

Improved Ghost Worker Fraud Detection System Using Artificial Neural Network

Promise Elechi

Department of Electrical Engineering
Rivers State University
Port Harcourt, Nigeria
elechi.promise@ust.edu.ng

Sunny Orike and Soye Friday Bille

Department of Electrical Engineering
Rivers State University
Port Harcourt, Nigeria
orike.sunny@ust.edu.ng

Abstract – Payroll fraud has been identified as one of the major problems in Nigeria’s civil service sector and it is considered as a huge drain to the economy. Over 40% of the government total recurrent expenditure is channeled to personnel cost, and ghost workers milks a large portion of this recurrent expenses. The primary aim of this work is to design and implement an improved fraud detection system for ghost workers using artificial neural network. The methodology adopted was a forensic analysis using an artificial neural network model and measuring the rules whose weight of computation to fire an artificial neural network model and further analysis and design using software development life cycle for payroll design architecture was used. Microsoft Visual Studio 2010 tools for web application which contains C# programming language, HTML, CSS, SQL server and JavaScript was used to implement the application development. The results showed web pages with a dataset of transactions to check for fraud among civil service employees and classification of the staff payroll transaction data by the trained model which detects whether a staff is a ghost worker or not. However, this forensic model application will help all department of the civil service commission to reduce the rate of fraud that is rampant in almost all the ministries of the federal republic of Nigeria and to sustain national development.

Keywords-*fraud; SQL; ANN; Detection; Software; Ghost worker.*

I. INTRODUCTION

Fraud has been identified as one of the problems of Nigeria. This is not peculiar to one sector of the economy rather it is a major problem in the public as well as the private sectors. It has not only snapped public trust in government, but has cost the nation billions, ranging from deteriorated infrastructure to unrealized public projects and corrupt management of public companies leading to delayed development in the civil service.

This menace is considered a national challenge and a drain on the national economy. The former Nigeria’s Minister of Finance, Kemi Adeosun, during the budget padding revealed the detection of over 23,000 ghost workers in the federal bureaucracy amongst the Ministries, Departments and Agencies (MDAs). It is important to note that the number revealed was less than the 62,893 reported by Dr. Ngozi Okonjo-Iweala, erstwhile Finance Minister during her time. The ghost worker effect is high, and its related impacts are

alarming. Recent records and discoveries of ghost workers in some of the States in Nigeria include, 5,000 for Plateau; 8,000 for Kano; 850 for Anambra; 9,300 for Kebbi; 71 for Osun; and 7,000 for Niger. By estimation, the remaining States have between 100,000 to 151,000 ghost workers. These figures indicate that there may be at least one ghost worker to three legitimate workers in country’s public service. It had been reported that 71,135 ghost pensioners exist on the federal government payroll [1].

The number of employees in an organization has increased such that business owners are unable to identify all their employees by name, in person or even by what they do in the organization. Ghost worker is a fictitious worker added to the payroll and the wages or salaries paid to that name are collected by the perpetrator. The scheme is usually done in large organizations, particularly when employees are spread over several geographic locations and where the payroll function is centrally controlled. This scheme is possible when the officer authorizing wage payment recognizes the ghost worker on the payroll list or that officer could be the perpetrator. Hiding the fraud is not necessary as the payments are made and recorded as legitimate operational expenses [2] and [10].

Ghost worker usually exists as an identity in payroll records, although the ghost employee may be a real person who does not work for the organization. This is a scam that is usually successful if the perpetrator of this unethical act has access to the organization's systems, hence, it is typically an inside job. The scheme works if:

1. The ghost’s identity is successfully added to the organization's payroll records.
2. The system is set up to make payments to the ghost possible, for wages, or other forms of payments, like expense reimbursements and extra work hours; and
3. All payments made to the ghost is concealed, especially from all existing control mechanisms.

Ghost workers represent dysfunctionality in governance. This is to be considered an abnormality in a functional self-accounting and self-auditing bureaucratic system, which may need urgent remedial measures. A ghost worker is a virtual worker existing in the payroll records, not in the employment of the

public service institution. They exist only in name and thus deny other citizens by collecting wages and entitlements that are not due for them. Some are engaged in one job but collect wages from two or more pay points.

This work is aimed at designing an improved fraud detection system for ghost workers using artificial neural network.

