IT-Driven Approaches to Fraud Detection and Control in Financial Institutions

Alo .U.Rita and Henry.F. Nweke

Department of Computer Science,Ebonyi State University Abakaliki, Nigeria
*Email:- auzomarita@yahoo.com

Abstract

Financial Institution of this 21st Century has witnessed an increase in the incidence and sophistication of fraud, waste and abuse in organisational processes. To ensure sustainability, there is need to fight fraud effectively and continually, dynamically and intelligently monitoring organisational processes. This paper, IT-driven approach to fraud detection and control in financial institutions x-rays the fundamentals of fraud detection and monitoring dwelling on IT-driven approaches to fraud detection and control. The paper further gave typically applications examples using unsupervised learning in Neural Networks, case-based Reasoning, genetic algorithm and fuzzy logic. Its derivable benefits were x-rayed. It has been concluded that there is need for an IT-driven approach to fraud detection and control as a workable alternative to curb the increase and sophistication of fraudsters.

KEYWORD: Data Mining, Artificial Neural Network, Fuzzy Logic, Case Based reasoning

1.0 Introduction

Although there is a growing interest in creating fraud detection and monitoring engines because fraudulent activities in financial institutions has been on the increase in the last decade thus posing a source of worry to managements of organizations. The increase in the level of fraudulent activities engenered by innovations in modern technology, the emergence of global superhighways of communication has resulted in loss of billions of dollars world-wide each year through internet and financial fraud [8]. The global economic crime survey of 2009 suggests that close to 30% of companies worldwide were reported as victims to fraud in the past years [16]. Financial institutions of this 21st century having witnessed an increase in the incidence and sophistication of fraud, waste and abuse of organizational processes, need to fight fraud effectively, continually, dynamically and intelligently monitor organizational processes.

Fraud is a million dollar business that involves one or more persons who intentionally act secretly to deprive another of something of value for their own benefit. Though as old as humanity, it can take an unlimited variety of different forms. In recent years, the development of new technologies has also provided further ways in which criminals may commit fraud [5]. Instances of fraud can be similar in content and appearance but usually not identical [15]

Fraud detection is the ability to identify fraudulent transactional activities as they occur. Olufidipe refers to it as a deceit deliberately practiced in order to gain some advantage dishonestly [13]. Comer calibrates the scale of fraud in an organization as a reflection of the ability of management to reduce to a barest minimum such risks [6]. Financial fraud in general is referred to as a deliberate fraud committed by criminals that injures investors and creditors through false representation to gain an unjust financial advantage. These financial frauds range from corruption, advanced fee fraud, money laundering, counterfeiting, illegal charge transfers, computer credit card frauds, contract scam etc

In view of this, monitoring which is a surveillance system, is used by those responsible for event processes to access, process and find out if there are unexpected risks and proactively respond to such [9]. Monitoring is done during the implementation of the control intervention and hence involved with checking the timing at which certain events occur and the rate at which feedbacks are sent to control ensuring that standards are adhered amongst personnel and customers adopting appropriate process control methods pre-testing developed processes with appropriate test data and documenting results

Monitoring the process therefore requires an appraisal of organizational strengths, weaknesses, opportunities and threats in a turbulent and competitive environment. In essence, it accesses the progress and effectiveness of processes, considers the cost and efficiency, x-rays which changes are needed and helps to plan more effectively for the future [15]. Achieving fraud detection and control in this 21st century of increased globalization and competitive environment requires IT-driven approaches to fraud detection and control hence the need for this paper.

2.0 Approaches to Fraud Detection And Control

Uncovering financial fraud is a difficult task using normal audit procedures [13].

Firstly, there is a shortage of knowledge concerning the characteristics and manifestation forms of fraud. Secondly, given the periodic audit trailing, most auditors lack the needed experience and integrity to detect and expose fraudulent activities. Finally, the involvement of financial managers in perfecting fraudulent crimes has worsened matters. These limitations therefore suggest the need for additional analytical procedures for its detection. Increased IT awareness and usage in an IT driven economy requires IT-driven approaches to fraud detection and control. Some of these approaches are enumerated according to the computing technology, type, feature and examples as shown in Table 1.

