RESEARCH ARTICLE OPEN ACCESS

Predictions of On-Line Banks' Fraudulent: A Study on Discriminant Analysis.

Gabriel Joseph Mukungu[1], Gwangyong Gim[2]
[1]Doctoral Student, Department of Business Administration, Soongsil University.
[2]Professor, Department of Business Administration, Soongsil University.

Abstract
The information and communication technology (ICT) is growing fast in the community along with some risks. The on-line banks' fraudulent are increasing dramatically resulting in the loss of money and retardate the growth on economic development. Some law enforcers like Tanzanian police yet use the traditional prevention technologies to investigate these on-line banks' fraudulent cases mainly known as cyber crimes.
For the purpose of this study, 150 data were collected as cyber criminals and non criminals. 100 data were cyber criminals out of the total sample size. With the aid of the discriminant analysis tools predicted cyber criminals against non criminal that 88% were overall, sensitivity was 73.9% and specific was 100%. However, the results from cyber criminals relatively with their ages implicated that the overall was 80.7%, sensitivity was 91.7% and specific was 75.5%.
As technology advances, the law enforcers like police need to use the modern tools to predict these cases.
Key words: On-line banks' fraudulent, cyber criminal, discriminant analysis.

I. Introduction

1.1 The growing of ICT applications in Tanzania.

The advancements in information and communication technology (ICT), in particular, the growing use of mobile money transactions and electronic banking service in Tanzania have a profound effect on the unbanked customers at urban and rural areas. In this fact, most of human activities in Tanzania, specifically the mobile money transaction depends upon the power of information and communications technology to stimulate the growth on economic development in real- time businesses and reduced costs to meet with customer' need's satisfactions.

For example, currently in Tanzania, there are four companies offering mobile money transaction services in the market, known as Vodacom M-Pesa, Tigo Pesa, Airtel Money and Ezy Pesa (Zantel Z-Pesa). Some of the services available to the users of mobile money transactions are domestic money transfers, mobile payments (airtime top-ups, merchant payments, utility bill payments, and salary transfers), and mobile banking services like balance inquiries, withdrawals, deposits and credit services, figure [1-1].

[Figure 1-1] Send money by phone

Source: (M-pesa, 2014)

In addition, the National Microfinance Bank (NMB) offers a mobile money application, "Pesa Fasta" which allows its customers to use their mobile phones to send and receive money to any person in Tanzania, who have a bank account (NMB mobile,

2014).

On the other side, the on-line banks' fraudulent known as cyber criminals have used their knowledge, in one way or another, to fraud these transaction businesses by either accessing the identity of the rightful mobiles' owner, or by tricking the victim into sharing information (Fraud Alerts, 2012). In other words, this fraudulent offenders conduct their activities with highly multi-skilled knowledge to achieve their aimed goals like obtaining money or goods by fraudulent.

1.2 Cyber crimes

These offences include a broad range of different offences such as offences against the confidentiality, integrity and availability of computer data and systems, offences related to computer, offences related to copyrights, etc. Cyber crimes are criminal activities like other crimes but they are directly related to the illegal access of computers data bases manipulation or theft of stored on-line data appliances, or sabotage of equipment and data. In other words, these are crimes that include crimes that are wholly mediated by communication technology (Wall, 2007). The cyber crime is the crime whereby ICT plays a significant role to accomplish the offence omission, for example, Leukfeldt, Kentgens, Frans, Toutenhoofd, Stol, and Stamhuis, (2012) illustrate the penalization of 28 crimes, ranging from hacking digital systems and installing spy ware to fraud using Internet banking and cyber stalking to endanger the government property.

It is from this logic that creates a new mind for the law enforcers like police to find and explore the modern application of investigative techniques like discriminant analysis to overcome the problems of cyber crimes like on-line banks' fraudulent.

1.3 Motivation of the Study

The discriminant analysis technique is aimed for cyber crime- solving as a predictable technique tool to overcome the incident of banks' fraudulent and future forecasting another serious crime at low- cost and in real time (Sullivan & Perry, 2004). This statistical analysis provides adequate satisfactory security and peace to the community to assist the economic growth and social development to the community of Tanzania.

Main Objective of this study was to explore the modern predictable technique to overcome the on-line banks' fraudulent (cyber crime- solving).

