Developing a fraud prediction model: Application of artificial intelligence methods using firm-specific data and locational factors

Jim Ericson B. Kung
De La Salle University
Manila, Philippines
jim_kung@dlsu.edu.ph

Jeneva Marielle Apolinar
De La Salle University
Manila, Philippines
jen.marielle@gmail.com

Jill Irish Ramirez
De La Salle University
Manila, Philippines
ramirezjill@yahoo.com

Winona Rebadomia
De La Salle University
Manila, Philippines
rebadomia_winona@yahoo.com

ABSTRACT

Corporate fraud, just like any irregularity, can start insignificant; but when blown out of proportion, may cause devastating effects beyond anyone's expectations. After several financial crises and economic collapses, the cycle of corporate malpractice continues to plague the business world, with people still failing to grasp its significance. This study emphasizes the need for early detection, thus aiding in keeping fraud occurrence to a minimum. This research sought to develop a forecasting model that could predict the occurrence of fraud in companies based on publicly available financial and locational information, specifically: current ratio (CR), total asset turnover (TATO), return on assets (ROA), debt to asset ratio (DAR), current asset to total asset ratio (CATA), corruption perception index (CPI) and gross domestic product (GDP). The model was estimated through the use of logistic regression and artificial intelligence (AI) models – specifically, fuzzy logic and neural network. These AI models, albeit being more commonly used in the context of engineering and computer science, have proven to be accurate in terms of predictive modeling. Furthermore, the study evaluated which among the three models is the most accurate in predicting fraud occurrence. Logistic regression was used to analyze the independent variables to determine their significance. Findings showed that all the variables, apart from TATO and CATA, were significant. Among them, CR, TATO, DAR, and CATA have a positive effect on the probability of fraud occurrence while ROA, CPI, and GDP have a negative effect on fraud occurrence. The three prediction models were then developed and compared to determine which is best for use by shareholders. Results showed much promise, with the neural network model proving to be the most precise, given the limitations of the study. It produced an accuracy rate of 74%, based on the F-score computed from the confusion matrix. 14.9% were non-fraudulent observations erroneously classified as fraudulent. 12.5% were fraudulent observations erroneously classified as non-fraudulent. This gives the model a total error rate of approximately 27%. The results demonstrated that the model serves its function effectively in predicting fraudulent financial statements, proving that the model could be of assistance in mitigating the widespread effects of fraud.

Keywords: fraud, artificial intelligence, locational factors

INTRODUCTION

From the high-profile accounting scandals of Enron Corporation to the seemingly harmless corporate decisions of Lehman Brothers, history has taught us time and again how devastating the consequences of unethical business practices can be.

Fraud can be defined as the intent to misrepresent, conceal, or omit truths for the purpose of deception or manipulation to the financial detriment of an individual or an organization (Akinyomi, 2010). The propensity and manifestation of fraud vary among the different levels of an organization, with most employees only being able to carry out simple theft, while corrupt managers are able to embezzle large amounts of funds or manipulate financial statements for their benefit.

In recent years, multi-billion dollar companies have been a hub for fraudulent activities; the most popular of them all being Enron. The publicity of Enron's failure paved the way for the need for stronger controls, policies and procedures in a company. As newer technology becomes available, these controls, policies and procedures are brought up to date. However, as these evolve, the fraudsters evolve with it, updating old methods and finding out ways to bypass illegal practices through the use of loopholes. This has raised the need for new and innovative ways to utilize existing technology to prevent and detect fraud.

As pervasive and inherent as fraud go, the researchers were incited to check if it were possible to develop a model that would aid in predicting the occurrence of fraud, emphasizing the need for early detection as a way to keep fraud to a minimum.

This study seeks to develop an effective and accurate fraud prediction model that will help predict if a company is susceptible to high fraud risk using financial metrics and locational factors. It answers three specific problems:

- How do financial metrics specifically, current ratio, total asset turnover, return on assets, debt to asset ratio, and current assets to total assets ratio - help predict the financial statement fraud occurrence?
- Do locational factors - Gross Domestic Product (GDP) and Corruption Perception Index (CPI) - have a significant impact on the frequency of fraud occurrence?
- Which among these methods - Regression analysis, Fuzzy Logic, and Neural Network - is the most effective and the most accurate in developing a fraud prediction model?

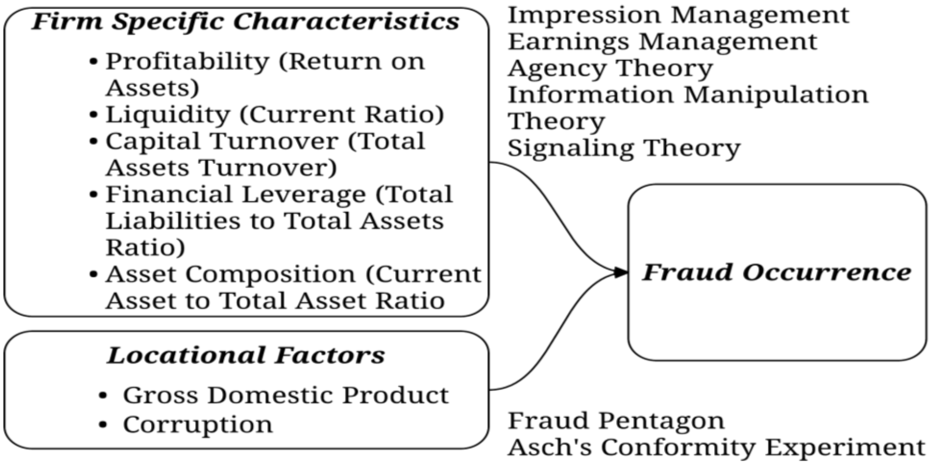
The objectives of this study are consistent with the main goal of developing a fraud prediction model. With this, the study aims:

- To determine which, among Regression, Fuzzy Logic, and Neural Network, is the most effective in developing a fraud prediction model;
- To establish the impact of financial metrics, specifically *current ratio, total asset turnover, return on assets, debt to asset ratio, and current assets to total assets ratio*, to occurrence of financial statement fraud, and
- To assess whether locational factors - *GDP* and *CPI* - have any significant impact on fraud occurrence.