Table 1: IT-driven Approaches to fraud detection

Technology	Type	Feature	Examples
Data miming	Decision Tree	Identify relationships that are not obvious	Find all multiple withdrawals above ₦500,000 for individual accounts
Text Mining	Thesaurus	Uses word and free form checking	Find all 'Account Information' in the institutions transaction processing system, check clearing system, email and customer management system
Statistics Suite	Models	Analytic Evaluation	The incident in question is ₦500,000 withdrawal , it's a new account and its 5 minutes to closing, so the statistical likelihood of fraud is high
Business Rules	Rules	"IF---THEN scenarios"	"IF the incident is a ₦500,000 withdrawal AND , it's a new account 'THEN' send the incident to the fraud investigating unit
Artificial Neural Network	Formulas	Identify patterns of behaviour using historic data	For every ₦500,000 withdrawal by a new account and its 5 minutes to closing, then automatically check for fraud
Fuzzy logic	Fuzzy Sets	Model partial truths	If ₦500,000 is a threshold for suspicion, ₦499,999 withdrawal should be examined
Case-based reasoning	Genetic Cases	Compare incidence to library of examples	For all withdrawals that are 5 minutes to closing with specific characteristics, GO TO the cased based reasoning library and check against concrete fraud samples to see if this incident is similar
Optimization Suites	Constraints	Apply multiple constraints to an evaluation to enable the best possible solution where no perfect solution exists	Only check for fraud (if 1=Y AND/OR, 2=Y BUT 3>five minutes), otherwise let the transaction go through
Genetic Algorithm	Evaluation Functions	Manage multiple variable types and complex outcomes	Using a library of known incidents, continue to find and update it with better examples of fraud(and replacing the proper ones in the library) resulting in a new, better re-combination of known fraud incidence
Goal-based Multi-agent	Correlation Tree	Find abnormal behaviour regardless of prior definition	Detection without human intervention that multiple withdrawals of ₦499,999 are fraud due to fraudster trying to stay under upper bound limit

3.0 Typical Applications

Neural Networks

Neural networks consist of many interconnected neurons or processing elements with familiar characteristics such as inputs, synaptic strengths, activation, outputs and bias. To develop an artificial neural network (ANN), one develops a mathematical model that best describes the biological systems functionality. Acomputer can then simulate the model fairly quickly and some degree of confidence may be gained with regard to its operation and functionality. To put it simply, an artificial neural network is a model that emulates a biological system. It is equally important to note that biological systems undergo learning especially by means of their neural networks. With the principle of emulation in mind, one can therefore talk of learning in artificial neural networks.

The particular learning technique of interest here is the unsupervised learning. It features first a trained artificial neural network (ANN). During the training session, the neural net receives as input patterns and organizes the patterns into categories. When at the application stage a stimulus is applied, the neural net provides an output response indicating the class to which the stimulus belongs.