From the main objective, specific objectives are:

- To identify variables to be involved in the discriminant analysis to predict cyber criminals.
- To build a model for predicting the cyber criminals.

1.4 Questions

i What are the convenient variables to be used in discriminant analysis to predict cyber criminals by Tanzanian police force?

ii Is the discriminant technique convenient tool for predicting cyber criminals by Tanzanian police force?.

II. Literature Review

On- line banks' fraudulent detection tools

The on-line banks' fraudulent prediction can be termed as the supervised. These data are the one which used a database of known fraudulent to build a model which predicts suspects.

The law enforcers like police investigators normally use the existing information from a set of fraudulent or non-fraudulent to develop models to predict fraudulent (Artís, Ayuso, & Guillén 1999; Weisberg & Derrig, 1998). This is emphasized by Belhadji, Dionne, and Tarkhani, (2000) that the classical statistical fraud detection would be used the fraudulent and non-fraudulent credit card data transactions to develop a model that would allow the classification of new claim's likelihood of being fraudulent.

However, other researchers elucidated that in order the discriminant analysis to be more powerful tools to improve its predictions, need to be combined with other algorithms, for example, the neural networks (Ripley, 1996; Hand, 1997; Webb, 1999). Another statistical analytics like tree-based on the algorithms such as CART can be combined with the discriminant analysis to improve its fraudulent predictions (Breiman, Friedman, Olshen & Stone, 1984; Quinlan, 1993). Similar researchers explained the same concept that, the combinations of some or all of these algorithms can be combined together using meta-learning algorithms to form the powerful algorithms that improve the best prediction in fraud detection (Chan, Fan, Prodromidis & Stolfo, 1999). Nevertheless, other researchers have identified the linear discriminant analysis and logistic discrimination to be effective proved tools for many applications, but more powerful tools as it has stated by (Ripley, 1996) because of their powers to determine the impact of multiple independent variables presented simultaneously to predict membership of one or other of the two dependent variable categories. The statistical analytics like discriminant and logistic have used successfully to predict money fraudulent (Hand, 1981; McLachlan, 1992).

In regard with this research, the linear discriminant technique was used two possible levels of the dependent variable. The on- line banks' fraudulent as a cyber criminal and non- fraudulent as a non cyber criminal was considered as the dependent

and independent variables respectively. The use of discriminant analysis values in iterative maximum likelihood estimation was utilized for the purpose of fitting logistic regression models to predict some fraudulent.

III. Methodology

3.1. Sample size

The appropriate sample size, the level of precision, the level of confidence intervals, and the degrees of variability in the attributes were being considered in this research design. The data of on-line banks' fraudulent (cyber criminal) and non cyber criminal employed for this research were collected from the Tanzania police force database.

The Tanzania police force consists criminal database of varies crimes like, cyber crimes, robbery, armed violence, burglary, house breaking, rape, murder, etc. For the purpose of this study; 150 data were collected as on- line banks' fraudulent (cyber criminals) and non criminals. 100 data were cyber criminals out of the total sample size from the year of 2007 to 2010.

3.2. Data Sources

The aim of this research was to predict on-line banks' fraudulent (cyber criminals) from non-criminal within the community of Tanzania related to their ages. The sources of these data were taken as primary data from Tanzania police force data bases to be tested by discriminant analysis to predict the on-line banks' fraudulent (cyber criminals) effects upon the society.

3.3. Data designing

The data designing was based on the statistical analysis theory of discrete and ordinal. According to Weisberg & Derrig, (1998) stated that, the required statistical analytical for fraud investigations were discrete and ordinal, but with no natural metric scale with the useful variable obtained from the subjective assessments of a "claims-wise". In order the collected data to be computed by discriminant algorithms were designed in the form of dichotomous through the training data by the aid of learning algorithm to build a desirable model for testing data accuracy. In regard with this research, data collected were on- line banks' fraudulent (cyber criminal) as dependent variable and non- cyber as the independent variable of which was assigned 0 and 1 minimum and maximum range respectively. The ages of the on- line banks' fraudulent (cyber criminal) were grouped from 20-50 years and 51- 70 as 0 and 1 minimum and maximum range respectively to aid the training data with the learning algorithm to develop the suitable models for testing data accuracy.

