulent e-mails. The result of the model formulation is presented in the following section.

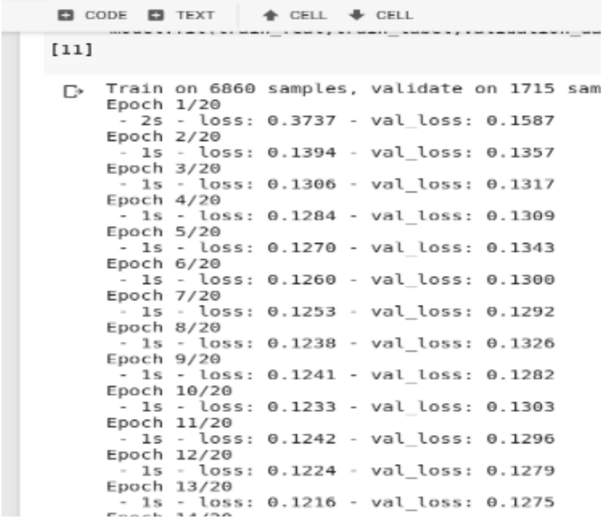


```
ined,seaman,serious,conciliabule,most,husky,vermuteter,hood,prompt,cualqui
0.0,0.0,0.0,1.0,0.0,0.0,1.0,0.0,1.0,0.0,0.0,1.0,1.0,0.0,1.0,0.0,0.0,0.0,0.
0.0,1.0,0.0,1.0,1.0,0.0,1.0,1.0,1.0,1.0,1.0,1.0,1.0,0.0,1.0,0.0,1.0,0.0,1.
0.0,0.0,0.0,1.0,1.0,0.0,1.0,1.0,1.0,1.0,1.0,1.0,1.0,1.0,1.0,0.0,1.0,0.0,0.
0.0,0.0,0.0,0.0,0.0,0.0,1.0,0.0,0.0,1.0,1.0,1.0,1.0,0.0,1.0,0.0,0.0,0.0,0.
0.0,0.0,0.0,0.0,0.0,0.0,1.0,0.0,0.0,1.0,1.0,1.0,1.0,0.0,1.0,0.0,0.0,0.0,0.
0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,1.0,0.0,1.0,0.0,0.0,0.0,0.0,0.0,0.0,0.
0.0,0.0,0.0,1.0,0.0,0.0,1.0,0.0,0.0,0.0,1.0,1.0,1.0,0.0,1.0,0.0,0.0,0.0,0.
0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,1.0,0.0,1.0,0.0,0.0,0.0,0.0,0.0,0.0,0.
0.0,0.0,0.0,1.0,0.0,0.0,1.0,0.0,1.0,0.0,1.0,1.0,0.0,1.0,1.0,1.0,0.0,0.0,0.
```

Fig. 7. Final Dataset containing Reconciled Features from both Datasets.

B. Results of Formulation and Simulation of Fraud Detection Model

The results of the process of model formulation and simulation for fraud detection involved the use of the dataset containing the extracted 12831 features. The MLP network which depended on the back-propagation algorithm was developed using 2 simulations. The experiment was performed by using the dataset in such a manner that the dataset was split into training/testing dataset proportion of (70/30) % and (80/20) % such that there were 6002/2573 and 6860/1715 for training and testing respectively. Fig. 8 shows a screenshot of the interface of the simulation process using the Python machine learning library for 20 epochs.



```
Train on 6860 samples, validate on 1715 sam
Epoch 1/20
 - 2s - loss: 0.3737 - val_loss: 0.1587
Epoch 2/20
 - 1s - loss: 0.1394 - val_loss: 0.1357
Epoch 3/20
 - 1s - loss: 0.1306 - val_loss: 0.1317
Epoch 4/20
 - 1s - loss: 0.1284 - val_loss: 0.1309
Epoch 5/20
 - 1s - loss: 0.1270 - val_loss: 0.1343
Epoch 6/20
 - 1s - loss: 0.1260 - val_loss: 0.1300
Epoch 7/20
 - 1s - loss: 0.1253 - val_loss: 0.1292
Epoch 8/20
 - 1s - loss: 0.1238 - val_loss: 0.1326
Epoch 9/20
 - 1s - loss: 0.1241 - val_loss: 0.1282
Epoch 10/20
 - 1s - loss: 0.1233 - val_loss: 0.1303
Epoch 11/20
 - 1s - loss: 0.1242 - val_loss: 0.1296
Epoch 12/20
 - 1s - loss: 0.1224 - val_loss: 0.1279
Epoch 13/20
 - 1s - loss: 0.1216 - val_loss: 0.1275
```

Fig. 8. Screenshot of Training and Testing Errors.