A. Background Study of Fraud

Fraud means giving a false or misleading account to manipulate or deceive an individual or organization. It creates a misjudgment or maintains an existing one as a way of inducing someone to make a contract. It has to do with the intentional enrichment of oneself by reducing the value of an asset in secret. Fraud by officers may be responsible for companies going through serious financial problems and even being bankrupt.

[3] reported that fraud as a problem is not a possibility but rather a probability. His explanation on the concept is that its prevention would be successful if there are collective efforts or decisions by a group of people as against individual effort, although this case is different where the group of people have the same interest. Where this is the case, the prevention of fraud may not be successful. Conversely, the group is influenced by the dominant decision maker capable of deciding everything. Russel, 1978 cited in [4] stated that the fraud is a generic term used in various ways.

B. Who is Committing a Fraud?

Organizations will be successful in their anti-fraud policies when focused on the right areas. It is immensely needful for organizations to identify those likely to commit fraud and the circumstances under which individuals may do so. For instance, opportunists could be successful in taking the chance to enrich themselves from programs at the early stage(s) of development, hence having limited control. Also, senior managers may resort to unethical means to achieve their goals while under intense pressure to meet high targets.

Fraud assumes different forms as well as degrees, hence, making courts to contend with only few rules for its defeat. It is safer for a generally acceptable definition so that individuals would find ways of committing frauds which might not fall within the essential ingredients of such acceptable definition as a way of evading justice. [5] identified fraud as a general term that encompasses numerous means which human ingenuity can devise, resorted to by one individual to take advantage over another by false representation. According to [6], fraud is an act of deception practiced intentionally by gaining unfair and unlawful advantage to the detriment of another. Accounting fraud refers to an act of deliberately falsifying records of accounting, such as sales and cost records to boost the net income or sales figures. This can subject the company and the executives involved to civil lawsuits [7]. Accounting frauds are usually committed by company officials to reverse

loss or to meet earning expectations from the shareholders and the public.

II. MATERIALS AND METHOD

A. Materials

For effective design and operation of the proposed system, the following hardware specifications were used:

- Core I-series laptop
- The minimum Random-Access Memory (RAM) should be 2024MB
- The system should have a hard disk of at least 40GB and CD-ROM drive
- The mouse, keyboard and printer are also required.

And for software requirement includes:

- Operating system - Certified distribution of Windows.
- Front end- Microsoft Visual Studio 2010.
- Back end- SQL Express Management Studio

B. Method

Data Mining refers to the secondary data analysis of observational data sets to find unsuspected relationships and to encapsulate the data in better ways that are both understandable and useful to the owner. The data mining relationship exercise is referred to as models or patterns. Data Mining typically deals with data that have already been collected for some purpose other than the data mining analysis. Therefore, the objectives of the Data Mining exercise play no role in the data collection strategy.

However, for the detection of payroll ghost worker fraud, the Data Mining approach for the detection of fraud in the Rivers Civil Service record, a pattern discovery association rule model will be formulated and combine with an artificial neuron perceptron model.

C. Techniques of Pattern Rule and Artificial Neural Network Model

A data set is a set of measurements taken from some process or environment. Using a collection of m objects as a case study which will imply m measurements. In this case, one may think of the collection of the measurements on k objects as a form of $k \times m$ data matrix. The n rows represent the n objects on which measurements were taken for example employees, payrolls, or individual objects observed in the HR record, such as Employee_Id and Email_Id. Such rows may be referred to as individuals, entities, cases, objects, or records depending on the context.

Artificial Neural Network (ANN) is a vastly simplified model of the form of a biological network. ANN is usually referred to as neural network (NN). [8] and [9], defined Artificial Neural Network as a computational model which is inspired by the

structure and functional aspect of biological networks. ANN shows better modeling technique for payroll record of workers data sets with non-linear conjunction, and thus for both data fitting and prediction strengths.

The fundamental processing element of ANN is an artificial neuron (or commonly a neuron). A biological neuron accepts inputs from other sources, integrate them, carry out generally a nonlinear process on the result, and then outputs the last result.