FRAMEWORK

Figure 1 exhibits the different variables that are probable determinants of fraud occurrence. It also shows the different theories that supports the effects of these variables to occurrence of fraud. These are divided into 2 classifications, according to the factors the variables under each group pertain to.

Figure 1. *Schematic Framework of the Study*



LITERATURE REVIEW

Nature and Occurrence of Fraud

Perpetrators. Contrary to one might think, fraud can be performed by anyone in an organization. According to a 2012 study by the Association of Certified Fraud Examiners (ACFE), they are commonly aged thirty-one to forty-five (31-45) and the instances of male fraudsters are higher than that of females. This is due to the assumption that more men still hold higher positions than women (Coenen, 2008). The report also shows that fraud committed by executives cost greater and harder than fraud committed by a regular employee. However, there are still those fraudulent acts committed by external persons. Other motives may also include but is not limited to, nature and scope of the job, tools/trainings provided, reward recognition system and ethical climate (Singleton, 2010).

Victims. Fraud is most prevalent in companies where there are no controls, no trust, no ethical standards, no profits and no future (Singleton, 2010). Most commonly victimized industries include the banking and financial services sector, the government and public administration sector and the manufacturing sector (ACFE, 2012).

Theories of Occurrence of Fraud

The manifestation of fraud cases has impelled researchers to study these cases and provide theories that will understand the occurrence of fraud, which will be used to prevent fraud from occurring. One of the most notable theories that explain fraud is the fraud triangle, which identified the three key elements of fraud – opportunity, motivation and rationalization. *Opportunity* is an environment or an impermanent situation that allows for fraud to be committed. *Motivation* is also known as the motivation to actually perpetrate the fraud. *Rationalization* is a typical state of conformity, where the individual replaces their own desires for what they believe to be the “greater good”, especially of the organization. The fraud triangle was further developed to the fraud diamond wherein a new element to committing fraud was added, namely capability. *Capability* means having the skills and knowledge to actually perform the fraud. The fraud diamond was also further developed into the fraud pentagon, an exploratory research where external influences were considered in the occurrence of fraud.

Methods of fraud

There are many techniques used by individuals in order to deceive intentionally financial statement users for financial gain. The Association of Fraud Examiners (ACFE) enumerated three examples of the methods in which fraud can be committed, one of which is the asset misappropriation. In this, an individual misuses or steals an organization's assets. The second type is the corruption. It involves the dishonest or illegal behavior especially by powerful people. This may include accepting inappropriate gifts, manipulating business transactions, etc. Lastly, ACFE considers fraudulent statements as a method to commit fraud and by far, the most common method of fraud by high profile companies. This simply include fictitious revenues, improper sales cut-off, failure to record an expense from a current transaction, etc.

Red Flags

Financial metrics. Numerous studies have been conducted in an effort to find effective methods of identifying indicating potentially fraudulent activities. One of proposed methods is ratio analysis.

Various researchers have tried to determine whether or not financial ratios computed from financial statements of fraudulent companies significantly differ from those of nonfraudulent companies. Persons, O. (2011) identified the ten financial ratios and variables commonly used to measure a firm's financial condition falling under the following categories: financial leverage, profitability, asset composition, liquidity, capital turnover, size, and overall financial position. The author matched samples of fraud and non-fraud firms on the basis of industry and time period. The resulting model correctly identified a large percentage of the fraud firms.

Locational Factors. Corrupt practices are seen to be more prevalent in many emerging and transitional economies today (Wilhelm, 2002). One of the prevailing reasons, as stated by Wilhelm (2002), is that in less developed countries, the cost of living many be high compared to the low salaries. Countries having a higher income may be able to devote more funds to fighting corruption. Corruption puts a negative effect on trust as this hinders commercial activity. This would result to people being less likely to buy, sell or invest in companies with a high level of corruption. This leads to less economic growth, and these forces managers to use "creative accounting" in an effort to attract investors. In addition to corruption, scandals and unethical behavior also reduces efficiency and fairness in markets; thus, decreasing the ability of the firms to adapt to complexity and change (Korsgaard, Schweiger, & Sapienza, 1995; McAllister, 1995).

Another factor to be considered is the level of external regulation. In the United States, Paul Sarbanes and Michael Oxley drafted the Sarbanes-Oxley Act (SOX) in 2002 to protect investors by improving the accuracy and reliability of corporate disclosures made to the public. It established a higher standard for corporate accountability by placing heavier penalties for acts of wrongdoing. It has also drastically changed the role corporate executives by requiring all financial reports to include an internal control report, which is the responsibility of the management, such as sections 302, 404, 802, and 807.

Statistical and Artificial Intelligence Models

Regression. Regression models are statistical tools used for investigating the linear relationship among variables (Gujarati & Porter, 2009). As stated by Sharma and Panigrahi (2012), the logistic regression model is normally used in literature since it can detect fraud up to 95.1% accuracy with significant expectation effect. Although, discrete variables must be transformed into a continuous value which is a function of the probability of the event to occur (Gujarati & Porter, 2009).

Bell and Carcello (2000) also developed a model based on logistic regression. In their study, several fraud risk factors are considered, some of which being: weak internal control environment, company growth, profitability, aggressive attitude of management, interaction with the environment, etc. Spathis (2002) has also employed the logistic regression method in order to come up with a fraud detection model. He used two different input vectors that contain financial ratios. Their report showed that the accuracy rate of the said model is beyond 84%. This simply proves that there really is a potential in detecting fraudulent activity by just using a company's published financial statements.