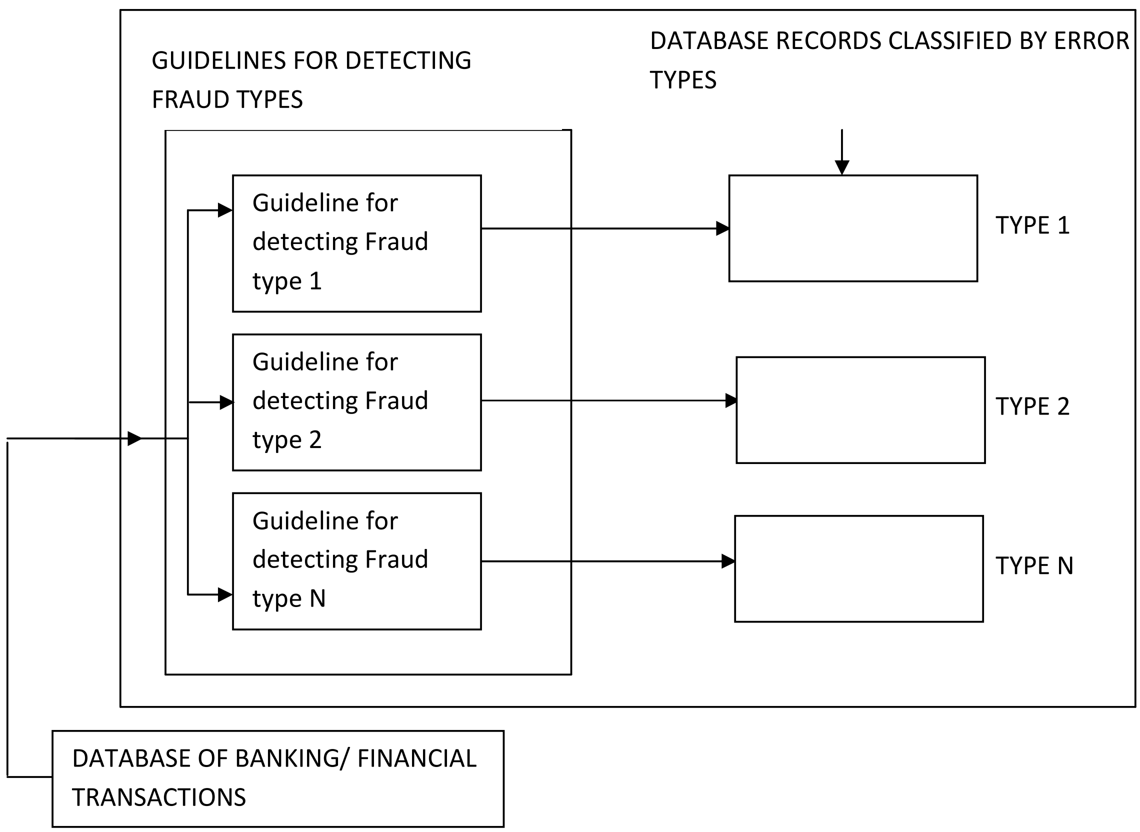

Fig 1 Artificial Neural Network Based Unsupervised Learning

Fig.1 shows the architecture of an artificial neural network based financial fraud detection system trained with the unsupervised learning technique. During the training session, a database of banking/financial transaction records is applied as input to the system one record at a time. Each record is classified into category (1, 2, … n) following clearly defined guidelines for determining financial fraud types. Typically, each fraud type has a software module that detects it. Each software module keeps in a file all database records that manifest the type of error it is trained to detect.

If the database of financial records contains known errors of different types including all the error categories of interest, then one could look at classified output to check that each record fell into the appropriate error type. Because all records would not be erroneous, category (n) or the last group could represent records without error. If it is found that a record did not fall into any of the category and is not error free, a new error class is created for it. Similarly, if more than one software module identifies a particular software record as belonging to the category of fraud they are designed to detect, the guideline for fraud detection used by such modules are adjusted until each record falls into only one category. This completes the learning/training session.

Cased Based Reasoning (CBR)

Cased based reasoning is a problem solving and learning paradigm that is fundamentally different from other major Artificial Intelligence approaches [1]. Instead of relying solely on general knowledge of a problem domain or making associations along generalized relationships between problem descriptors and conclusions, CBR is able to utilize the specific knowledge of previously experienced, concrete problem situations (cases). A new problem is solved by finding a similar past case and reusing it in the new problem situation. A second important characteristic of CBR is that it is also an approach to incrementally sustained learning since a new experience is retained each time a problem has been solved hence making it immediately available for future problems. Anderson [4] has shown that people use past cases as models when learning to solve problems particularly in early learning. Kolodner [10] equally supports this by indicating that the use of past cases is a predominant problem solving method amongst experts.

Another important feature of case based reasoning is its coupling to learning. The driving force behind this approach comes from machine learning. This is shown in the CBR cycle in Fig 2.

The cycle has four main processes. They are Retrieve, Reuse, Revise and Retain. A new problem is solved by retrieving one or more previously experienced case, reusing the case in one way or the other, revising the solution based on reusing a previous case and retaining the new experience by incorporating it into the existing knowledge base (case base).

