Towards a Machine Learning Approach for Earnings Manipulation Detection

Bilal Dbouk* and Iyad Zaarour

ABSTRACT

Manuscript type: Research paper.

Research aim: This paper aims to enhance Earnings Manipulation Detection (EMD) by applying the Bayesian Naïve Classifier (BNC), a supervised machine learning approach which evaluates and compares the manual auditors' method with a widely applied mathematical model (Beneish model).

Design/ Methodology/ Approach: The data set consists of financial statements of 53 companies over years 2006 to 2009. Three data sets were created: training data set using financial statements from 2006 to 2007, and two test data sets made up of financial statements from 2008 to 2009. The Beneish model and the manual auditors' method are used to test for EMD accuracy as well as to evaluate results. In the process of testing and comparing the two methods, a new layer of supervised machine learning technique namely the BNC is introduced.

Research findings: The analysis of results for the EMD shows that the Beneish model outperforms the manual auditors' method. The results also reveal a higher classification rate (86.84 per cent) when using the Beneish model as compared to the manual auditors' method (60.53 per cent). This difference indicates that the manual auditors' method is less effective in detecting earnings manipulation.

Theoretical contribution/ Originality: The main contribution of this research is the introduction of the supervised machine learning

* Corresponding author: Bilal Dbouk is a Tax Controller at the Lebanese Ministry of Finance (MOF), at the Large Taxpayers Office (LTO) and a doctoral student at the Doctoral School of Law, Political, Administrative, and Economic Sciences (EDDSPAE), The Lebanese University (LU), LU Bldg, Sin-El-Fil, Beirut-Lebanon. Email: BilalD@finance.gov.lb

Iyad Zaarour is an Assistant Professor at the Faculty of Economics and Business Administration, The Lebanese University (LU), LU Campus, Hadath, Beirut-Lebanon. Email: i.zaarour@ul.edu.lb

approach as a new layer in the framework of EMD. This approach can be used to broaden the scope for auditors, forensic accountants, tax controllers, and other manipulation detectors who are involved in the auditing procedures.

Practitioner/ Policy implication: The results of this study will help regulators and practitioners to re-define their overall strategic decision-making in detecting accounting manipulations.

Research limitation: This study devised a framework for EMD using the Beneish model and machine learning approach for companies operating in wholesale liquid fuel industry. Further studies need to be conducted to examine the application of such methods in other industries. In addition, future studies may need to assess the potential of other manipulation detection models which could be applied together with different machine learning tools.

1. Introduction

The last decade has witnessed an upsurge of financial fraudulent behaviours with high profile financial frauds committed by large companies such as Enron, Lucent and Harris Scarfe, Waste Management and Worldcom (Yue, Wu, Wang, Li, & Chu, 2007). This phenomenon has led to an imperative need for effective financial accounting fraud detection. The audit profession recognises the lack in current auditing methods. The continuous integration of technological advancement and businesses calls for a higher tech-savvy procedure for detecting financial frauds. This is because the normal auditing procedure is time consuming as huge volumes of data need to be analysed. This burdens auditors in today's digital economy (Sharma & Panigrahi, 2012). In that regard, effective methods and analytical procedures which can complement and strengthen the various audit analytical procedures when running an audit assignment, is imperative.

In this concern, the Beneish model, a mathematical model that uses financial ratios and information which are extracted from a company's financial statement was considered. This model, when appropriately applied, could highlight possible areas of concern in financial statements where an auditor could use to classify a company as a manipulator or non-manipulator (Hogan, Rezaee, Riley, & Velury, 2008). The informa-

tion which are constructed from a company's financial statement are then used to calculate and create the M-Score which could be used to describe the level of earnings manipulation of the company. The importance of the Beneish model in earnings manipulation detection is further emphasised when it was employed to assess the financial statements of big companies in the West that have failed, such as Enron and WorldCom (Mahama, 2015). In 2004, the report of the Association of Certified Fraud Examiners (ACFE) also set up the requirement for Certified Public Auditors (CPAs) to adopt the model or its ratios during audits when implementing the Statement on Auditing Standards No. 99 (SAS 99). This requirement was established as a step to ensure that financial statements are free from material errors or manipulations (Nwoye, Okoye, & Oraka, 2013).

In addition to the Beneish model, a new approach in earnings manipulation detection namely machine learning was introduced. Unlike mathematical models, the machine learning approach incorporates sophisticated techniques such as data mining to gain more insights into data analytics (Hogan et al., 2008). Data mining has the capability of addressing the weakness of mathematical modelling as it is able to extract useful information from large data sets. Due to this characteristic, data mining could be applied to facilitate the auditors in extracting and discovering the hidden patterns noted in massive volumes of data merely by using the most optimal and cost effective option. This technique is made possible with the development and availability of Modern (Risk-based) Auditing and Computer Assisted Audit Techniques (CAATs), both of which are supported by the Business Intelligence (BI) systems and Generalized Audit Software (GAS) (Hematfar & Hemmati, 2013). Data mining has increasingly been a subject of interest in the auditing process and a variety of techniques including the Bayesian Naïve Classifier (BNC), Logic Base Algorithm (Decision Trees, Rule Base Algorithm), Neural Network Based Machine Learning (Single Layered Artificial Neural Network, Multi-layered Neural Network) and Support Vector Machines (Kotsiantis, 2007; Anto & Chandramathi, 2011) are developed for this purpose. Despite the existence of the various techniques, scholars (e.g., Kotsiantis, 2007; Anto & Chandramathi, 2011) have argued that no single learning algorithm can uniformly outperform other algorithms with all data sets. Yet, the decision to use the algorithm depends on the type of classification problem and the accuracy of other algorithms that best fit the phenomenon. In a similar vein, some researchers (Saad, Zaarour, Bejjani,

& Ayache, 2012; Kotsiantis, 2007) have also indicated that the BNC is a well-known representative of statistical learning algorithms which has a high classification accuracy. It also contains an effective classification tool which is easy to interpret and it is widely used in detecting banking and financial frauds.