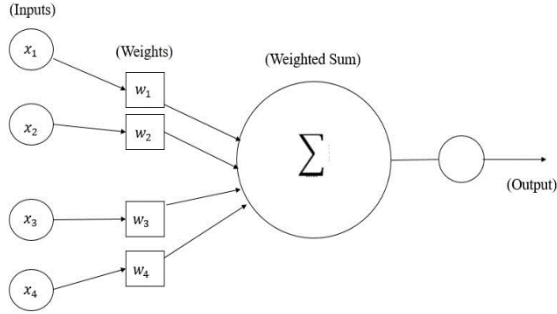


Fig. 1: Perceptron Model

D. Formalizing Mapping

$P \rightarrow Q$ with support count set to the threshold value of one and it determines how often a rule is applicable to a given data set while the confidence determines how the given BVN code for employees that appeared in the payroll transaction is authentic.

From the Data set as shown in Table 1, the number of transaction and attribute of employees are represented as an item set for each transaction record.

Table 1: Data Set for Payroll information (Sample)

T/No	BVN_Code	Emp_No	Email_Id	Last Name	First Name	Income	Month	Year
1	4500002101	PSP002	Mark.coy@psp.com	Mark	Coy	56000	March	2015
2	3509021102	PSP120	John.cole@psp.com	John	Cole	75000	March	2015
3	4500002101	PSP109	Mark.paul@psp.com	Mark	Paul	60000	March	2015
4	3509021102	PSP008	Kim.sawyer@psp.com	Kim	Sawyer	60000	March	2015
..
N								

The Association of the employee properties are formally defined below:

$P \rightarrow Q$

$\{\text{Month, Year}\} \rightarrow \{\text{BVN_Code}\}$

$\{\text{March 2015}\} \rightarrow \{4500002101\}$

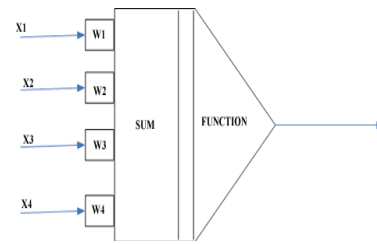
$$\text{Support} = S(P \rightarrow Q) = \sigma(P \cup Q)/N, \quad (1)$$

Where: σ is Support count, P is Month, Year, Q is BVN_Code,

Therefore, the association of the BVN_code with the Year and Month of Payroll transaction in the record is given as a threshold function unit value that is the minimum value of one to determine whether there is a ghost worker or not.

$$\{ \text{Fraud}(TP_i) = \sigma(P \cup Q) \} = \begin{cases} 0 & TP_i > 1 \text{ then Fraud Exist} \\ 1 & TP_i = 1 \text{ then No Fraud Exist} \end{cases} \quad (2)$$

Q represents the BVN_Code for tracking the existence of employee data properties mapped to their respective monthly allowances. (Duplicate Verifications)



$$\text{Output} = F [\sum_{i=1}^4 X_i \cdot W_i] \text{ Fraud/Not Fraud}$$

Fig. 2: Function of Artificial Neuron

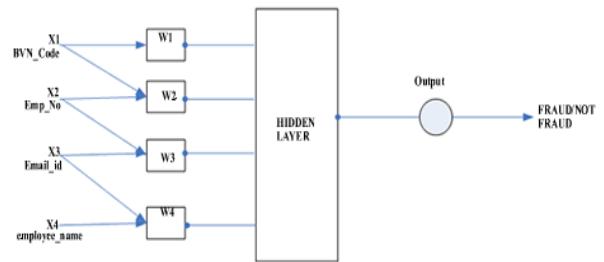


Fig. 3 Fraud Detector Model with ANN

Nodes (neurons) in input layer only acts as buffers for the delivery of the input data $X_i = 1, 2, \dots, n$ to nodes in the hidden layer. Each processing node in the hidden layer sums up its input data X_i after weighting them with the strengths of the connections W_i from the input layer and calculates its output as a function of the sum (shown in Figures 2 and 3).

$$\text{Output of Neuron} = F \left[\sum_{i=1}^4 X_i \cdot W_i \right] \quad (3)$$

Where input parameter of the data set are the properties of each worker in the payroll record and they are tracked in the hidden layer of the node connections with a weight variation and the sum of the input combination after weighting is passed to the output function of the perceptron for activation of the neuron. The threshold value required for firing the neuron is compared with the minimum threshold value for the pattern rule model obtained for the association model rule. The neuron fires if there is fraud from ghost worker detection otherwise the transaction is not a Fraud.