IV. Models Development and Hypotheses

4.1 The development of models

In this research the goodness of data fit as well as validity was very useful to the model built hypotheses from the discriminant models to predict on- line banks' fraudulent (cyber criminals) against non criminal's respect to their ages.

4.2 Discriminant model

The on- line banks' fraudulent (cyber criminals) and non- criminal respects with their ages were computed as dependent and independent variables through training data into testing data to build desirable models for accuracy. The on- line banks' fraudulent (cyber criminals) and non- criminal were considered as dependent and independent variables respectively. 0 and 1 were computed as minimum and maximum ranges respectively.

4.3 Hypotheses

In regard with this research, it is hypothesized that the:

i The discriminant technique has the positive effect over the traditional tools used by Tanzania police force to predict online fraudulent crimes.
ii The young offenders have positive effects over the Tanzanian community.

V. The Validation of Discriminant Model.

The training data and learning algorithm were mainly used for building model to verify the accuracy of a predictive data test. When the input variables are many then, the possible ways of combining them is necessary in order to build the model "fit" or the variable data of this method should be over trained. In other words, the model should reflect natural noise to the input data more than actual statistical between the input and output data predictions. To overcome this problem, it is adversely to split these data set into as training set and a testing set so that the output variable will be well-known for each set (Fayyad & Irani, 1993). However, this method fitted the model to the training data where the model was then applied to the data test to observe whether it had about the same predictive accuracy as on the training data or not. This effect gave the accuracy of the model generated by the method.

Results of hypotheses

The analyzed results with their corresponding outputs were resulted from the discriminant analyses.

Hypothesis 1

The discriminant technique has the positive effect over the traditional tools used by Tanzania police force to predict online fraudulent crimes.

Table 5-1 Wilks' Lambda

Wilks' Lambda

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1 dimension0	.395	136.912	1	.000

From the multivariate test—Wilks' lambda, the calculated value was .395. Because of p was < .05, then the model was good fit for the data.

[Table 5-2] Standardized Canonical Discriminant Function Coefficients

Standardized Canonical Discriminant Function Coefficients

	Function
	1
Non- criminals	1.000

The discriminant function (DF) yields out of the equation.

DF=1.000*Noncriminal .....(1).

Using this equation (1), one's scores based on non- criminals can calculate their score on the discriminant function which results the functions at group centroids.

[Table 5-3] Functions at Group Centroids

Functions at Group Centroids

Cyber criminal or Non- criminal?	Function
	1
Non- criminal	1.331
Cyber criminal	-1.134

Unstandardized canonical discriminant functions evaluated at group means

DF = 1.331*non criminal + -1.134*cyber criminal .....(2)

Using the equation (2) above, can calculate the Discriminant Function (DF).

The centroid is the mean values for the discriminant scores for a particular group. The means for a group on all the functions are the group centroids.

If someone's score on the discriminant function is closer to 1.331, then those answers are probably non- criminal. If the score on the DF is closer to -1.134, then the data probably is the cyber-criminal. In practical terms, generally this cut score can be calculated by find their mean in equation (3).

Cut Score =(1.331+ -1.134) / 2 .....(3).

=0.197/2

= 0.0985.

If an individual person's score on the DF (calculated by plugging in their scores on non-criminal and cyber criminal equation above is above 0.0985, then probably is non- criminal. If their DF score is below 0.0985, then probably is the cyber criminal.

[Table 5-4] Classification Statistics

Classification Results[b,c]

		Cyber criminal	Predicted Group Membership		
			0	1	Total
Original	Count	0	51	18	69
		1	0	81	81
	%	0	73.9	26.1	100.0
		1	.0	100.0	100.0
Cross-validated	Count	0	51	18	69
		1	0	81	81
	%	0	73.9	26.1	100.0
		1	.0	100.0	100.0

a. Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.

b. 88.0% of original grouped cases correctly classified.

c. 88.0% of cross-validated grouped cases correctly classified.

The [table 5-4] above shows the overall information about the predicted cyber criminal group membership:

- Overall % correctly classified = 88.0%
- Sensitivity was 51/69 (73.9%)
- Specificity was 81/81 (100%)

From these results; the predictive value (PPV) and negative predictive value (NPV) are calculated below:

- PPV = 51/(51+0)
  =51/51
  = 100%
- NPV = 81/ 81+18
  = 81/99
  =81.8%

Hypothesis 2

The young offenders have positive effects over the Tanzanian community.