Given the blending point of machine learning and mathematical modelling, this study incorporates both the mathematical model and the machine learning approach for the purpose of devising a framework for earnings manipulation detection (EMD). This study specifically aims to assess a mathematical model that has been widely used in EMD, namely the Beneish model to establish a dependable indicator of accounting manipulation for fraud detection procedures by showing the power of the model when compared to current normal auditing procedures. The classification power of the Beneish model was tested and then compared with the results obtained through the traditional manual auditors' method. Unlike prior studies, a new machine-learning technique, Bayesian Naive Classifier (BNC) is added as a layer. It is used as an approach to assess the Beneish model and the manual auditors' methods procedure. For this to happen, the classification of the financial data is applied. Since the BNC is best fitted for the current study's EMD classification problem, it would also serve as the new supervised machine learning layer for assessing the results. The M-Scores and the manual auditors' results extracted from this study are then set as the parameters of the constructed BNC network. The data sets of this research consist of 53 financial statements taken from Lebanon's largest companies involved in the wholesale liquid fuel industry, over years of 2006-2009.

This paper is organised as follows. Section 2 reviews prior literature on traditional and modern audit methods, earnings manipulation, Beneish model and supervised machine learning technique. Section 3 explains the methodology employed. Section 4 presents the findings and discussion. Section 5 brings the paper to conclusion.

2. Literature Review

2.1 Traditional vs Modern Audit Methods

Prior to the 1980s, traditional audit methods relied on detailed verifications of accounts which did not have any sampling or testing techniques in preparing a complete balance sheet that is free from errors. As time

goes by and economies grow, auditors assessed the truth and fairness of the financial statements of companies based on the sampling techniques developed (Lee & Ali, 2008). With time, the upsurge in the number of transactions and complexity of companies have led to the importance of internal control systems. In the early 1980s, those systems were found to be more expensive and auditors started to make greater use of analytical procedures (Salehi, 2008). Starting from the mid-1980s, modern auditing or risk-based auditing was adopted where more focus was given to those areas that were more likely to contain errors. Auditors started to examine audit evidences that were derived from a wide variety of sources. Consequently, auditors moved to detect, report and assess fraud explicitly in conformance to regulators' increasing concern on corporate governance issues.

Traditional auditing methods which include manual verifications, document vouching by sampling, inventory counts, and the use of simple statistics and/or ratios (Kuenkaikaew & Vasarhelyi, 2013) are regarded as conservative. They provide little value in the modern business environment as they are slow and backward (Verner, 2012). In addition, one of the big weaknesses of traditional auditing methods is that they cannot completely fulfill audit verification needs. This is because most of those methods, particularly substantive tests, are done manually with limited sample data (Kuenkaikaew & Vasarhelyi, 2013). Unfortunately, today, many auditors are still using the traditional manual auditing methods. They use their analytical procedures by incorporating vouching in every transaction, casting, computing basic statistics and ratio analyses which further complicate the process of auditing under modern economy.

In today's environment, modern auditing methods such as continuous audit (CA) or real time audit or risk-based audit or predictive audit can solve traditional manual audit work problems (Kuenkaikaew & Vasarhelyi, 2013). Continuous audit allows predictive audit which not only considers historical data, but also uses analytical method to predict the expected future outcome of process performance for improved control and faulty transaction prevention. While con-tinuous audits allow auditors to make good decisions, most auditors are still trapped in bad decision-making due to information ambiguity represented by data insufficiency and complexity (Utami & Nahartyo, 2016). The advancement of information technology (IT) however, has made the continuous auditing process more effective and efficient as it helps to speed up the audit process through automation. Moreover,

modern technology also enables auditors to extend the process to cover the whole population in a timely manner. With the advances and continuous development of technology, auditors are expected not only to enhance the credibility of the financial statement, but to provide value added services as well as advising management (Lee & Ali, 2008). Auditors today need to implement modern analytical methods and to benefit from the acceleration and automation of business IT for the purpose of predicting the expected future outcome of the performance process of every transaction.

In addition to computing basic statistics and ratio analysis that are widely used through the traditional approach, the continuous audit approach incorporates sophisticated methods such as data mining and machine learning techniques (Omar, Koya, Sanusi, & Shafie, 2014). This could assist auditors in fraud detection and prevention better. Rather than just storing data, machine learning and data mining techniques facilitate in ensuring the computers behave intelligently. Thus, data, mathematical models and machine learning tools play an important role in the continuous audit method. This is because continuous audit method relies on prediction model that will learn behaviour from historical data and uses this to predict outcomes. In this regard, different machine learning techniques can be applied to data to create predictive models (Kotsiantis, 2007). Therefore, more machine learning oriented methods and continuous auditing concepts are incorporated into new audit methods. Many fraud identification proactive approaches, under BI tools, have primary considerations to address for the purpose of minimising fraud in organisations (Omar et al., 2014). Consistently, this trend reshapes auditing towards a more risk-based auditing, CAATs and CA methods. CAATs using computer technology, machine learning and BI may make it possible to test the entire population rather than a sample. Further, as CA can create new procedures in conducting financial audits, it enjoys advantages over traditional audit methods (Moorthy, Mohamed, Gopalan, & San, 2011).

According to Sayana (2003) and Tumi (2013), Generalized Audit Software or General Purpose Audit Software (GAS) are the main types of CAATs that comprise data investigation, statistical tools and sampling techniques. This software can perform tests on missing sequences, statistical analysis and calculations. Thus, this tool could assist auditors in detecting misstatements in financial statements for the purpose of achieving the general audit objectives of validity, completeness, ownership, valuation, accuracy, disclosure and classification of system

outputs. Moreover, with the traditional method, auditors have the exhausting task of manually verifying many accounts such as payroll and earnings without enough coverage. This makes the audit process a time consuming task. These flaws, however, could be addressed with the GAS, since the tool permits the coverage of all the doubtful cases with minimal effort and time (Sayana, 2003).