E. Forensic Analysis of Data model for Detecting Ghost Worker

Recall from the pattern rule model, an implication of the payroll transaction item sets for the workers on the payroll record given and mapped to the bank code (BVN) for tracking.

$$\{\text{Month, Year}\} \rightarrow \{\text{BVN_Code}\}$$

Item Sets are the worker's attributed properties in the dataset like $\{\text{BVN_code, Emp_No, Email_Id, LastName, FirstName, Income, Age, Month, Year, Date}\}$

Given that the threshold value of $\sigma(\text{PUQ})$ is upheld and the No Fraud associated with the BVN_code and duplication then the equation combined with the neural network for pattern similarities with the computation of weight given below.

From equation 3,

$$\text{Output of Neuron} = F \left[\sum_{i=1}^4 X_i \cdot W_i \right] \quad (4)$$

$$\text{Output of Neuron} = W_1X_1 + W_2X_2 + W_3X_3 + W_4X_4 \quad (5)$$

Weight of Connection from the strength order are the Human Resource Master Record Verification, W1, the Payroll Information Master File W2, Payroll Transaction Record W3, Payroll Validation Record W4.

Neuron Output = $f(\text{Summation})$ if Output ≤ 1 then No Fraud, that can match with the pattern rule minimum threshold value. Otherwise Fraud is detected due to ghost workers and the same matched with the pattern rule.

If (Fraud (TPi) = 1 and $f(\text{Summation of node weight}) \leq 1$) Then No Ghost Worker Found with the binary classification set to logic 1,

Otherwise Ghost Worker Found, binary classification set to logic 0

This classification is the grouping of workers that are authentic with their records traced to the master files and the false record of employee whose records are duplicated and cannot be traced to the human resource master file.

The sigmoid function is used to calculate the output of each network layer and is defined as follows:

$$g(\text{Summation}) = g(x) = \frac{1}{1 - e^{-x}} \quad (6)$$

The squared error function is defined as follows:

$$E = \frac{1}{2} (y - g(x))^2 \quad (7)$$

Where $g(x)$ is the network's output prediction and y is the instance class label. The weights of the neural network determined from the derivative of the squared error function. The derivative of the error function with respect to a weight is given by:

$$\frac{dE}{dw_i} = (y - f(x))f'(x)a^i \quad (8)$$

Where w_i are the weights for the i th input variable, x is the weighted sum of the inputs, and are the inputs to the neural network. This computation is repeated for each training instance, and the changes associated with a weight, w_i are added up, multiplied by the learning rate (small constant), and subtracted from the w_i 's current value. This is repeated until the changes in the weights become very small.

F. Software Development Life Cycle for the Payroll and Ghost Worker Detection

Software Development Life Cycle (SDLC) is a framework that describes the activities performed at each stage of a software development project. It involves system analysis, design, implementation, maintenance and disposal of the system.

1. System Analysis

System analysis entails detailed study of the existing system, leading to specifications of a new computer-based payroll system. In this stage, data are collected on the available files, decision points and transactions handled by the existing system. Interviews, on-site observations and questionnaire are the tools used for system analysis of present system.

2. Existing System

Payroll is the financial record of employee's basic salary, allowances, deductions and net pay. Ministries of the Rivers State civil service is a body saddled with the responsibility of carrying out the policies of the Rivers State government in relation to infrastructural development and social service delivery. The Ministries have a pool of employees, contract staffs, administrative and casual staffs. Where there is existing payroll system, every calculation is done by various financial units of each ministries, departments and agencies of the Rivers State civil service and the operation not fully computer-based. Major operations in the existing system are done manually. The process of paying an employee is very tedious, time consuming and frequent padding of training budgets and absentee workers or ghost employees to the payroll records by the units exposed to alterations.

3. Proposed System

The proposed system is a web-based solution for a ghost workers payroll fraud detection using Artificial Neural Network. The base of the proposed system is a database, which stores all the information pertinent to employees, and income payment. The system for the validation of personnel records is the data set for and existing record to be tracked for frauds used to test the model for performance and usage efficiency.

4. Feasibility Analysis

Feasibility study is conducted to see the advantages of automated system over manual system operation. It includes the merits and demerits of the existing

system and proposed solution. The objective of feasibility study is to divide the task and time in proper way, implement and then terminate it effectively. The schedule feasibility has been decided at the time of synopsis submission by defining time for each step and the schedule work-flow diagram shows the progress, the schedule for each process in a project and the total project time.