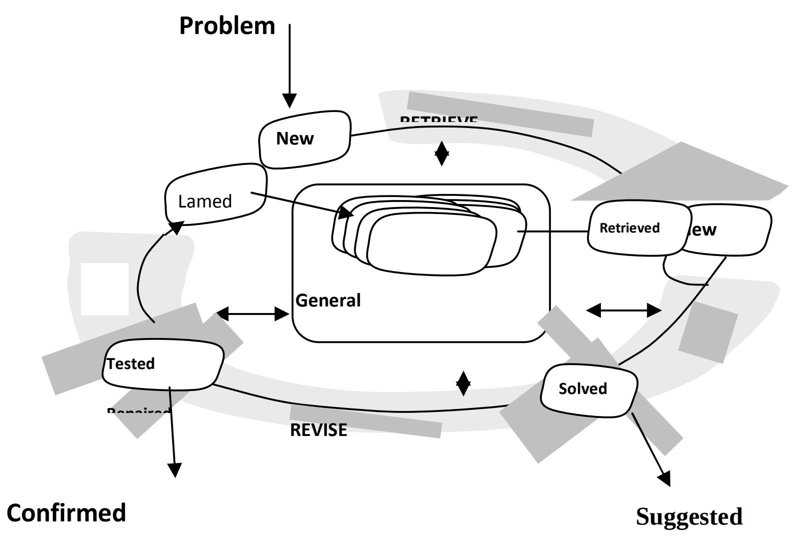

Fig. 2: Case Base Reasoning (CRB) cycle

Genetic Algorithm

Genetic Algorithm is a stochastic, heuristic and search optimisation technique used in computing to determine the exact or approximate solution to optimization problem. The algorithm is based on biological evolution and genetics which was developed by Holland in 1975[14]. They are variety of papers on the features of Genetic algorithm with different structures which can be modified to suit the problem at hand.

Genetic algorithm can effectively reduce the number of variables and search time to produce a predictive and effective model. The algorithm starts with a set of population spanning the entire search space, then the fitness function is evaluated for each population to determine the one with the best fitness function, that generate the a new set of population [11]. There are three different methods of generating populations with the best fitness function, which include reproduction, crossover and genetic mutation.

Reproduction passed on the next generation a population with the best fitness function. Crossover produces new offspring by bringing together the vectors of two population in the current generation while mutation produce a new population by altering the genetic make up of a population in the current generation. The algorithm continue until the best fitness function is reached or the maximum number of generation is gotten [10].

There are obvious advantages of using genetic algorithm in problem search and optimisation. Firstly, it studies the fitness functions and problems without differentiating them. This is useful when there are no comparable differences between the fitness functions and the problem at hand and also when the fitness function cannot be written in closed form. Secondly, genetic algorithm studies whole population at a point instead of individual unit of a population. This helps to find multiple optimal solution in the research space [10]. Other advantages of genetic algorithm were listed by Randy and Ellen Haupt [17], they include:

- optimize with continuous and discrete variables.
- simultaneously searches for wide sampling of the cost surface.
- deal with large number of variable
- does not require derivative information
- optimize variable with extremely complex cost surface
- provide a list of optimum variable not just a single solution
- may encode the variable so that the optimization is done with the encoded variable.
- works with numerically generated data, and experimented data or analytical functions
- genetic algorithm is mainly suited for parallel computation.

Fuzzy Logic

Fuzzy Logic starts with the concept of fuzzy set. Ali Alidoosti et al [2] defines Logic set as a collection of element in the universe of discourse where the boundary of the set contain ambiguous, Vague and unclear data. The fuzzy set is denoted by a member function which aligns each element in the universe of discourse with a value at interval [0,1]. Fuzzy set has no clear defined boundary [2], it contain with only a partial degree of membership. The concept provides a mathematical formulation that can characterise the uncertain parameter involved in risk analysis method. Fuzzy set can take on different degree of membership on continuous interval [0,1] where "0" means belonging to no membership and "1" means belonging to full membership.

Fuzzy IF-THEN Rule also called Fuzzy implication or conditional statement assume the form:

if x is A then y is B

A and B represent the linguistic values defined by fuzzy set on the universe of discourse x and y. x is A is know as Antecedent or premises and y is known as Consequence or condition. The rules are interpreted by evaluating the input and applying the necessary fuzzy operators. The premises is true, then the conclusion is true [17].

Fuzzy Reasoning (Approximate Reasoning) is an inference method used to derive conclusion from a set of Fuzzy IF-THEN rule and one or more conditions. The base rule is a modus ponen two value logic that can be used to infer the truth of a proposition B from the truth of A and the implication A→B.

For example, if A is identified with " the Mango is yellow" and B is identified with result is applied to the consequence, if the "the Mango is ripe" then if it is true that Mango is yellow, it is also true that " mango is ripe".

Premise 1 (fact)
x is A
premise 2 (rule)
if x is A and y is B
Consequence (conclusion)
y is B