Hematfar and Hemmati (2013) compare risk-based and traditional auditing and their effect on the quality of audit reports. Their results of the t-test show that risk-based auditing creates significantly more valid and reliable results. It also produces significantly higher quality reports. Thus, it is important for auditors to be knowledgeable and IT savvy as this skill will facilitate them in using IT tools and techniques such as CAATS to detect misstatements and to avoid the possibility of issuing an erroneous opinion (Vasile-Daniel, 2010).

Figure 1 depicts a modern audit method which introduces a new layer of machine learning in the framework of earnings manipulation detection (EMD). The traditional manual approach is still needed despite the disadvantages. Figure 1 shows that traditional methods can be modernised through the application of mathematical models and machine learning techniques. Adding a new layer of machine
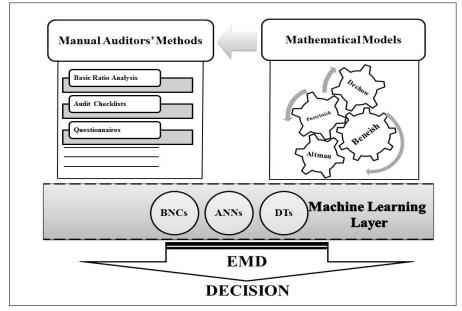

Figure 1: A Diagram of Machine Learning Layer for EMD

learning technique such as the Bayesian Naïve Classifiers (BNCs), Artificial Neural Networks (ANNs) or Decisions Trees (DTs) after using mathematical models such as Beneish, Pustylnick, Dechow, etc. within (represented by the Left Arrow) audit methods will strongly enhance auditors' decision in the framework of EMD.

2.2 Earnings Manipulation

The last few decades have seen a significant growth in academic research which focused on the areas of earnings manipulation. According to Paolone and Magazzino (2014), earnings management represents keeping the accounting practices within the limits of legality following the accounting rules and principles set by the Generally Accepted Accounting Principles (GAAP). However, earnings manipulation involves violating such rules and principles by applying aggressive earnings management over the boundaries of GAAP through intervention in the external financial reporting process with the intent of obtaining some private gains (Paolone & Magazzino, 2014; Drabkova, 2016). Earnings manipulation is an illegal exercise and an accounting fraud. Despite stringent accounting rules and standards set by GAAP, companies are still involved in earnings manipulation by abusing the flexibility given to them by GAAP when reporting their earnings (Sankar & Subramanyam, 2001; Miller, 2009).

Prior studies on this topic provide fruitful discussions of accruals accounting and the development of models based on accruals. Various mathematical models and financial ratios which could be utilised to determine the level of earnings manipulation are discussed, with several models highlighted. Some of these include models developed by Healy (1985), De Angelo (1986) and Jones (1991). Beneish (1997; 1999) added a new dimension to the literature of earnings manipulation by developing the probit and logit models (mathematical models) which use a set of different ratios in addition to the accruals. The Beneish model uses variables such as day's sales in receivables, sales margins and asset quality to identify companies that are involved in earnings manipulation. By incorporating such variables, this model is able to provide a broader view of earnings quality (Pustylnick, 2011). The incident of the Enron scandal in 2001 has led to a higher appreciation of the Beneish model. Scholars such as Hariri and Brewer (2004) and Simpson (2016) noted that the use of Beneish model could have signalled Enron as a problematic company, earlier. In a similar vein, some scholars

(Beneish, Lee, & Nichols, 2013; Kighir, Omar, & Mohamed, 2014) argued that the Beneish (1999) model is a strong and important instrument which should be used by auditors to detect earnings manipulation. They highlighted the capability of the model in separating those accruals which are more likely to persist from those whose accruals are more likely to reverse. Thus, the current study uses the Beneish model as an EMD tool. The Beneish model is then compared to the traditional audit method. In comparing these two models, a machine learning technique will be added as a new layer.

2.3 The Beneish Model: An Application Review

Various literature have discussed tools that are utilised in the framework of earnings manipulation detection. For instance, Bell and Carcello (2000) have applied logistic regression in their study and found that a simple logistic regression model outperforms the manual audit process performed by auditors. Some researchers (Nigrini & Mittermaier, 1997; Durtschi, Hillison, & Pacini, 2004; Cleary & Thibodeau, 2005) applied Benford's Law which detects deviations through comparing the frequency of digits in a given data set between actual and expected manipulations. A few scholars (Green & Choi, 1997; Lin, Hwang, & Becker, 2003; Koskivaara 2004) utilised the ANNs model as a tool to compare between actual and expected account balances. In more recent studies, however, researchers (Pustylnick, 2011; Mahama, 2015) associated bankruptcy with earnings manipulation. They indicated that the Altman Z-score could be used as a tool to flag financial statements manipulation. Similarly, Hogan et al. (2008) provided a systematic review of earnings manipulation detection. They reported that tools such as regression analysis, non-financial information, digital analysis and ANNs' models are heavily utilised to support the continuous audit processes.

DÁmico and Mafrolla (2013) focused on four models namely, Jones's model (1991), Modified Jones Model (MJM) (1995), Kothari, Leone, and Wasley's model (2005) and Beneish's model (1991) to detect fraudulent and non-fraudulent financial statements of 40 Italian public listed companies over the period of 1990-2009. Of these four models, they found that only Beneish model is statistically significant in predicting earnings manipulation. Similarly, Miller (2009) used the Binary Logistic Regression to test for Miller Ratio and MJM power to detect earnings manipulation. He found that both models fail to predict

the EMD at a statistically acceptable level of confidence. These findings indicate that the Beneish model has a strong potential in detecting earnings manipulation over other models.

Notwithstanding these empirical works, prior studies that incorporated the Beneish model solely have provided supporting evidence of the model's potential. For instance, Nwoye et al. (2013) used the Two-Way ANOVA test with the Beneish model to detect earnings manipulation of manufacturing companies in Nigeria over the years of 2002-2006. The study extracted data of audited annual reports and they found that the Beneish model contributes substantially to the detection of fraudulent financial statements. Within the context of developed countries, Paolone and Magazzino (2014) tested the probability of earnings manipulation of 1809 leading Italian companies in various industries (textile, food, clothing, automotive and metallurgic) between the period of 2005-2012. They found that the Beneish model is able to detect 929 companies out of the total of 1809 to have a high probability for earnings manipulation, with the clothing industry appearing to attain the highest percentage of manipulation (66.9 per cent).