5. Use Case Scenario for the Detecting Ghost Worker

This use case scenario describes how the ghost workers can be tracked on the intelligent model illustrated in Figure 5 for a payroll management system with a data set specified.

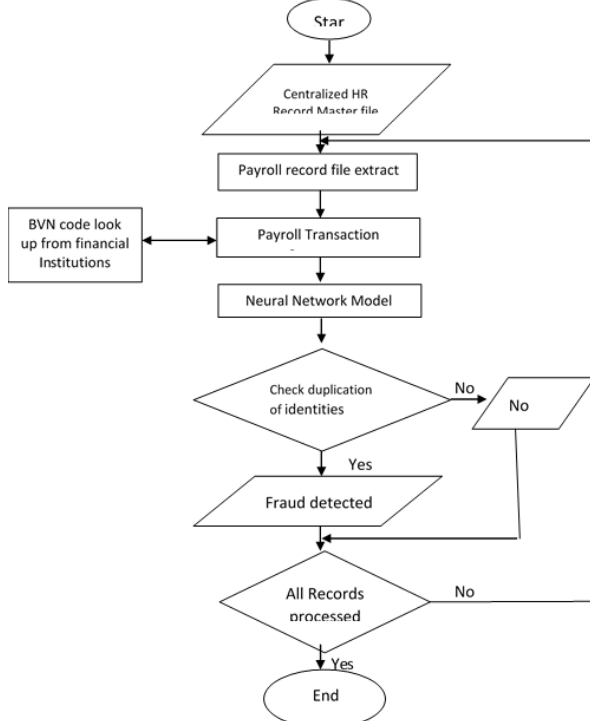


Fig. 4: Scenario Diagram for Fraud Architecture

G. Conceptual Design Model for the Application Model

Conceptual model design document is a power communication tool that consist of concepts and ideas which helps people to understand the application model.

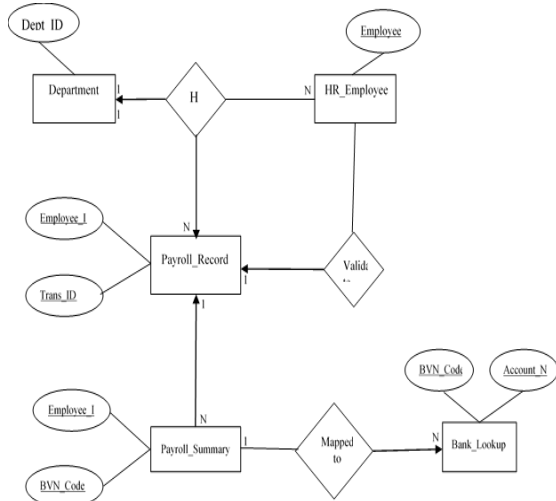


Fig. 5: ER Diagram

H. Data Flow Diagram for the Payroll System

The operation of the detecting model and processes can be analyzed using the data flow diagram as shown in Figure 6.

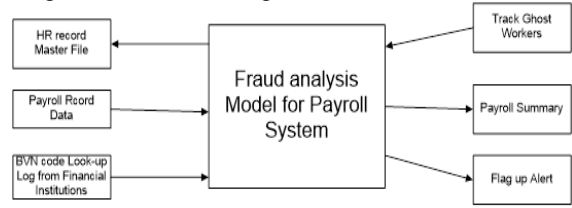


Fig. 6: DFD Level 0

I. Logical Design

A comprehensive as well as unambiguous definition of the solution’s logical elements from the user and functional point-of-view is presented by the logical design. This is a design in the absence of encumbrances of architecture, infrastructure and technology. A logical design can identify and define all the objects as well as their behaviors, relationship and attributes within the scope of the solution as shown in Figure 7.