Neural Network. Neural network is a computational model that is loosely based on the neuron cell structure of the biological nervous system (Janztzen, 1998). This computational model functions similarly to the natural neurons in the human brain. The process of neural network consists of an input, which can be compared to a synapse in the human brain. Such inputs are multiplied by weights and then computed with a mathematical function, which determines the activation of a neuron (Gershenson, 2013). The calculation of the weights and the hidden layers will lead to the creation of the output.

Neural network offers many advantages relative to other artificial intelligence that can be used in detecting financial fraud. For one, neural network is adaptive- the system's limitations automatically adjust in order to come up with the correct output for a given input. Thus, this approach less likely is to be affected by accounting manipulations (Feroz, 2000). Second, neural networks can approximate nonlinear functions and naturally models multivariable systems.

A previous study revealed that input vector consisted of financial ratios and qualitative variables, was more effective when fraud detection model was developed using neural network. The model was also compared with standard statistical methods like linear and quadratic discriminant analysis, as well as logistic regression methods (Fanning, 1998).

Another paper illustrated the application of artificial neural networks to test the ability of selected Statement on Auditing Standards (SAS) No. 53 red flags to predict the targets of the SEC investigations. (Feroz, 2000) SAS No. 53 lists several categories of red flags including personnel, financial and audit-oriented red flags. The ANN model successfully distinguishes reporting violators from matched control firms in 81% of the possible cases.

Fuzzy Logic. Fuzzy Logic is a mathematical technique developed by Prof. Lotfi Zadeh in 1965 that is used in evaluating the innate vagueness and imprecision of natural language, as well as the extremely qualitative and ambiguous thinking of humans. Numerous other studies have applied fuzzy logic in various areas of management, business and finance.

In 1998, Deshmukh and Talluru created a rule-based Fuzzy-reasoning system for assessing the risk of management fraud using red flags as linguistic variables that fall under three main categories: condition, motivation, and attitude. The final results proved to be less accurate than previously tested statistical methods.

Ammar, Wright, and Selden (2000) created a multilevel Fuzzy rule-based system in ranking state financial management. Using fuzzy set theory, they successfully developed a model to measure the effectiveness and rank the performance of state financial management. A similar study was conducted by Chai, Hoogs, & Verschueren (2006). They proposed a method to convert a binary rule-based decision model derived using genetic algorithm into a model that generates fuzzy scores with the objective of ranking company financial statements in the set of potentially fraudulent companies.

METHODOLOGY

The study utilized two pools of data, one for the fraudulent companies and the other for the non-fraudulent companies. The fraudulent population used in this study consists of all publicly-listed and delisted companies with a *reported* material incidence of fraud. The non-fraudulent population consists of companies without any reported material fraud case in their entire existence during the time this study was conducted.

This study used a combination of convenience sampling, wherein members of the population are selected based on their opportune accessibility, and simple random sampling, wherein each member of the population has an equal opportunity of being selected. In total, the group has identified more than one hundred (100) fraudulent companies, from which the researchers have selected forty-five (45), taking into consideration the number of companies per region – Asia, North America, and Europe. The same procedures were used for the non-fraudulent companies, giving a total of ninety (90) companies in the sample. These companies are matched according to the years, industry, region, and any financial ratio within the +/-10% range for any given year. For the purposes of developing the statistical and artificial intelligence models, thirty (30) pairs of fraudulent and non-fraudulent companies were used as the modeling data. The remaining fifteen (15) pairs were divided into training data, for the neural network model, and testing data. Financial information of each company over four (4) years was collected, giving a total of three-hundred and sixty (360) observations.

An analysis of the financial performance of the companies was performed based on the financial metrics of that company covering a period of four years. The researchers have set Year 4 as the year when the fraud was uncovered and Years 1, 2 and 3 as the immediate 3 years prior to the date of the fraud. The financial metrics used for analysis are:

- Current Ratio (Liquidity)
- Total Assets Turnover (Capital turnover)
- Return on Assets (Profitability)
- Total Liability-Total Asset Ratio (Financial Leverage)
- Current Asset-Total Asset Ratio (Asset Composition)

Aside from the financial metrics, locational factors based on the geographical location of the company, specifically the Gross Domestic Product (GDP) and the Corruption Perception Index (CPI), will be included in the study.

In order to achieve the objectives of this research, three models were estimated: the panel logistic regression, the Adaptive Neuro-Fuzzy Inference System (ANFIS), and the back-propagation algorithm neural network. The study utilized two statistical software in estimating the models – *Stata*, for estimating the panel logistic regression, and *Matlab*, for estimating the fuzzy logic and neural network models.

A logistic regression analysis was utilized to determine which factors would have a significant impact on the frequency of fraud occurrence. It uses the maximum likelihood approach in estimating the model (Gujarati & Porter, 2009), wherein calculus is used in finding the smallest possible deviance between observed and predicted values. The logistic regression uses different iterations to get the smallest and possible deviance or best fit.

After estimating the model, the coefficients provided is the odds ratio. The bigger the difference between one (1) and the observed odds ratio, the stronger the relationship of the dependent and independent variable is. When the odd ratio is above 1, increasing the independent variable increases the odds that the dependent variables equal 1. Inversely, when the odds ratio is below 1, increasing the independent variable decreases the odds that the dependent variable equals 1. On the other hand, when the odds ratio equals to 1, the independent variable has no effect on the dependent variable (Reyna, 2014).