On the Malaysian front, Kamal, Saleh, and Ahmad (2014) assessed the reliability of the Beneish model in detecting earnings manipulation of Malaysian public listed companies. They used a sample of 17 public listed companies which had been charged by the Securities Commission Malaysia for misstating their financial statements, from the years 1996-2014. Of the 17 listed companies, 14 committed earnings manipulation. Their study concludes that the Beneish model is a reliable tool for detecting firms that are engaged in fraudulent reporting and misstatements. In another study conducted by Omar et al. (2014), the Beneish model and ratio analysis were applied to observe four financial indicators namely Profitability, Operating Efficiency, Liquidity and Coverage and Funding Structure which were utilised as tools to detect financial fraud committed by Megan Media Holdings Berhad (MMHB). Their study suggests that auditors could use the Beneish model to perform audits for a reasonable assurance that financial statements are free from material misstatements.

2.4 Supervised Classification – Machine Learning Techniques for EMD

Previous literature provides an extensive review about the supervised classification techniques of machine learning (Kotsiantis, 2007; Anto & Chandramathi, 2011) with machine learning being established as a

subset of Artificial Intelligence (AI). According to Russell and Norvig (1995), AI could be defined as a system of thinking rationally, like human beings. They also argued that AI could be viewed from two different perspectives. First, from the human centered view, AI needs to be approached using empirical science and experimental confirmations. Second, from the rationalist view, AI needs to involve a combination of mathematics and engineering. The AI term, however, was then changed to business intelligence (BI) in the domain of business applications. Here, BI represents a wide area of applications and technologies for collecting, storing, analysing and providing access to information, with the aim of improving business process modelling quality (Nedelcu, 2013). Thus, a great importance was emphasised in recent years, as a result of incorporating probability theory into reasoning and machine learning of AI models (Braz, Amir, & Roth, 2008).

Similarly, machine learning involves many applications or algorithms to ensure that computers learn to behave more intelligently by generalising rather than just storing data. Data mining (DM) appears to be the most significant domain of machine learning (Kotsiantis, 2007). Despite the different definitions established for DM and for the supervised classification of machine learning, the concept is the same. In this concern, Turban, Aronson, and Liang (2005) defined DM as a semi-automatic process that uses statistical techniques, mathematics, AI, and machine learning to extract and identify potential knowledge and useful information that is stored in a large database.

Kotsiantis (2007) illustrated that every case in any data set used by machine learning algorithms is represented by a set of variables called features which could be in the form of continuous, categorical data or binary data. Any case in the data set that is given labels (categories) is classified as supervised. This feature makes supervised machine learning different from unsupervised machine learning where under the latter category, the cases will be unlabelled. Thus, supervised machine learning could be referred to as the process of learning a set of rules from cases (examples in a training set) or more generally speaking, creating a classifier that can be used to classify and predict fraud when incorporating new cases.

Applying the machine learning approach can improve the efficiency of the system and the design of machines. According to Kotsiantis (2007), the process of applying supervised machine learning to a real-world problem includes various steps such as problem definition, data identification, data pre-processing, definition of training set, selection

of algorithm, training parameters, and evaluation with a test set. In this concern, Kotsiantis (2007) and Anto and Chandramathi (2011) reviewed different supervised machine learning techniques such as Logic Base Algorithms (Decision Trees, Rule Based Algorithms), Neural Network Based Machine Learning (Single layered ANNs, Multi-layered ANNs), Statistical Based Machine Learning (BBNs and BNCs), Instance-based Learning, and Support Vector Machines for classification. Following this train of thought, some scholars (Zhang & Zhou, 2004; Dikmen & Kukkocaoglu, 2010; Sharma & Panigrahi, 2012; Patil & Sherekar, 2013) suggested that these supervised machine learning techniques could be used as tools to assess the risk of fraud and manipulations in financial statements. Expanding on this, Gupta and Gill (2012) developed a data mining framework for financial statement data assessment to prevent and detect fraud.

Despite the various types of algorithm suggested and developed for the DM approach as noted in the literature, most scholars (e.g., Diware, Borhade, & Ringe, 2016) argued that no single learning algorithm can uniformly outperform other algorithms over all data sets. Yet, decisions on the type of algorithm to be used should be made based on the type of the classification problem and the accuracy of the applied algorithm that fits best. Likewise, many researchers (e.g., Kotsiantis, 2007) also hold the view that the most suitable supervised machine learning technique for a classification problem is the BBNs and the BNCs. In this regard, a BBN is defined as a graphical model for probability relationships among a set of variables called features (Kotsiantis, 2007). Accordingly, Saad et al. (2012) and Saad, Zaarour, and Zeinedine (2013) indicated that the BBNs are powerful tools for knowledge representation and inference, under conditions of uncertainty. Other researchers (Kirkos, Spathis, & Manolopoulus, 2007; Zaarour et al., 2015) noted that the BBNs are the most well-known representative of statistical learning algorithms which have high classification accuracy, layered architecture with hidden variable hierarchy and a probabilistic inference. In contrast, the BNC is the simplest structure of the BBNs (Kotsiantis, 2007) and it has been noted as an effective classification tool which is easy to interpret and is widely used in banking and financial/claim fraud detection (Elkan, 2001a). Moreover, Zhang (2004) found that the BNC is one of the most efficient and effective inductive learning algorithms for machine learning and data mining. His results indicated that the cancelling effect of dependencies will not influence classification. In

this case, the BNC appears to be the optimal classifier. Consistently, Domingos and Pazzani's (1997) study, cited in Kotsiantis (2007), showed how some studies found the BNC to be a superior tool as compared to other learning schemes. Its effectiveness can be seen even on data sets which have substantial feature dependencies. The BNC tool has a short computational time for training in addition to other significant consequent computational advantages.