The Fig. 8 presented the time taken for each model to be built, the error made by the training data and the error made by the testing dataset for each epoch/iteration. The simulation was completed when the lowest possible error rate was determined during the iterations. The screenshot shows that the simulation had the error rate at the 14th epoch thus terminated and returned the model simulated for validation. The results of the error rate for the training and testing datasets presented in Fig. 8 is presented the plot in Fig. 9. This shows the behavior of the model using the training and testing dataset for model validation for model performance evaluation.

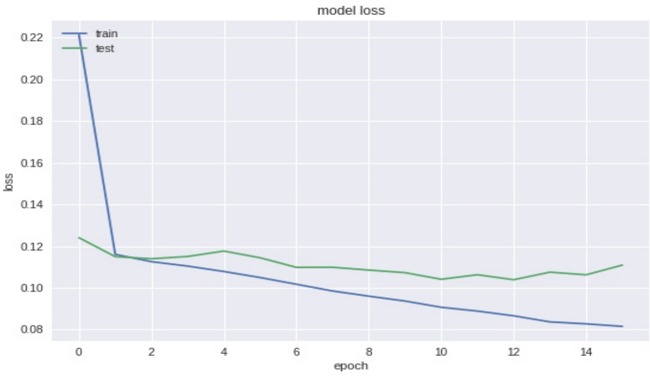
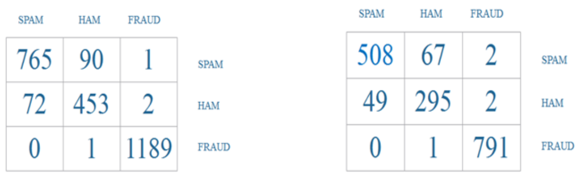


Fig. 9. Graphical Plot of Training and Testing Errors.

The plot showed that the model attained lower error rate within the training dataset compared to within the testing dataset due to the fact that the training dataset was adopted for model development. Since the dataset collected had been split into 2 sets of training/testing dataset 70/30 and 80/20, both datasets were adopted for the simulation and validation of the fraud detection algorithm for this study. In the dataset consisting of 30% testing dataset, there was 837 Spam (32.3% of Spam), 544 ham (21.14% of Ham) and 1192 Fraud (46.33%) e-mails which totals to 2573 (30% of datasets) while in the dataset containing the 20% testing dataset there were 557 Spam (32.49%), 363 Ham (21.17%) and 795 Fraud (46.36%) e-mail datasets.

C. Results of the Evaluation of Performance of the Model

By using the 30% of the dataset containing the extracted features from e-mails to build the MLP model the following were observed; out of the actual 837 spam mails, 765 were correctly classified while 72 were incorrectly classified as ham mails respectively; out of the actual 544 Ham mails, 453 were correctly classified while 90 and 1 were incorrectly classified as Spam and Fraud e-mails respectively: while out of the actual 1192 Fraud e-mail, 1189 were correctly classified while 1 and 2 were incorrectly classified as Spam and ham mail respectively. In total there were 2407 correct classification out of 2575 owing for an accuracy of 93.55% as shown in the confusion matrix in Fig. 10 (Up).



SPAM	HAM	FRAUD	
765	90	1	SPAM
72	453	2	HAM
0	1	1189	FRAUD

SPAM	HAM	FRAUD	
508	67	2	SPAM
49	295	2	HAM
0	1	791	FRAUD

Fig. 10. Results of Simulation using 30% (Left) and 20% (Right) Testing Dataset.

By using the 20% of the dataset containing the originally extracted features from e-mails to build the gradient-based BP algorithm the following were observed: out of the actual 557 spam mails, 508 were correctly classified while 49 were incorrectly classified as ham mails respectively; out of the actual 363 Ham mails, 295 were correctly classified while 67 and 1 were incorrectly classified as Spam and Fraud e-mails respectively; while out of the actual 795 Fraud e-mail, 791 were correctly classified while 2 and 2 were incorrectly classified as Spam and ham mail respectively. In total there were 1594 correct classification out of 1715 owing for an accuracy of 92.94% as shown in the confusion matrix in Fig. 10 (Down).

IV. Discussions

The results of the simulation and validation of the classification models developed were presented on a 3 by 3 confusion matrix which presented the correct and incorrect Table I: Results of the Evaluation of the Performance of the Model classifications made during model validation. The results presented in the confusion matrix was used to determine the values of the performance evaluation metrics that were used as a basis of choosing the most effective classification model for fraud detection.