Another model developed in this study is the Neural Network model. For this study, the researchers used the Backpropagation Algorithm, specifically the Cascade-Forward Backpropagation and Feed-Forward Backpropagation, which is one of the most studied and used algorithms for neural networks learning. The Backpropagation method uses output errors in the output layer to estimate the error in the direct leading layer, and use the leading layer's error to estimate the previous layer and so on and so forth. In the learning process, each neuron changes its joint weights according to specific rules and finally makes output closer to the expected output (Chao Xi & Erihe, 2013). The main difference between the Cascade-Forward and the Feed-Forward is that the Cascade-Forward only follows one straight path by creating weighted connections from input layer to hidden layer and finally to the output layer. On the other hand, the cascade forward backpropagation model is similar to the feed-forward but it has an additional weighted connections from the input layer to the output layer and the succeeding hidden layers (assuming there are more than 1 hidden layer).

The final artificial intelligence model developed in this study is the Adaptive Neuro-Fuzzy Inference System (ANFIS). Still fundamentally a fuzzy logic controller, ANFIS combines the self-learning capabilities of Artificial Neural Networks (ANN) and the human-like decision-making using membership functions of the Fuzzy Logic Inference System (FLIS). Membership functions describe the degree of which an input belongs in a fuzzy set. For example, a temperature of 37.3°C may be classified as partially *high* and partially *normal*. The fuzzy sets are *high* and *normal,* and their membership functions map the input (temperature) to a value ranging from 0 to 1 for both sets. One means complete membership to the fuzzy set. Zero means that the input is not a member of the fuzzy set. Any value between one and zero means partial membership to the fuzzy set. Meanwhile, fuzzy rules are if-then rules comprised of the premise and the conclusion. For example, in the context of this study on fraud, a very basic fuzzy rule can be 'if profitability is *high*, liquidity is *low*, capital turnover is *high*, financial leverage is *low*, asset composition is *current*, and locational factors are *risky*, then risk of fraud is *high*'. Using the ANFIS application in MATLAB, the user can input a data set containing the input variables and corresponding output variable and automate the formulation of rules and membership function, depending on the configuration set by the user.

In selecting which of the discussed models would be most effective in predicting the occurrence of financial statement fraud, the models developed were tested using a new set of fraudulent and non-fraudulent companies that are not part of the sample used to develop the models. The results of which will be analyzed using the following confusion matrix to assess if the models estimated may be used in predicting fraudulent companies.

Table 1. Confusion Matrix

		Predicted	
		Positive	Negative
Actual	Positive	True Positive	se Positive (Typ Error)
	Negative	se Negative (Type Error)	True Negative

PRESENTATION OF FINDINGS, ANALYSIS, AND IMPLICATION

Regression results

The logistic regression model was first estimated for the study with the following equation:

$$fraud = \beta_0 + \beta_1 CR + \beta_2 TAT + \beta_3 ROA + \beta_4 TLTA + \beta_5 CATA + \beta_6 CPI + \beta_7 GDP + \varepsilon$$

Table 2. Coefficients of the model based on the logistic regression result

Variable	Marginal Effects	P-Value
	7193 *	
	3808	
	)935 **	
	2357 *	
	4471	
	)4842 **	
	K10^-7 **	

**significant at 1% level
*significant at 5% level

The result showed that all of the variables, apart from TATO and CATA, were significant. Among them, CR, TATO, DAR, and CATA have a positive effect on the probability of fraud occurrence while ROA, CPI, and GDP have a negative effect on fraud occurrence.

The implications of these variables are shown below. Note that the implications provided below are only possible scenarios given the result of the regression.

Table 3. Summary of Implication

Variable	Implications
CR	Typical fraud schemes involve the overstatement of assets and understatement of liabilities, which may be done through timing differences, improper asset valuations, concealed liabilities and expenses, improper disclosures, and others (ACFE, 2014).
TATO	Revenues are commonly overstated to promote the image of a company, which may be done through the recording of fictitious revenue, timing differences, premature revenue recognition, and others (ACFE, 2014).

ROA	Expenses may be overstated to account for asset misappropriation through billing schemes, expense reimbursement schemes, check tampering schemes, payroll schemes, and others (ACFE, 2012).
DTA	Asset misappropriation schemes will initially overstate liabilities, which may be accomplished through billing schemes, expense reimbursement schemes, check tampering schemes, payroll schemes, and others (Hall, 2001).
CATA	Firms with higher receivables and inventories have a higher risk of undetected false financial statements (Simunic ,1980).
CPI	The concept of Normative Influence states that people will change their behavior and actions to conform with the majority (Larsen, 1990).
GDP	During times of economic distress, people will be more inclined to commit fraud in order to keep their businesses afloat. (Australian Institute of Criminology, 2011).

The marginal effects results show that DTA has the highest positive effect on the probability of fraud occurrence, having a 34.02% increase in the possibility of fraud for every increase of 1.0 unit in the value of the variable. TATO has the second highest effect on the probability of fraud occurrence (7.34%), followed by CR (5.57%) and CATA (0.4447%).

The marginal effects results show that ROA has the highest negative effect on the probability of fraud occurrence, having a 201.09% decrease in the possibility of fraud for every 1.0 unit increase of the ROA. CPI has the second highest effect (12.05%), followed the GDP (6.38X10^-5%).

Creating and Testing the models

After estimating and analyzing the logistic regression model, the neural network and the fuzzy logic models were created using the program *Matlab*. These models will then be analyzed and compared against one another, on how well they are able to predict correctly the occurrence of fraud in a company. This will be done through the use of the confusion matrix.

Before showing the results of the analysis, it must first be noted the that the logistic regression model was not analyzed further since the model was only able to provide 14 results out of 56 that are between 0 and 1. Thus, making the logistic regression model unable to be further analyzed in the confusion matrix.

The summary of the analysis of the confusion matrix are shown below.