Given the argument, this study therefore, uses the BNC which is composed of Directed Acyclic Graph (DAG) with only one parent (representing the unobserved node) and several children (corresponding to observed nodes) and it has a strong assumption of independence among child nodes in the context of the parent. In comparison, the BBN is a graphical model for noting probability relationships among a set of variables. Learning through a BNC is divided into learning of the DAG structure of the network, and learning of the determination of its parameters through conditional probability tables (CPT) (Kotsiantis, 2007). The main reason for using the BNC on a mathematical model such as the Beneish model is its great features. For instance, the BNC takes into account prior information about a given problem in terms of structural relationships among variables. The BNC is suitable for data sets that have small features and it can achieve maximum prediction accuracy. Furthermore, the learning of the BNC can be trained very quickly and it can be used easily as an incremental learner (Diware et al., 2016). The BNC is very transparent and easily applied even by users like physicians who may find that probabilistic explanations replicate their way of diagnosis (Kotsiantis, 2007). In addition, the tool is also naturally robust for identifying missing values, an important matter, since missing values are simply ignored in computing probabilities. Therefore, they have no impact on the final decision.

Thus, the small number of features or variables in an EMD problem supplied by the financial ratios of the Beneish model as attributes for the learning process, is suitable for the BNC. This study applied the Supervised Classification when using the BNC for detecting earnings manipulation on financial statements.

3. Methodology

In this study, supervised machine learning was applied to a real life problem namely, earnings manipulation of financial statements. In

this approach, different variables including financial ratios of a widely applied mathematical model for earnings manipulation detection namely the Beneish model were utilised. These variables represent the parameters of the BNC. Following Kotsiantis' (2007) study, several steps including problem definition, data identification, data pre-processing, definition of a training set, selection of algorithm, training parameters, and evaluation with a test set were employed in the process of applying the supervised machine learning to a real-world problem. These steps are summarised in Table 1 and explained as follows.

3.1 Stage 1 – Problem Definition

The problem stems from different aspects that appeared over time as prescribed in the introduction and literature review sections. One crucial aspect revealed by the high number of financial frauds detected is the increase in the financial fraudulent behaviour which can cost a country's economy billions of dollars annually (Kamal et al., 2014). Another aspect of the problem is the existence of traditional audit methods like manual verification of accounts and old sampling techniques which are no longer adequate in view of today's high technology and the continuous integration between technological advancements and businesses. Since auditors may have limited analytic capabilities, BI and machine learning approaches can help to strengthen and maximise the power of auditors in the framework of EMD as a major component of the analytical audit procedure. This will allow auditors to overcome information ambiguity and make accurate decisions. In this regard, a mathematical model widely applied for EMD was employed. This is then compared to the Traditional Auditors' Methods through the application of a supervised machine technique, namely the BNC, by using a computer software tool.

3.2 Stage 2 – Period Selection

In this study, data were extracted from the financial statements for four consecutive years starting from 2006 to 2009. They were extracted from the database of a Lebanese governmental financial audit institution using its audit software. The years of 2006 to 2009 represented the only available closed financial period of the historical financial statements where the audit institution did the audit.

3.3 Stage 3 – Data Collection

Since one of the risk factors that could elevate the chances of committing financial statement fraud is the nature of the industry (Hogan et al., 2008), the Lebanese liquid fuel industry was selected. This sector has the largest number of firms which have undergone the audit process based on the audit institution's database. In this regard, the industry represents the ideal setting for earnings manipulation detection research. The samples of this study comprise 53 large companies in the wholesale liquid fuel industry, with each company representing a separate case. All companies have the approximate business size, i.e. a turnover of more than ten billion Lebanese Pounds per year (around 7 million US Dollars) with the same legal status of "*Société Anonyme Libanaise*" (S.A.L) or Joint-Stock companies.

The data used for analysis in this study consist of balance sheets, income statements and cash flow statements of the 53 samples. The extracted data were not publicly available. The researchers were authorised by the Lebanese Ministry of Finance to use the data for academic/research purposes.

3.4 Stage 4 – Data Preparation

This stage consists of data transformation, cleaning and filtering. During this preparation stage, null attributes were not removed because they represent the missing values. Subsequent to the data cleaning and filtering process, a total of 19 informative attributes were gathered. These attributes comprise: Tangible Fixed Assets; Total Receivables; Total Current Assets; Total Debt; Total Assets; Net Income; Total Long Term Debts; Total Current Liabilities; Working Capital; Total Sales; Total Cost of Goods Sold; Gross Profit; Sales General & Administrative Expenses; Total Depreciation; Total Expenses; Profit or Loss From Operations; Non-Operating Income; Results Before Income Tax; and Income Tax Payable.

3.5 Stage 5 – Data Selection

As shown in Table 2, the attributes of Total Receivables and Total Sales were used to calculate the ratio of Days' Sales in Receivables Index (DSRI). The attributes of Total Sales and Total Cost of Goods Sold were used for Growth Margin Index (GMI). The attributes of Total Current Assets and Total Assets were used for Assets Quality Index (AQI). The attributes of Total Sales were used for Sales Growth Index (SGI). The

Table 1: Approach Stages

Stage No.	Stage Name	Stage General Features	Stage Attributes
1	Problem definition	Financial misrepresentation Tax evasion Financial fraud detection	Earnings management Earnings manipulation Financial data assessment
2	Period selection	Time management – yearly figures Audited financial statements Closed accounting/audit results	2006-2007-2008-2009
3	Data collection	Financial statements Quantitative data Single source	Balance sheets Income statements Cash flow statements
4	Data preparation	Transformation Cleaning Filtering	Oracle-based system Excel spreadsheets Records matching and deletion 1400 companies/177 attributes per year/all industries 177 attributes per year – filtered till became necessary 19 input variables
5	Data Selection	Sampling/attributes Companies' type Industry identification	53 companies Joint-stock legal nature Liquid fuel industry
6	Data partition	Training-data set Test-data set Expert data-set (traditional audit-system results)	2006-2007 training data-set 2008-2009 test data-set 2008-2009 expert data-set

Table 1: (continued)