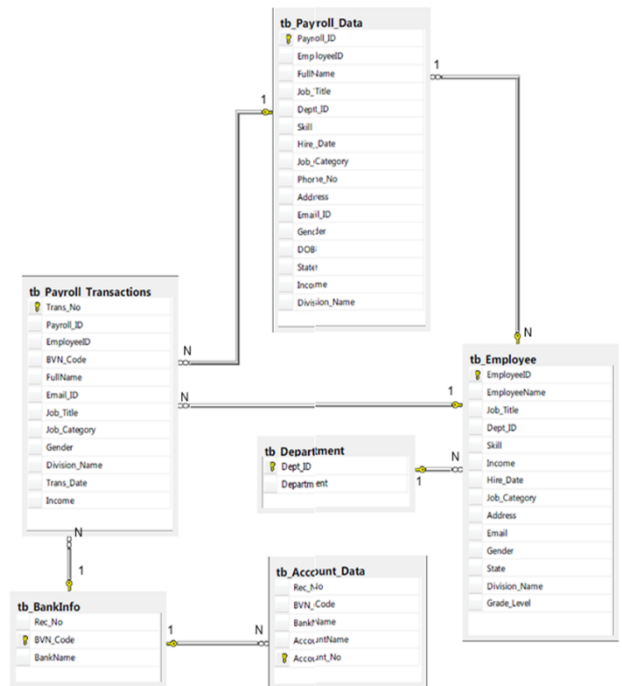


Fig. 7: Logical Design

J. Model Evaluation and Performance

The performance of the neural network model is obtained using the measures of detection rate/precision and accuracy from the classified dataset. Parameters used for the model performance evaluations include the following:

True Positives (TP): True positives depict transactions as fraud are correctly classified as fraud in the classified data.

True Negatives (TN): True negatives show classify legitimate transaction as normal predictions, that is ‘No Fraud’ exist from the classified data.

False Positives (FP): This means a false alarm by the model. No fraud exists due to retirement.

False Negatives (FN): This means fraudulent transaction classified as normal behavior. Fraud exists due to retirement. The model performance can be evaluated using the following equations.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FN+FP} \quad (9)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (10)$$

$$\text{Recall, } r = \frac{TP}{TP+FN} \quad (11)$$

III. RESULTS AND DISCUSSION

A. Results

Figure 8 shows the dialog file upload for loading the staff payroll data into the system for tracking and identities verification whose duplicate is in the central storage data center and each record identified and approved.

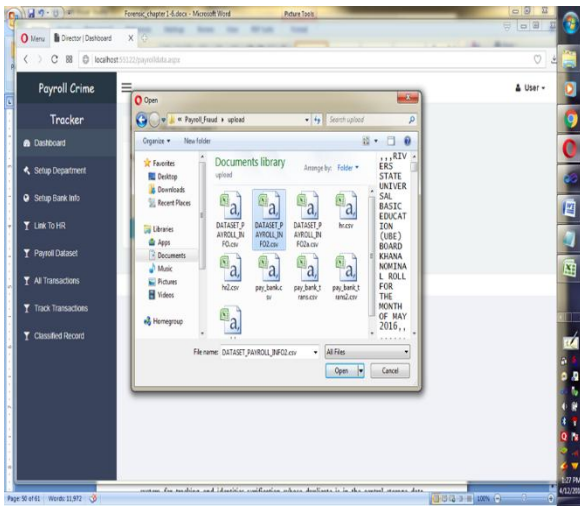


Fig. 8: Upload of Payroll Dataset File to Track

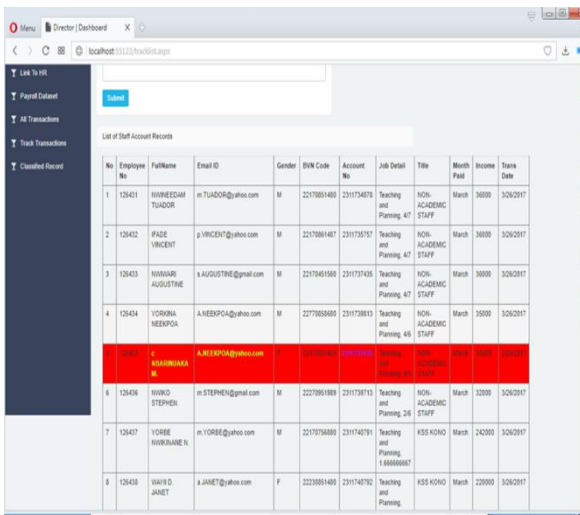


Fig. 9: Detected Transaction

Figure 9 shows the red mark for emergency point for illegal transaction which is not true, and verification of transaction was not confirmed with the center data for payroll record and record contradicting

each other with the neural model detecting the staff number not in the record as a ghost worker.