Table 4. Summary of Analysis of the Confusion Matrix

	NN1*	**NN2****	**Neuro Fuzzy**
Sensitivity (recall)	0.75	1	0.6786
Specificity	0.7143	0.2857	0.7857
Likelihood Ratio +	2.625	1.4	3.1667
Likelihood Ratio -	0.0952	0	0.4091

Positive Predictive Value (Precision)	0.7241	0.5833	0.76
Negative Predictive Value	0.7407	1	0.7097
F-Score	0.7368	0.7368	0.7170

*Cascade-Forward Backpropagation
**Feed-Forward Backpropagation

Both the sensitivity and specificity show how well the models classify fraudulent and non-fraudulent firms in the testing dataset. As seen from the table above, the NN2 model has the highest sensitivity having a value of 1. This means that the NN2 was able to 100% identify companies with fraudulent financial statements. NN1 ranks second, with a sensitivity of 75%, followed by Neuro Fuzzy, which has a sensitivity of 67.86%.

The order of ranking was reversed when checking for the specificity. The Neuro Fuzzy has a 78.57% specificity. This means that the Neuro Fuzzy was able to correctly classify 78.57% of the non-fraudulent firms correctly. The NN1 and NN2 obtained a specificity of 71.43% and 28.57%, respectively.

The likelihood ratios, both positive and negative, show how likely the model classifies a fraudulent company as fraudulent in comparison to a company who did not commit fraud. The positive likelihood ratio would be best having a value greater than 1 while the negative likelihood ratio would be best having a value less than 1.

Looking at the positive likelihood ratio, it can be seen that all the models have a positive likelihood value greater than 1. This means that all of the models would more likely classify fraudulent firms as fraudulent rather than non-fraudulent. The Neuro Fuzzy model obtained the highest value of 3.17, NN1 with a value of 2.63 and NN2 with a value of 1.4.

The negative likelihood ratios of the models all have a value less than 1. This means that all fraudulent firms are less likely to be classified as non-fraudulent. This tells the same story as the positive likelihood ratio but the models would be ranked inversely with the negative likelihood ratio. The model that has the best negative likelihood ratio is the NN2, with a value of 0, followed by NN1 (0.095), and Neuro Fuzzy (0.409).

After looking at how well the models predicted the testing data set, the study will further analyze both the positive and negative predictive value of the model.

The Neuro Fuzzy has the highest positive predictive value with a score of 76%. This means that the 76% of all the predicted fraudulent companies was actually fraudulent. NN1 follows the Neuro Fuzzy with a value of 72.41%. This means that the model was 72.41% right on all the companies that the model predicted as fraudulent. Last is NN2 with a score of 58.33%. This means that NN2 was only 58.33% accurate in predicting fraudulent companies to be fraudulent.

On another note, NN2 had the highest negative predictive value of 100%. This means that all the non-fraudulent firms predicted by the model all are correctly non-fraudulent. NN1 had a negative predictive value of 74.07%, which means that NN1 was 74.07% correct on all the companies predicted as non-fraudulent. Lastly, the Neuro Fuzzy has a negative predictive value of 70.97%. This means that from all the non-fraudulent firms the Neuro Fuzzy model predicted, 70.97% of those are truly non-fraudulent.

To conclude the analysis of the confusion matrix, the F-score is computed. The F-score is seen as the measure of accuracy in statistical analysis of binary classification. As seen from the results, both NN1 and NN2 have the same F-score, which is also the highest, with a value of 73.68%. This means that both the NN1 and NN2 have an accuracy of 73.68%. On the other hand, the Neuro Fuzzy model has an F-score of 71.70%. This means that the Neuro fuzzy is 71.70% accurate.

Implications of the Confusion Matrix

Forecasting models have traditionally been categorized into three main groups: statistical models, theoretical models, and Computational/Artificial Intelligence models.

According to literature, 64% of case studies used statistical models such as the discriminant analysis model, logit model, probit model and decision tree, 11% used theoretical models such as the hazard model and credit risk model, and 25% used artificial intelligence models such as artificial neural networks, fuzzy logic and genetic algorithm. In terms of analytical capabilities, statistical methods rely on the precision, reliability and accuracy of the variables used, whereas artificial intelligence methods are able to process and interpret inaccurate data and tolerate uncertainty and approximation. Moreover, in contrast to the regression model which takes the effect of each variable independently, the artificial intelligence models takes into consideration the combinations of the values of the variables. This is why the artificial intelligence models were able to recognize a pattern in the values of the modeling data to an acceptably accurate degree.

Comparatively, the results of the models are acceptably accurate, considering that, according to available research, the accuracy of traditional credit scoring models used by financial institutions for forecasting the bankruptcy of customers range from approximately 72% to 77.5% (Dadios, 2012). The accuracy rates of all the artificial intelligence models fall within this range, meaning they are reliable enough for the use of professionals in assessing the fraud risk of companies. Among them, NN1 proves to be the best model for the use of the general risk-averse stakeholder, as it has the highest accuracy rate (74%), and it ranks second in term of both sensitivity and specificity. Additionally, NN2 had the highest possible sensitivity score of 1, meaning it correctly predicted all of the fraudulent companies; therefore, users who are only concerned with the prediction of fraudulent companies may prefer to use this model, as is with the case of regulators who are selecting companies to investigate. The neuro-fuzzy model may also have some use for certain stakeholders who are more specific when it comes to the accuracy of predicting non-fraudulent firms, as its specificity ranks first among the three with 78.57%.

CONCLUSION

Financial metrics aid in predicting the occurrence of fraud because such monitor the behavior of the financials of a company. For example, a sudden significant increase in revenue without a compensating increase in assets might indicate an occurrence of fraud. Of the six financial ratios we used as variables for this paper, the *current ratio, total asset turnover, debt to asset ratio, and current asset to total asset ratio,* have a direct relationship with the occurrence of fraud. In other words, as these ratios increase, the probability occurrence of fraud also increases. When these ratios decrease, the probability occurrence of fraud also decreases. On the other hand, the *return on asset ratio* has an inverse relationship with fraud. As the ratio increases, the probability occurrence of fraud decreases. When it decreases, the probability of fraud occurrence increases.