Stage No.	Stage Name	Stage General Features	Stage Attributes
7	Earnings manipulation detection	Beneish model Traditional audit framework	M-Scores attributes 8 final M-Scores ratios/approach variables Manual auditors' results
8	Machine learning technique selection	Supervised classification	BNC technique [class definition & labelling; manipulator-m; non manipulator-nm]
9	Learning	2006-2007 training data-set [53 companies including those with missing values]	DM tool Original training 2006-2007 data-set cases Case simulation with missing values /estimation maximization (EM) algorithm
10	Testing, Analysis, and Evaluation	2008-2009 test data-set (Based on Beneish model) 2008-2009 expert-test data-set (Based on the traditional audit-system results) [38 companies data, excluding missing class values in the test set]	Confusion matrix Classification rates/error rates Performance matrix/sensitivity

Table 2: The Beneish Model Ratios

No.	Ratio Symbol	Ratio Name	Formula	Rationale (Beneish et al., 2013)
1	DSRI	Days' Sales in Receivables Index	$(\text{Accounts Receivables}_t / \text{Sales}_t) / (\text{Accounts Receivables}_{t-1} / \text{Sales}_{t-1})$	Shows distortions in receivables that can result from revenue inflation
2	GMI	Gross Margin Index	$((\text{Sales}_{t-1} - \text{Cost of Sales}_{t-1}) / \text{Sales}_{t-1}) / ((\text{Sales}_t - \text{Cost of Sales}_t) / \text{Sales}_t)$	Deteriorates margins that predispose firms to manipulate earnings
3	AQI	Asset Quality Index	$(1 - (\text{Current Assets}_t + \text{PPE}_t) / \text{Total Assets}_t) / (1 - (\text{Current Assets}_{t-1} + \text{PPE}_{t-1}) / \text{Total Assets}_{t-1})$	Captures distortions in other assets that can result from excessive expenditure capitalisation
4	SGI	Sales Growth Index	$(\text{Sales}_t) / (\text{Sales}_{t-1})$	Manages the perception of continuing growth and capital needs predispose growth firms to manipulate sales and earnings
5	DEPI	Depreciation Index	$(\text{DE}_{t-1} / (\text{DE}_{t-1} + \text{PPE}_{t-1})) / (\text{DE}_t / (\text{DE}_t + \text{PPE}_t))$	Captures declining depreciation rates as a form of earnings manipulation
6	SGAI	Sales, General & Administrative expenses Index	$(\text{SGA}_t / \text{Sales}_t) / (\text{SGA}_{t-1} / \text{Sales}_{t-1})$	Captures decreasing administrative and marketing efficiency through larger fixed SGA expenses that predisposes firms to manipulate earnings
7	LVGI	Leverage Index	$((\text{LTD}_t + \text{Current Liabilities}_t) / \text{Total Assets}_t) / ((\text{LTD}_{t-1} + \text{Current Liabilities}_{t-1}) / \text{Total Assets}_{t-1})$	Shapes earnings manipulation when increasing leverage tightens debt constraints and re-disposes firms to manipulate earnings
8	TATA	Total accruals to total assets index	$((\text{WC}_t - \text{WC}_{t-1}) - (\text{Cash}_t - \text{Cash}_{t-1}) + (\text{ITP}_t - \text{ITP}_{t-1}) + (\text{Current Portion of LTD}_t - \text{Current Portion of LTD}_{t-1}) - \text{DE}_t) / (\text{Total Assets}_t)$	Captures manipulations where accounting profits are not supported by cash profits

Note: Where PPE = Plant, Property and Equipment; DE = Depreciation and Amortisation Expense; SGA = Sales, General and Administrative Expenses; LTD = Long Term Debt; WC = Working Capital; ITP = Income Tax Payable; t = current year; $t-1$ = previous year;
M-Score = $-4.84 + 0.920\ \text{DSRI} + 0.528\ \text{GMI} + 0.404\ \text{AQI} + 0.892\ \text{SGI} + 0.115\ \text{DEPI} - 0.172\ \text{SGAI} + 4.679\ \text{TATA} - 0.327\ \text{LVGI}$

attributes of Total Depreciation and Tangible Fixed Assets (PPE) were used for Depreciation Index (DEPI). The attributes of Sales General & Administrative Expenses and Total Sales were used for Sales, General & Administrative Expenses Index (SGAI). The attributes of Total Long Term Debts, Total Current Liabilities, Total Assets, and Total Current Liabilities were used for Leverage Index (LVGI). The attributes of Working Capital, Total Long Term Debts, Cash, Total Assets and Income Tax Payable were used for Total Accruals to Total Assets Index (TATA).

In this stage, all the 19 attributes, hereby termed as input variables, were used to derive the M-Scores for the four consecutive years. Table 2 also depicts the formula of the different ratios and M-Scores applied. The eight financial ratios which were calculated then formed the independent variables of this study. They consist of DSRI, GMI, AQI, SGI, DEPI, SGAI, LVGI, and TATA.

In this study, 53 times eight dimensional table values for each year's data set were generated based on the financial statements of the 53 companies. The M-Score calculations were based on time *t* and *t*-1 for the years 2006-2007 and 2008-2009. The time series of 2007-2008 however, was dropped so that more knowledge can be derived from the training data as well as to avoid interdependence between the training data set and the test data set.

3.6 Stage 6 – Data Partition

In the context of this study, a partition was created between two data sets where data extracted from the financial statements of 2006-2007 were categorised as training data set and data extracted from the financial statements of 2008-2009 were used as test data set. The process of calculating the M-Scores for the two data sets helped to prepare the data for supervised calculations and evaluation using the BNC. This process was conducted through the application of a computer software tool, namely Netica. (Another 2008-2009 test data-set was used, based on the manual traditional audit-system results as described through the following stage).

3.7 Stage 7 – Earnings Manipulation Detection

3.7.1 The Beneish Model (M-Score Data)

The Beneish model was introduced by Messod Beneish in 1997 and it was then updated in 1999 by adopting a weighted exogenous sample maximum likelihood (WESML) probity model for the purpose of detect-

ing manipulations. It is also known as M-Score model. The M-Score model is a mathematical model that uses eight financial ratios to identify whether a company has manipulated its earnings (Beneish, 1997; 1999). This model is similar to Altman's Z-Score but it is more directed towards detecting possible earnings manipulation by companies rather than the companies' solvency status.