[Table 5-5] Wilks' Lambda

Wilks' Lambda

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1 dimension0	.605	74.143	1	.000

The multivariate test—Wilks' Lambda .605 of Chi-square value of 74.143 has $p < .05$, indicating that the model was good fit for the data.

[Table 5-6] Classification Results

Classification Results[b,c]

		Age	Predicted Group Membership		
			0	1	Total
Original	Count	0	44	4	48
		1	25	77	102
	%	0	91.7	8.3	100.0
		1	24.5	75.5	100.0
Cross-validated	Count	0	44	4	48
		1	25	77	102
	%	0	91.7	8.3	100.0
		1	24.5	75.5	100.0

a. Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.

b. 80.7% of original grouped cases correctly classified.

c. 80.7% of cross-validated grouped cases correctly classified.

From the [table 5-6] above shows that:

- The overall, percentile correctly classified was 80.7%
- Sensitivity was 44/48 (91.7%)
- Specificity was 77 / 102 (75.5%)

From these results; the positive predictive values (PPV), and negative predictive values (NPV) are calculated:

- PPV = 44/(44+25)
  =44/69
  = 63.8%
- NPV = 77/ 77+4
  = 77/81
  =95.1%

VI. Implications, limitations, Discussion and Conclusion

6.1 Implications of the study

Among implications from these research findings are considered below:

6.1.1 Discriminant results implications

Results from predicted on- line banks' fraudulent (cyber criminals) against non criminal, the 88% was the overall; sensitivity was 73.9% and specific was 100%. However, the results from cyber criminals relatively with their ages implicated that the overall was 80.7%, sensitivity was 91.7% and specific was 75.5%.

These results of on- line banks' fraudulent (cyber criminals) against non- criminal characteristic tests respect with their ages are useful to the community under the following considerations questions:

Sensitivity: How likely the test detects the presence of someone's characteristic from others characteristic?

Specificity: How likely the test detects the absence of someone's characteristic from others without the characteristic?

Positive predictive value: How likely someone with positive test results to the actual characteristic?

Negative predictive value: How likely someone with a negative test does not result to actual characteristic?

6.1.2 Theoretical implications

A logistic regression analysis was computed and predicted on- line banks' fraudulent (cyber criminals). All these model tests were statistically significant, indicating that the on- line banks' fraudulent (cyber criminals) were tested reliably distinguished from non- criminals respect with criminal's ages.

6.2 Limitations

This research was constrained with some numbers of limitations.

First, its data were only collected from the police station rather than courts.

Courts are ones, which have many criminal databases than police stations because the convictions of criminal are done at courts of laws. In this sense the courts are expected to have reliable data.

Second, this research did not distinguish who is the exactly young person under the law from the adult. For the purpose of this research, it was treated the age of 20-50 to be young people and 51- 70 adult ones, which were not justified by laws.

Third, moreover, the method used in this test was only based on the discriminant analysis to predict cyber criminal and the test results were based on the independent random sample selection of collected data from police stations. Moreover, it was based on the highly sensitivity of the inclusion of outliers with standardized canonical discriminant function coefficients.

6.3 Discussion and Conclusion

With the aid of training data and learning algorithm to build a desirable model for testing data accuracy the researcher developed two hypothesis:-

The first hypothesis: was examining whether the discriminant technique has the positive effect over the traditional tools used by Tanzania police force to predict online fraudulent crimes. According to the results, of the first hypothesis predicted that the overall on- line banks' fraudulent (cyber criminals) against non- criminal was 88%, sensitivity was 73.9% and specific was 100%.

Second hypothesis: was examining whether the young on- line banks' fraudulent (cyber criminals) have positive effects over the Tanzanian community.

The results showed that young offenders with the ages between 20- 50 years old are mostly involved with cyber activities. The results were implicated that the overall was 80.7%, sensitivity was 91.7% and the specific was 75.5%. Some researchers elucidated that, many young people are more involved in fraudulent because of the venturing responsibilities. According

to Ottih, (2000) and Okia, (1994) stated that many young people with the ages of 31-40 years are mostly found involved in fraud, followed by those of 41 years and above because of their serious venture responsibilities.