The results also displayed possible techniques for companies to commit fraud in their financial accounts. For example, current ratio is increased with overstating assets or understating liabilities. On the other hand, a company can show a high total asset turnover ratio by falsified high revenues. High debt to asset ratio may result from asset misappropriation. Current asset over total asset ratio is increased with manipulation of current assets, especially receivables and inventory. A low return on asset ratio may indicate overstatement of assets, billing scheme, expense reimbursement scheme, etc. To complete, asset misappropriation is the most common form of fraud that can be committed by a company.

The locational factors, Gross Domestic Product (GDP) and Corruption Perception Index (CPI), were assessed to have a significant impact on fraud occurrence. Both have an inverse relationship with the probability of the occurrence of fraud. Regarding CPI, when a company is included in a country in which the perceived level of corruption is high (low CPI), the probability of fraud occurrence is also high. Regarding GDP, when a company is included in a country with low GDP, there is a higher incentive for a company to perpetrate fraud to maintain an image.

As for the two Artificial Intelligence models developed in this study, the results proved to be promising. Based on the f-score computed from the confusion matrix, two of the neural network models created, *NN1* and *NN2*, and the *neuro-fuzzy model* all had accuracy rates of over 70%, although NN1 and NN2 both scored the highest with 74%. Their differences lie in their sensitivity and specificity, whether they correctly predicted more positives or negatives. NN1 was the most accurate in terms of the f-score (74%) and ranked second in both sensitivity (75%) and specificity (71%). NN2 also had an f-score of (74%) and ranked first in sensitivity (100%) and last in specificity (29%). The Neuro Fuzzy model ranked slightly lower than the neural network models in terms of accuracy (72%) and ranked first in specificity (79%) and last in sensitivity (69%). In general, stakeholders would be more inclined to use the model that balances risk-aversion and precision; therefore, NN1 was selected as the best model for the general risk-averse user. However, depending on the user's needs in terms of sensitivity and specificity, NN2 and the neuro fuzzy model may still prove to be useful in certain situations.

RECOMMENDATIONS

For corporate directors and shareholders

The management should design a strong set of internal controls that would alleviate the risk of fraud occurrence. In addition to such controls, the management could also integrate the use of the model established in this study in their internal control procedures. Earlier detection of possible fraudulent behavior is more beneficial to the company as it saves them time and cost.

The researchers would like to emphasize that the models developed are merely additional controls that the company might want to take upon themselves to implement. These models do not intend to replace existing internal control procedures, but merely as a supplement to strengthen such procedures.

This study has also proved that the Corruption Perception Index (CPI) is significant in the prevalence of fraudulent activities in a country/region. If corporate directors decide to expand their businesses overseas, the perceived corruption level in that country/region should be taken into consideration, despite that country having the right market and a good credit rating.

For investors

The study covered the profitability, liquidity, capital turnover, financial leverage and the asset composition of a company; and through the methods carried out, was able to show that these metrics do have an effect on the risk of fraud occurrence in a company. The red flags identified through this study may be of assistance to investors as they perform preliminary investigations on the companies they want to invest in. Risk-averse investors would want to avoid investing in companies that have a high fraud risk because this will affect their returns in the long term.

For financial institutions

As financial institutions assess a business' credit risk, liquidity risk also forms a huge part in the decision on whether they will allow businesses to make loans, the documentation and legal covenants they will require, and how much they are going to lend out. This study has shown that liquidity, measured by the current ratio, is a significant variable in determining whether there is an occurrence of fraud in a company. In fact, the study showed that high liquidity might indicate high fraud risk. Although being highly liquid may prove to be attractive in the eyes of financial institutions, they should remain skeptical until further assessments are made.

For government agencies

For government agencies like the Bureau of Internal Revenue (BIR), knowing whether a company is into practicing fraudulent activities may be prima facie evidence on whether they are paying the right amount of taxes. As shown in the red flag analysis, a low return on assets ratio may indicate high fraud risk. Having a lower return on assets ratio may mean having little to no net income, thus enabling the company to pay lower taxes. The BIR may further investigate situations like this in order to ensure that companies are not evading taxes.

For the accounting profession

The models used in this study are especially relevant in the realm of risk-based auditing, forensic accounting and fraud examiners in general. The red flags identified in the study can be used as guidance in identifying possible risk of misstatements during the initial risk assessment stage of audit planning. It will help them become more efficient in their work as they can devote more time and resources in obtaining audit evidence in those accounts that have higher fraud risk. Forensic accountants and fraud examiners specialize in uncovering illegal financial practices in companies. In this study, we have a model that has a 74% accuracy rate in predicting fraudulent activities in a company. This model may be integrated with their procedures as well. It can act as a baseline in determining how high of a risk a company is in terms of being viewed as fraudulent.

For the general public

The study may also help in reducing the level of poverty in a country. Fraud, especially corruption, is a major factor in the fight of a nation against poverty. Predicting the occurrence of fraud can aid organizations in monitoring whether the operations performed internally are objective and fair. This would foster employment security by being a detective control that will help management detect fraud occurrence and mitigate its effects, which will then lead to lower possibility of businesses closing down. Additionally, the models can be used to check for improper disclosures of income and assets, which have a big impact on government revenues, which can be used for country development (i.e. alleviation of the poor and infrastructure development).

For the academe and future research efforts

The study has shown that the Artificial Intelligence models are actually relevant to the world of fraud prediction and detection, as NN1 produced the best percentage of accuracy with the when it comes to predicting fraudulent activities in a company. Integrating discussions on the relevance of Artificial Intelligence models may prove to be beneficial in the long run. The academe can also now think of pursuing other similar studies that might make use of the Artificial Intelligence models. In the case of this study, engineering and computer science methods can actually be intertwined with problems that are in the realm of accounting.