B. Classification of Outcomes

Fig. 10: Classification of Payroll Transaction

The Figure 10 had classified the original transaction and fake transactions with the yellow marked as ghost worker fraud with value 1 and the original one as Not fraud with value 0. The binary classifier for detection is a function that fires from the value obtained from the neural network model combined with the association pattern rule. From the 71 transactions, 7 records transactions were classified as illegal records, which signify ghost worker fraud.

C. Model Performance Evaluation Results

In classification problems, the main source of performance evaluations is a confusion matrix.

Table 2: Confusion Matrix of the Classified Outcomes

	n=71		
	Predicted: Not Fraud	Predicted: Fraud	
Actual: Not Fraud	TN=61	FP=0	61
Actual: Fraud	FN=2	TP=8	10
	63	8	

Table 2 shows that the neural network model developed in this project is 97 % accurate in model performance and a precision of 100%.

D. Testing and Debugging

Every unit developed in the implementation phase are integrated into a system after testing each unit. The application programming code of the whole system is tested by analysts and programmers as a way of finding correct errors. The new system is considered part of the daily activity of the organization during installation. Application software is installed on existing and/or new hardware; there is introduction of users to be part of the new system and trained.

IV. CONCLUSION AND RECOMMENDATION

A. Conclusion

This work has developed an improved fraud detection system for ghost workers using artificial neural network. The solution was implemented using C#, HTML, CSS, SQL server and JavaScript with a data set of transactions to track and detect fraud among civil service employees. Payroll application developed in this project is accessible with an internet browser. It calculates, maintains and records the payroll information of employees. This solution provides a powerful tool that can be used to detect, track and eliminate payroll fraud in Rivers State civil service

B. Recommendation

From the Classification and result of our computation, it was revealed that forensic model analysis has significant relationship with fraudulent practices in Nigerian public institutions. This implies that as forensic analysis practices in Public institutions increase, fraudulent practices will reduce.

Again, the implementation strategy and monitoring systems of the Civil Service should be taken seriously to ensure that the Rivers State Civil Service would be service-and-result-oriented and geared towards the achievement of the nation's socio-economic development.

There should be more research and proposals focusing on real time fraud detection of payroll fraud via ghost workers. This will greatly reduce the ugly trend in the Rivers State Civil Service.

Finally, for the Rivers State government to succeed, the Civil Service should be organized in a way to champion the national change process initiating, guiding and managing the change.

REFERENCES

- [1] O.H. Obaze, Caging Nigeria's ghost workers, <http://sunnewsonline.com/caging-nigerias-ghost-workers/>, 7th March 2016. Visited 13/04/2018.
- [2] J. Finlay, ACFE Issues 2010 Report to the Nations on Occupational Fraud and Abuse, https://www.acl.com/pdfs/ACL_fraud_ebook.pdf. Visited 12/04/2018.
- [3] A. David, Low Level Fraud is a High-Level Issue, Supply Management, 2005. Retrieved from: <http://www.highbeam.com/doc/1P3-821992191.html>. Visited 25/03/2018.
- [4] A. Bello, Fraud Prevention and Control in Nigerian Public Service: The need for a Dimensional Approach. *Journal of Business Administration*, vol. 1, no. 2, pp. 118-133, 2001.
- [5] B. Okafor, Strategic Approach to Reduction of Employee, Theft Fraud and Embezzlement. *Nigerian Accountant*, vol. 37, no. 4, pp. 3-5, 2004.
- [6] J.C. Anyanwu, *Monetary Economics: Theory, Policy and Institution*, Onitsha: Hybrid, 1993.
- [7] L. Arokiasamy, and S. Cristal-Lee, Forensic Accounting: Public Acceptance towards Occurrence of Fraud Detection. *International Journal of Business and Management*. Vol. 8, no. 12, pp. 145-160, 2009.
- [8] K.U. Rani, Analysis of Heart Diseases Dataset using Neural Network Approach. arXiv preprint arXiv:1110.2626, 2011. Visited: 18/12/2017.
- [9] J.V. Tu, Advantages and Disadvantages of using Artificial Neural Networks versus Logistic Regression for Predicting Medical Outcomes, *Journal of clinical epidemiology*, vol. 49, no. 11, pp. 1225-1231, 1996.
- [10] P. Elechi and N.R. Saturday, An Interactive Virtual Classroom System for University Education, *Elect and Mech Eng:Open Access* 1:1, 1-7.