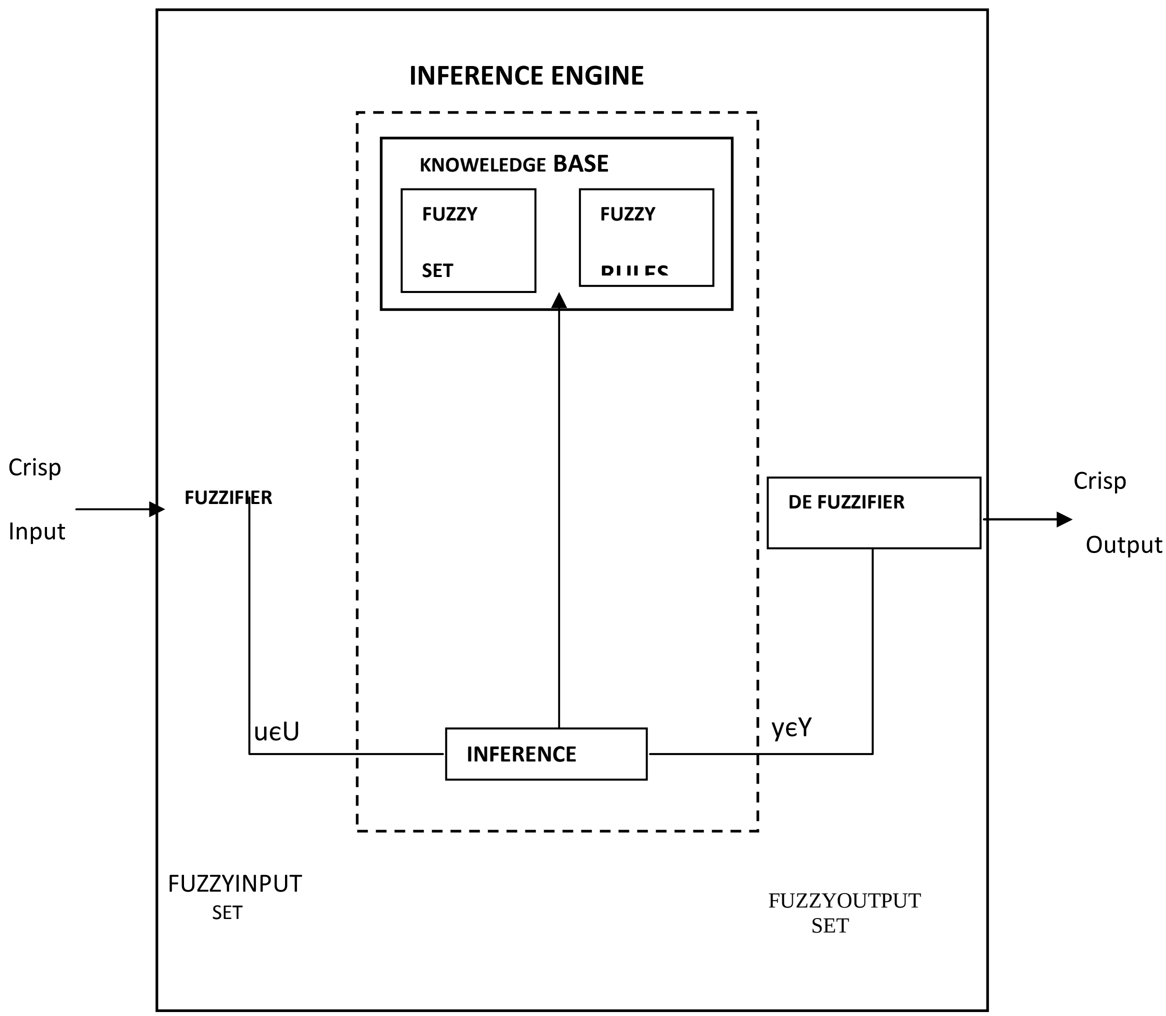

Fig. 3: **INFERENCE ENGINE**

There are different literatures that gave brief definition of Fuzzy inference system (FIS). Ali Alidoosti et al [2] describe FIS as the process whereby a given input set is mapped to an output set using fuzzy logic. According to Amrita Sarka et al [3], it is a nonlinear mapping of input vector into

Figure 3 shows the typical structure of Fuzzy Inference System that can be applied in fraud detection and control in a financial system. Chuen [7] explained the functions of the different components of fuzzy inference system. The components and their corresponding functions are:

Fuzzification Interface

- measure the value of input variables
- transform the values of crisp input variable into corresponding universe of discourse
- convert input data into linguistic values which can be viewed as fuzzy set label.

Knowledge base

- made up of knowledge about application domain, antecedent control goals, database and linguistic control rule base.
- the database provide definition for linguistic control rules and fuzzy data manipulation
- rule base shows the control goals and policy of the domain expert using linguistic control rule. it contain number of IF-THEN rule.

Decision-making/Implication part

- this is the kernel of Fuzzy Inference System;
- stimulate human decision-making based on fuzzy concept and inferring fuzzy control. action.