According to the Beneish model, a score (M-Score) greater than -2.22 indicates strong likelihood of a company being a manipulator (Beneish, 1997; 1999).

The functions and calculation methods of the eight independent variables (financial ratios) which have been determined for this study are represented in Table 2. These variables were derived from the financial statements and used to calculate a score which can detect earnings manipulation by using the formula of the M-Score as follows: M-Score = -4.84 + 0.920 DSRI + 0.528 GMI + 0.404 AQI + 0.892 SGI + 0.115 DEPI – 0.172 SGAI + 4.679 TATA – 0.327 LVGI (Beneish, 1997; 1999). Therefore, in order to identify earnings manipulation, M-Scores were calculated for each company by deriving the financial ratios of the Beneish model. These values assist in classifying a company into a manipulator or a non-manipulator. In the years of 2006-2007, the Beneish model was able to classify 18 cases of manipulators, 18 cases of non-manipulators and 17 missing cases due to incomplete data. In the years of 2008-2009, the Beneish model was able to classify six companies as manipulators of their earnings, 32 non-manipulators, and 15 missing cases.

3.7.2 The Traditional Audit Behaviour (Manual Auditors' Methods)

In looking at this method, the audit results acquired from the audit institution of all cases in the sample were applied. These results represent the audit system's outcome which was entered by the audit institutions each time an audit case file is closed. This method summarised all the manual methods used by an auditor to measure the manipulation in a given case. Restatement of earnings were mainly used to classify a case that is likely to be a manipulator of earnings. Traditional audit methods are perceived to be traditional because they allow auditors to reach a reasonable conclusion based on a limited sample of a case's parameters. This method relies on manual sampling and paper audit analysis such as basic ratio analysis, year-to-year financial data comparison, and tax performance comparisons rather than advanced mathematical models and simulations which are represented by CAATs, CA, GAS, and BI.

The classifications of the 53 cases extracted for this study were based on the results of each audit case. It is assumed that when the audit institution restated the accounts of a company's earnings figures, then this company is likely to be considered as a manipulator.

In the case of this study, there was no key performance indicator for the wholesale liquid fuel industry and firm type which could help an auditor or financial analyst expert to use as a benchmark to classify a company as being a likely manipulator. Thus, the auditor would not be able to classify whether a company had or had not manipulated its earnings even after running a basic ratio analysis. Moreover, there was no access to public information disclosing companies which had violated accounting standards or was sanctioned for manipulations. Thus, the EMD could only be measured by using the results extracted from the Audit Case Selection Tabs of the Audit System which were used during the audit programme of the audit institution for the period 2006-2009 for each of the 53 companies. Following the auditors' amendments of the financial (earnings) results, after the audit assignment, an auditor is obliged to issue and archive the system on an adjustable table. This is to illustrate that the original reported financial statement is under change or amendment. This process reveals a material restatement in the company's accounts. In other words, the company is obliged to modify its financial statement as the audit reveals that there is a mis-interpretation in its original financial statement. Thus, the company is considered to have manipulated its earnings.

Accordingly, the audit results of the 53 companies were extracted. Based on this, perception of the company's previous performance (prior information about the classification problem needed for BNC) and their manipulation existence were scrutinised. Analysis of the manual procedure revealed the identity of the actual manipulator. In this regard, a total of 18 companies had manipulated their accounts with 20 being non-manipulators and 15 missing cases. This occurred after the missing cases were excluded from this test set so as to match the exact cases in the test set of the Beneish model (the same 38 cases were tested for more precision in the classification).

3.8 Stage 8 – Machine Learning Technique Selection

As discussed in the literature, machine learning is based on the training information provided to the learning system of the environment (external trainer). Machine learning consists of learning a set of rules

from cases (examples in a training set) or creating a classifier that can be used to generalise new cases. In this concern, the correct machine learning technique algorithm has to be chosen for the supervised machine learning within the EMD framework. Consequently, the Bayesian Naïve Classifier (BNC) was selected. This selection was based on its incremental power in classifying real life problems within a framework like the EMD framework. In the context of this study, a new layer of machine learning using the BNC was introduced as a means to extend the procedure of the manual auditors' methods. This machine learning technique captures a modern approach which is similar to that previously presented and prescribed in Figure 1.

At the level of the machine learning layer, the BNC Network type was constructed for the purpose of distinguishing the manipulator companies from the non-manipulator companies. The complete data sets of 2006-2007 were then used as training data for constructing the BNC Network. At this level, no findings were introduced for the BNC Network as it was ready for supervised classification through learning and testing.

3.9 Stage 9 – Learning

Learning is a very useful property of the BNC. It occurs when an expert builds the BNC Network that has been refined by learning from the data. Though vast amounts of data should make learning from a data set more possible and efficient, this is usually not the case (Heckerman, 1998). In the current study, the 2006-2007 training data sets were used for learning the BNC Network of the 53 cases. Expectation Maximization (EM) algorithm, as shown in Figure 2, was utilised using a computer software tool, namely Netica. The characteristics of the BNC Network which are susceptible to missing values allowed these data to be included in learning the network. The EM algorithm makes it possible to overcome this problem of missing values (parameters) in the BNC Network. This algorithm allows parameter estimation to be noted in probabilistic models even with incomplete data. The algorithm includes a process which alternates between the step of guessing a probability distribution over completions of missing data, given the current model known as the E-step and the re-estimation of the model parameters using these completions which is known as the M-step (Do & Batzoglou, 2008). The attributes used in the learning process encompass the eight financial ratios of the Beneish model (DSRI, GMI, AQI, SGI, DEPI, SGAI, LVGI and TATA).
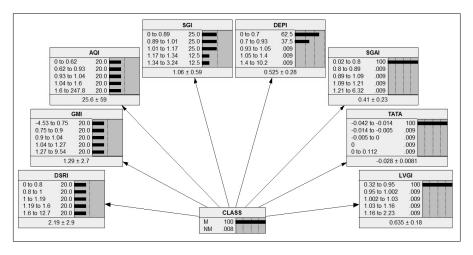
Figure 2: Learned-training Data Set Using EM Algorithm – The BNC Network