Fraud occurrence can only be painted accurately based on the number of cases tackled at a time. With this, the researches recommend that the sample size be increased in order to accommodate other regions such as South America. Having a larger sample size will increase the accuracy of the models and the results in general. Including other regions may also better explain the significance of location when it comes to fraud occurrence. Future studies can also take into consideration the possibility of breaking up the study per industry or per country, instead of being aggregated into one big international bubble, as shown in this research. It will limit the scope of the study, making it very relevant to that specific country or industry as deeper details can be obtained. The researchers of this study also recommend the usage of other Artificial Intelligence models, such as the Genetic Algorithm model. It would be another basis of comparison, aside from the three (3) models already discussed.

SIMULATION

For better appreciation, the researchers applied the best fraud prediction model obtained to three (3) possible real-life situations while presenting the results of the other fraud prediction models estimated in the study for reference purposes.

Case 1:

Mark Lopez is a CPA working for an established audit firm in the Philippines. The busy audit season is just starting and Mark had wanted to try a more efficient way to perform audit procedures on his two (2) clients in the Philippines. By being able to predict possible fraudulent activities undertaken by his clients, Mark felt that this would help him organize his workload by focusing his attention on the company that has possible fraudulent accounts first before the others. Relevant financial data for the year 200A from the two companies are listed below (in millions). *Note: Company A's data has been taken from a real life fraudulent company (One.Tel). Company B's data has been taken from a real life non-fraudulent company. Both company data have not been used before in this*

	Company A	*Company B*
Return on Assets	0.08187	0.07756
Current Ratio	1.17864	1.25739
Total Asset Turnover	1.65646	0.47591
Financial Leverage	0.56445	0.55877
Asset Composition	0.22850	0.16113

Results

	Company A	*Company B*
NN1 Model	0.97688	0.29255
Prediction	**Fraud**	**Non-Fraud**

The results have shown that Company A has the possibility of being classified as fraudulent, while Company B, the opposite. Both companies have been correctly predicted by the model. Given that Company A has a higher possibility of having illegal practices, Mark, should focus his attention on the said company first in order to bring into light the fraudulent practices being undertaken by the company, if there really were after substantive procedures, as soon as possible.

***Results of the other models, for reference purposes:*

	Co. A	*Prediction*	*Co. B*	*Prediction*
Regression	0.53399	**Fraud**	0.458181	**N-Fraud**
NN2	0.32118	**N-Fraud**	0.061485	**N-Fraud**
N-F	4.39	**Fraud**	2.47	**Fraud**

Case 2:
3JW Company is a financial Institution in the Philippines that provides mutual fund services. One of the company's mutual funds invests solely on stocks. Given this, it makes it susceptible to dire consequences if such investments come from a fraudulent company. 3JW Company would want to mitigate the risk of providing zero to negative returns to its investors. The firm enlists the help of experts in order to possibly predict the involvement of two (2) potential investors' in fraudulent activities, before letting them participate in the mutual fund. Relevant financial data for the year 200B from the two companies are listed below. *Note: Company A's data has been taken from a real life fraudulent company (Texaco). Company B's data has been taken from a real life non-fraudulent company. Both data sets have not been used before in this study.*

	*Company A**	*Company B*
Return on Assets	(0.202786)	0.091353
Current Ratio	1.674041	1.530610
Total Asset Turnover	0.455172	1.472051
Financial Leverage	0.341832	0.574235
Asset Composition	0.437548	0.461633

Results

	Company A	*Company B*
NN1 Model	0.67225	0.99036
Prediction	**Fraud**	**Fraud**

The results have shown that both companies have the possibility of being classified as fraudulent. Company A has been correctly predicted as fraudulent. Company B, however, has been predicted to be fraudulent despite it not being fraudulent in the first place. This goes to show that 3JW Company should further investigate or deny the participations of both companies as a way to mitigate fraud risk.

***Results of the other models, for reference purposes:*

	Co. A	*Prediction*	*Co. B*	*Prediction*
Regression	0.53399	**Fraud**	0.458181	**N-Fraud**
NN2	0.32118	**N-Fraud**	0.061485	**N-Fraud**
N-F	4.39	**Fraud**	2.47	**Fraud**

Case 3:

Run Against Tax Evaders (RATE) is a program initiated by the DOF & BIR to investigate and prosecute individuals and/or entities engaged in tax evasion and other criminal violations of the National Internal Revenue Code of the Philippines. The fraudulent activities covered by the RATE program includes offenses relating to income, offenses relating to deductions and other violations such as making false entries in book and records to reduce tax liability. John Gonzales is a BIR employee under the RATE program conducting a preliminary investigation on Company XXX, to establish whether there is prima facie existence of fraud. Company XXX was brought into attention by an anonymous phone call through one of BIR's offices. Relevant financial data for the year 200C from Company XXX is listed below. *Note: Company XXX's data has been taken from a real life fraudulent company (Rite Aid). This dataset has not been used in the study before.*

	Company A*
Return on Assets	(0.09681)
Current Ratio	1.75905
Total Asset Turnover	2.11758
Financial Leverage	0.85105
Asset Composition	0.42844

John's preliminary investigation consisted of verifying whether the allegations towards Company XXX are plausible. He enlisted the help of experts to help him determine the possibility of fraud in Company XXX through its available financial data. Once the prima facie existence of fraud has been established, a formal investigation may be initiated through the issuance of a Letter of Authority (LOA) against the company.

Results

	Company XXX
NN1 Model	0.99822
Prediction	**Fraud**

The results have shown Company XXX has the possibility of being classified as fraudulent. It has been correctly predicted as fraudulent. The anonymous phone call proved to be a good tip as the company being questioned is actually predicted to be fraudulent. After site visits, interviews and other necessary procedures to substantiate the possibility of fraud, BIR can now launch a formal investigation on Company XXX.