As recalled earlier, 70% and 80% of the collected dataset consisting of 6000 and 6860 e-mails respectively were used for training all the models that were simulated while 30% and 20% of the dataset consisting of the remaining 2575 and 1715 e-mails were used for testing the model during model validation based on the results presented in the confusion matrix. The results of the study showed that MLP model for fraud detection developed using 30% of datasets for testing had a better performance compared to using the 20% of dataset for testing as shown in Table I. It was observed that the accuracy of the model was increased by a value of 0.61% owing to the ability to classify 2407 e-mails correctly out of 2575 e-mails.

The results of the True Positive (TP) rate which revealed the proportion of actual e-mail correctly classified revealed that the model was able to correctly identify 91.4%, 83.3% and 99.7% of the Spam, Ham and Fraudulent e-mails.

The results of the False Positive (FP) rate which revealed the proportion of e-mails incorrectly classified showed that only 5.2% of Ham/Fraud mails were misclassified as Spam, 3.6% of Spam/Fraud mails were misclassified as Ham while 0.1% of Spam/Ham were misclassified as Fraud. The Precision which showed the proportion of predicted mails which were correct showed that 89.4%, 86% and 99.9% of mails predicted as Spam, Ham and Fraud were correctly classified. Fig. 11 shows a graphical plot of the performance of the fraud detection model proposed for this study.

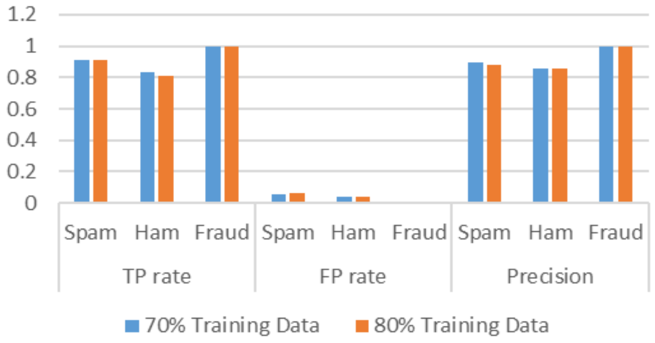


Fig. 11. Graphical Plot of Performance Evaluation.

The study showed that adopting the use of the fraud detection model formulated and simulated using the 30% testing dataset, a better performance compared to the model developed using the 20% testing dataset was observed. The model using the 30% testing dataset had an increased accuracy, TP rate and precision and a reduced FP rate compared to using the 30% testing dataset.

TABLE I: RESULTS OF THE EVALUATION OF THE PERFORMANCE OF THE MODEL

Training Data	Correct	Accuracy (%)	TP rate			FP rate			Precision		
			Spam	Ham	Fraud	Spam	Ham	Fraud	Spam	Ham	Fraud
70% Training Data	2407	93.55	0.914	0.833	0.997	0.052	0.036	0.001	0.894	0.86	0.999
80% Training Data	1594	92.94	0.912	0.813	0.995	0.06	0.038	0.001	0.88	0.853	0.999

V. Conclusion

The study presented a fraud detection model using a MLP which adopted the use of a back-propagation algorithm for model formulation using e-mail datasets. Two (2) datasets which consolidated served as the source of features extracted from the e-mails. By combining the features extracted from the CLAIR dataset with the features in the Spambase dataset, a single dataset required for the detection of fraud was presented.

It was concluded that by using a dataset containing a larger proportion of testing dataset (30%) produced a better result than using a lesser proportion (20%), during the model formulation and simulation using the back-propagation algorithm of the MLP classifier. It was also concluded that by adopting this model for integration into existing e-mail server can further improve the identification of fraudulent e-mails rather than their dependence on spam detection algorithms.

The study proposes for future work, the adoption of the use of other Nature-Inspired Algorithms such as: Swarm Intelligent Algorithms. These algorithms can be adopted for the purpose of identifying the most relevant features within e-mails before the formulation of model using the classifiers.

References

[1] Behdad M, Barone L, Bennamoun M, French T. Nature-inspired techniques in the context of fraud detection. *IEEE Transactions on Systems, Man and Cybernetics – Part C, Applications and Reviews. 2012; 42*(6): 1273-1290.

[2] Tive C. 419 Scam, Exploits of the Nigerian Con Man: Bloomington, iUniverse; 2006.

[3] Reich P. *Advance Fee Schemes in Country and Across Borders.* Proceeding of Crime in Australia International Connections Conference organized by Australian Institute of Criminology; Melbourne, Australia; 2004.