.

scalar output, while Nikam S R et al[12], shows it is computing framework based on the concept of fuzzy set theory, IF-THEN rule and fuzzy reasoning. Fuzzy Inference System use fuzzy logic to define an output as a function of measured input, tolerating linguistic and imprecise data

Defuzzification interface

- transform the fuzzy result into crisp output;
- yield a non fuzzing control action from an inferred fuzzy control actions.

3.0 Derivable Benefits

The benefits that are derivable from using IT-driven approaches to fraud detection include:

- Increased functionality, efficiency and productivity of organisation.
- Ensure judicious use of resources and maximization of profits.
- Helps organisations to maintain a competitive edge among contemporaries.
- Reduction in organisational waste and abuse.
- Ensure dynamic and intelligent monitoring of organisational processes.

Conclusion

The need for IT-driven approaches to fraud detection and control cannot be over-emphasised. The increase and sophistication of fraud will no doubt continue to grow with improved technology. Hence the need to continually, dynamically and intelligently monitor event processes through IT-driven approaches. This would reduce organisational waste, monitor event processes and present a well-deserved and maximised profit.

.

Reference

[1] Aamodt .A & Plaza. E (1994) "Case-Based Reasoning: Foundational Issues, Methodological Variation and System Approaches", AI Communications, IOS Press, Vol 7 No 1 Pg 39-59

[2] Ali Alidoostia, Ali Jamshidib, Siamak Haji Yakhchalic, Mohammad Hossein Basirid, Ramin Azizie & Abdolreza Yazdani-Chamzinif (2012), "Fuzzy Logic for Pipeline Risk Assessmen"t, Journal of Science Letters, Vol 2 Issue 5 Pp 1707- 1716

[3] Amrita Sarkar, G.Sahoo & U.C.Sahoo (2012), "Application of Fuzzy Logic in Transport Planning", International Journal on Soft Computing (IJSC), Volume 3 Issue 2 Pp 1-29

[4] Anderson .J. R (1983) "The Architecture of Cognition", Harvard University Press, Cambridge

[5] Bolton .R. & Hand .D (2002) “Statistical fraud Detection: A review”, A Statistical Science 17(3) Pg 235-255

[6] Comer C.O (2004)“Fraud management, Detection and Control” unpublished project

[7] Chuen Chien Lee (1990), "Fuzzy Logic in Control System: Fuzzy Logic Controller- Part "I, IEEE Transaction on Systems, Man & Cybernetics Vol. 20 No. 2 Pp 21-36.

[8] Doughety Gary (2000) “Real-Time fraud Detection”, FleetBoston Financial Cooperation

[9] Kahn M.G, Barley T.C, Sterb S.A, Eraser V.J, Dunagan N.C (2001) "Statistical process ontrol methods for expert systems performance monitoring", Department of Miam, st. Louis Mo USA, session notes Pp 2-7. Online at http://www.who.n/medicines/organisations/par/programme.html.

[10] Kolodner . J (1988) "Reconstructive memory, a computer model", Cognitive Science Vol 7 Pg 281-328

[11] Nabel T Alshabatat (2012), "Optimizing the Natural Frequencies of Beam via Notch Stamping", Research Journal of Applied Science, Engineering and Technology, Volume 4, Issue 14 Pp 2030-2035.

[12] Nikam S R, Nikumbh P J & Kulkarni (2012), "Fuzzy Logic and Neuro Fuzzy Modelling", Conference Proceeding of International Journal of Computer Application, Volume NCRTC Issue 4 Pp 22-31.

[13] Olufidipe E.O (1994) “Fraud in Nigerian Economy and its implications” The Nigerian Banker.

14] Ozlem Gurunlu Alma & Elif Bulut (2012), "Genetic Algorithm Based Variable Selection for Partial Least Square Regression Using ICOMP Criterion", Asia Journal of Mathematics and Statistics, Vol 5 Issue 3 pp 82-92.

[15] Palshikar G.K (2002) “ The Hidden Truth: Fraud and their Control: A Critical pplication for business intelligence”, Intelligent Enterprise Vol 5 No 9 Pg 46-51

[16] Pricewaterhousecoopers LLP (2009) “2009 Global Economic Crime Survey” http:/www.pwc.com/crimesurvey

[17] Randy L Haupt & Sue Ellen Haupt (2004), "Practical Genetic Algorithm", 2nd dition, John Willy & Sons Inc., USA.