***Results of the other models, for reference purposes:*

	Company. XXX	***Prediction***
Regression	0.72316436	**Fraud**
NN2	0.37109	**Non-Fraud**
N-F	3.57	**Fraud**

REFERENCES

Association of Certified Fraud Examiners. (2014). *Financial Transaction and Fraud Schemes.* Retrieved from http://www.acfe.com/uploadedFiles/ACFE_Website/Content/.

Association of Certified Fraud Examiners. (2012). *Report to the Nations on Occupational Fraud and Abuse.* Retrieved from https://www.acfe.com/uploadedFiles/ACFE_Website/Content/rttn/2012-report-to-nations.pdf

Australian Institute of Criminology, 2011). *Fraud Vulnerabilities and Global Financial Crisis.* Retrieved from http://www.aic.gov.au/publications/current%20series/tandi/421-440/tandi422.html .

Akinyomi, J.O. (2010). The Role of Auditors in Fraud Detection, Prevention and Reporting in Nigeria. *Journal of Research in National Development*, *8*(1)

Ammar, S., Wright, R., & Selden, S. (2000). Ranking State Financial Management: A Multilevel Fuzzy Rule-Based System. *Decision Sciences*, *31*, 449-481.

Barney, J.B. & Hansen, M.H. (1994). Trustworthiness as a Source of Competitive Advantage. *Strategic Management Journal, 15*, 175-190.

Bell, T. B., & Carcello, J. V. (2000). A decision aid for assessing the likelihood of fraudulent financial reporting. *Auditing: A Journal of Practice & Theory*, *19*(1), 169-184.

Chai, W., Hoogs, B.K., & Verschueren, B.T. (2006). Fuzzy Ranking of Financial Statements for Fraud Detection. *International Conference on Fuzzy System*, 152–158.

Chao Xi & Erihe. (2013). Artificial neural networks and fuzzy logic in forecasting short term temperature. Telemark University College.

Coenen, T (2008). Essentials of corporate fraud. Hoboken, N.J: John Wiley & Sons.

Dadios, E. (2012). Fuzzy Logic: Emerging Technologies and Applications. InTech. Vol. 1. Pp.259-286

Deshmukh, A., & Talluru, L. (1998). A Rule-Based Fuzzy Reasoning System for Assessing the Risk of Management Fraud. *International Journal of Intelligent Systems in Accounting, Finance & Management*, *7*(4), 223–241.

Gershenson, C. (2013). Artificial Neural Network for Beginners. Retrieved March 1, 2014 fromhttp://arxiv.org/ftp/cs/papers/0308/0308031.pdf

Goldsmith, A.A. (1995). Democracy, Property Rights and Economic Growth. *Journal of Development Studies, 32,* 157-175.

Gujarati, D. & Porter, C. (2009). *Basic Econometrics, 5th ed.* Singapore: McGraw-Hill Higher Education.

Fanning, K., & Cogger, K. (1998). Neural Network Detection of Management Fraud Using Published Financial Data. *International Journal of Intelligent Systems in Accounting, Finance & Management, 7*(1), 21- 24.

Feroz, E.H., Kwon, T. M., Pastena, V., & Park, K.J. (2000). The Efficacy of Red Flags in Pedicting the SEC's Targets: An Artificial Neural Networks Approach, *International Journal of Intelligent Systems in Accounting, Finance, and Management*, *9*(3), 145–157

Hall, J. (2011). Accounting information systems. Mason, OH: Cengage Learning.

Jantzen, J. (1998). *Introduction to Perceptron Networks*. Technical University of Denmark.

Jones, T.M. (1995). Instrumental Stakeholder Theory: A Synthesis of Ethics and Economics. *Academy of Management Review 20*, 404-437.

Korsgaard, M., Schweiger D., & Sapienza H. (1995). Building Commitment, Attachment, and Trust in Strategic Decision-making Teams: The Role of Procedural Justice. *Academy of Management Journal 38*, 60-84.

Larsen, K. S. (1990). The Asch conformity experiment: Replication and transhistorical comparisons. Journal of Social Behavior & Personality 5(4), 163 – 168.

Mayer, P. (1995). Corruption and Growth. *The Quarterly Journal of Economics 110*(3), 681-712.

McAllister, D.J. (1995). Affect and Cognition-based Trust as Foundations for Interpersonal Cooperation in Organizations. *Academy of Management Journal 38*, 24-59.

Persons, O.S. (2011). Using financial statement data to identify factors associated with fraudulent financial reporting. *Journal of Applied Business Research, 11*(3)*,* 38 – 46.

Reyna, O. T. (2014). Predicted probabilities and marginal effects after (ordered) logit/probit using margins in stata. Retrieved from http://www.princeton.edu/~otorres/Margins.pdf

Sharma, A., & Panigrahi, P.K. (2012). A Review of Financial Accounting Fraud Detection based on Data Mining Techniques, *International Journal of Computer Applications*, *39*(1).

Simunic, D. (1980). The pricing of audit services: Theory and evidence. *Journal of Accounting Research Spring, 1980,* 161-190.

Singleton, T (2010). Fraud auditing and forensic accounting. Hoboken, N.J: John Wiley & Sons.

Smith, M. (1982). *Evolution and the Theory of Games*. Cambridge: Cambridge University Press.

Spathis, C.T. (2002). Detecting False Financial Statements Using Published Data: Some Evidences from Greece*. Managerial Auditing Journal*, 179-191.

Wilhelm, P.G. (2002). International Validation of the Corruption Perceptions Index: Implications for Business Ethics and Entrepreneurship Education. *Journal of Business Ethics, 35*(3).