[4] Australian Competition and Consumer Commission (ACCC). *Upfront Payment and Advance Fee Frauds* [Internet]. 2017. [cited on 2017 November 30]. Available from: https//www.scamwatch.gov.au /types-of-scams/unexpected-money/up-front-payment-advanced-fee-frauds on.

[5] Nizamani S, Memon N, Glasdam M, Nguyen DD. Detection of fraudulent emails by employing advance feature abundance. *Egyptian Informatics Journal.* 2014; *15*: 169-174.

[6] Fraud MR. Detection using supervised machine learning algorithms. *International Journal of Advanced Research in Computer and Communication Engineering.* 2017; *6*(6): 6-10.

[7] Carcillo F, Dal Pozzolo A, Le Borgne Y.-A, Caelen O, Mazzer Y, Bontempi G. *SCARFF: a Scalable Framework for Streaming Credit Card Fraud Detection with Spark.* [Internet]. 2017. [cited on 2017 September 23]. Available from: https://doi.org/10.1016/j.inffus.2017.09.005.

[8] Zhang H, Li D. Naïve Bayes Text Classifier. *Proceedings of the IEEE International Conference of Computing*. 2007; 708-713.

[9] Abu-Nimeh S, Nappa D, Wang X, Nair S. A Comparison of Machine Learning Techniques for Phishing Detection. *Proceedings of Anti-Phishing Working Groups 2nd Annual e-Crime Researchers Summit.* 2007; 60-69.

[10] Amor NB, Benferhat S, Elouedi Z. Naïve Bayes vs decision trees in intrusion detection systems. *Proceedings of the 2004 ACM Symposium - Applied Computing*. 2004; 420-424.

[11] Sculley D, Cormack G. Filtering email spam in the presence of noisy user feedback. *Proceedings of the 5th Email Anti-Spam Conference*, 2008; 1-10.

[12] Chandrasekaran M, Narayanan K, Upadhyaya S. Phishing E-Mail detection based on structural properties. *Proceedings of the 1st Annual Symposium on Information Assurance, Intrusion Detection Prevention*, 2006; 2-8.

[13] Kim DS, Nguyen H.-N, Park JS. Genetic algorithm to improve SVM based network intrusion detection system. *Proceedings of the 19th Conference of Advanced Information and Network Applications 2*, 2005; 155-158.

[14] Degang Y, Guo C, Hui W, Xiaofeng L. Learning vector quantization neural network method for network intrusion detection. *University of Wuhan University Journal of Natural Sciences.* 2007; *12*(1), 147-150.

[15] Su M.-Y, Yeh S.-C, Chang, K.-C, Wei H.-F. Using incremental mining to generate fuzzy rules for real-time network intrusion detection systems. *Proceedings of the 22nd International Advanced Information and Network Application Conference*. 2008; 50-55.

[16] Rehak M, Pechoucek P, Celeda P, Krmicek V, Grill M, Bartos K. Multi-agent approach to network intrusion detection. *Proceedings of the 7th International Joint Autonomous Agents and Multi-agent Systems Conference.* 2008; 1695-1696.

[17] Srivastava A, Kundu A, Sural S, Majumdar A. Credit card fraud detection using hidden Markov model. *IEEE Transactions on Dependable Secure Computers.* 2008; 5(1): 37-48.

[18] Sanchez D, Vila M, Cerda L, Serrano J. Association rules applied to credit card fraud detection. *Expert Systems Application.* 2009; 36(2): 3630-3640.

[19] Panigrahi S, Kundu A, Sural S, Majumdar A. Credit card fraud detection, a fusion approach using Dempster–Shafer theory and Bayesian learning. *Information Fusion.* 2009; 10(4): 354-363.

[20] Bolton RJ, Hand DJ. Statistical fraud detection, A review. *Statistical Science Journal.* 2002;17(3): 235-249.

[21] Zavvar M, Razaei M, Garavand S. E-Mail spam detection using a combination of particle swarm optimization and artificial neural network and support vector machine. *International Journal of Modern Education and Computer Science.* 2016; 7: 68-74.

[22] Choudhary M, Dhaka A. Automatic e-mail classification using genetic algorithm. *International Journal of Computer Science and Information Technologies*. 2015; 6(6): 5097-5103.

[23] Bahnsen AC, Stojanovic A, Aouada D, Ottersten, B. Cost sensitive credit card fraud detection using bayes minimum risk. *Proceedings of the 12th International Conference on Machine Learning and Applications*. 2013; 333-338.

[24] Chaudhary K, Yadav J, Mallick B. A review of fraud detection techniques, credit card. *International Journal of Computer Applications.* 2012; 45(1): 39-44.