The results from this research are realistic because the age of 20-50 years is where many Tanzanian is assumed to be serious with venturing responsibilities and not polygamist's culture of having a number of wives and the greater number of children as it might be considered by other people.

With the rapid advance growth of information and communication technology, the law enforcers like police need modern investigative model tools to overcome the traditional models that prevent cybercrimes and other serious crimes. These modern investigative model tools will help to develop methodologies against the prevailing new technologies as they emerge and become the subject of investigations process. Moreover, they will be used as the proactive ways to identify opportunities for the development and deployment of technology to support the work of investigators and capture cyber criminals. Therefore, Law enforcers like police will apprehend more culprits before the courts of laws with scientific evidence to prove beyond reasonable doubts.

References

[1] Artis, M., Ayuso, M. & Guillen, M. (1999). "Modelling Different Types of Automobile Insurance Fraud Behaviour in the Spanish Market", *Insurance: Mathematics and Economics* , 24 (1-2), pp. 67-81.

[2] Belhadji, E. B., Dionne, G., and Tarkhani, F. (2000). A model for the detection of insurance fraud*. *Geneva Papers on Risk & Insurance*, *25*(4), 517-538.

[3] Breiman, L., Friedman, J. H., Olshen, R. A. & Stone, C. J. (1984). *Classification and regression trees*. Belmont, California, U.S.A., Wadsworth Publishing Company.

[4] Chan, P. K., Fan, W., Prodromidis, A. L. & Stolfo, S. J. (1999). Distributed data mining in credit card fraud detection. *IEEE Intelligent Systems* 14(6), 67-74

[5] Fayyad, U. M. and Irani, K. B.(1993). Multi-interval discretization of continuous-valued attributes for classification learning, artificial intelligence, 13, 1022–1027.

[6] Fraud Alerts, (2012). Retrieved 25th November 2013 from http://www.financialtechnologyafrica.com/top-story /223/fraud-alerts-tigo-pesa tightens-mobile-money-security Hand, D.J. (1997). *Construction and Assessment of Classification Rules*.

[7] Chichester: Wiley. Hand, D.J. (1981). Discrimination and Classification. Hand, D.J. (1981). *Discrimination and Classification*. Chichester: Wiley.

[8] Leukfeldt, E. R., Kentgens, A., Frans, B., Toutenhoofd, M., Stol, W.Ph. and Stamhuis, E. (2012). Alledaags politiewerk in een gedigitaliseerde wereld. Handreiking voor delicten met een digitale component. Handbook of crimes with a digital component Den Haag: Boom Lemma Uitgevers

[9] McLachlan G.J. (1992). Discriminant Analysis and Statistical Pattern Recognition. New York: John Wiley and Sons

[10] M-pesa. (2010). Retrieved 5th November 2013 from https://www.google.co.uk/#q= Photo+of+M-pesa+send+pesa+by+phone

[11] NMB mobile. (2014). Retrieved 7th December 2013 from http://www.nmbtz. com/index.php?option=com_content&view= article&id=16 4&Itemid=205

[12] Okia- Anie, I A. (1994) "Venture Characteristics and success: A study of selected Entrepreneurships in Port Harcourt" Unpublished M.B.A thesis, Rivers state Univeristy of science and Technology, Port Harcourt. Nigeria

[13] Ottih, L. O (2000) Entrepreneuring: Toward Preparedness "Pearl publishers, Port Harcourt, Nigeria.

[14] Quinlan, J. R. (1993). *C4.5: programs for machine learning*. San Mateo, CA, Morgan Kaufmann.

[15] Ripley, B. D. (1996). *Pattern recognition and neural networks*. Cambridge University Press

[16] Sullivan & Perry. (2004). Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations

[17] Wall, D.S. (2007). Cybercrime: The Transformation of Technology in the Networked Age. Cambridge: Polity Press Webb, A.R. (1999). *Statistical Pattern Recognition*, London: Arnold

[18] Weisberg, H. I., & R. A. Derrig. (1998). Quantitative Methods for Detecting Fraudulent Automobile Bodily Injury Claims, Risques, 35: 75-101