The BNC Network shown in Figure 2 reflects the probabilistic relationships between class variable (earnings manipulation), with manipulator and non-manipulator as the parameters; and feature variables (DEPI, SGAI, TATA, LVGI, DSRI, GMI, AQI and SGI). These feature variables served to describe the characteristics of a company to be classified as manipulator or non-manipulator. In the BNC Network, the arrows (also called links) between any two nodes as shown in Figure 2 indicate that there are probability relationships which are known to exist between the states of those two nodes. However, the direction of the link arrows roughly corresponds to "causality". Thus, the BNC can easily be extended to computing utilities, given the degree of knowledge that is available in a situation. This appears to be the reason why it is popular in business and financial decision making as much as in scientific and economic modelling. As shown in the sample findings of 2 and 3 in Figure 3, changing the possible parameters of SGAI, TATA, and LVGI reveals the chances of companies manipulating earnings. This indicates that in learning the structure of the BNC Network, the algorithm chooses SGAI, TATA and LVGI as the feature attributes that are relevant to predict earnings manipulation.

However, Figure 4 shows that by adding positive findings such as LVGI to DEPI, SGAI and TATA, the state of earnings manipulation decreases significantly. These findings are interesting, because they reveal that the ratios of the Beneish model, except for GMI, AQI, and
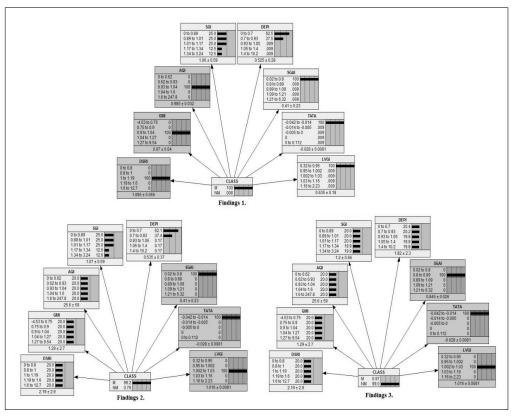
Figure 3: Sample Findings of AQI, GMI,DSRI, SGAI, TATA and LVGI – The BNC Network

DSRI, have significant relationships with fraudulent financial state-ments for the wholesale liquid fuel industry. For example, sample findings 3 of Figure 3 represent the probability of manipulation Class (C), given some values of the Beneish model's variables such as SGAI, TATA, and LVGI. Mathematically this is represented as *P(C| SGAI, TATA, LVGI)*. Nonetheless, according to findings 3 in Figure 3, the mathematical formula is represented as *P(C| SGAI=between 0.8 & 0.89; TATA=between -0.042 & -0.014; LVGI=between 1.002 & 1.03)*.

As can be seen in Figure 3, these findings demonstrate that the BNC Network predicted 99 per cent of the non-manipulator existence. How-ever, in sample findings 2 of Figure 3, the prediction of the manipulation existence is 99.2 per cent. But, as can be seen in the sample findings 3 of this figure, when giving the same values for TATA and LVGI as in
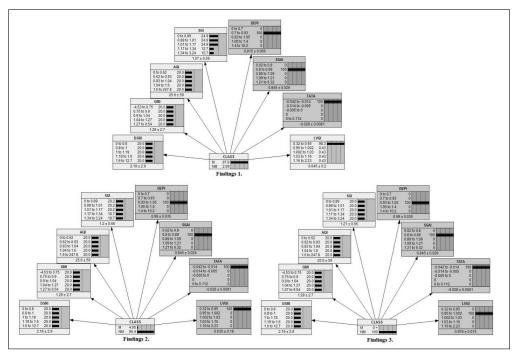
Figure 4: Sample Findings of DEPI, SGAI, TATA and LVGI – The BNC Network

findings 2 but with the exception of setting a different value for SGAI (by raising it to set in the interval of 0.02 and 0.8), the prediction of the manipulation existence is slightly decreased to 97 per cent indicating that SGAI affects manipulation under specific values. In this regard, an auditor may focus on the analytical procedure so as to direct the audit towards some indicators for the EMD. In a recent study (Ahmed & Naima, 2016), but under a different context from machine learning, using statistical independent t-test, TATA, SGAI, AQI and DSRI are found to be the ratios that have mainly caused probable manipulations. Thus, the BNC model seems like a good model that can be used to match this study's purpose.

3.10   Stage 10 – Testing, Analysis, and Evaluation

3.10.1 Beneish Model

After learning the BNC Network from the 2006-2007 training data set by using the EM algorithm, the 2008-2009 test data set acquired based

on M-Scores were used to test the learnt classifier so as to evaluate the classification results.

3.10.2 Manual Auditors' Methods

At this level, the detection test for accuracy based on the BNC was used to learn from the training data set. This data set was derived from the M-Scores. The test data set was derived from the manual auditors' results for the audited years covering 2008 and 2009. The analysis and evaluation of the results are presented in the following section.

In summary, the methodological process involved in this study comprises: (1) Extracting data from financial statements supplied by the audit programme of the audit institution. Following the process of filtering and cleaning data, 19 informative attributes were identified. (2) Using the 19 informative attributes to identify manipulation of financial statements, M-Scores were calculated for each company through the eight financial ratios gained from the Beneish model formula. (3) The Bayesian Naïve Classifier (BNC) was applied to build a model for the training and testing data. (4) Data were grouped into 2006-2007 training data set and two 2008-2009 test data sets. (5) After learning the training set, classification of the BNC network was tested over the M-Score test set and then over the manual auditors methods' results. (6) Confusion and performance matrices (sensitivity and specificity tests) were validated.

4. Findings and Discussion

This study tested and compared the performance of the BNC which was applied in the Beneish model and manual auditors' method. The performance of a test can be characterised in terms of its sensitivity, specificity and threshold; and decision-making is strongly affected by the interpretation of the test results (Bradford, Kunz, Kohavi, Brunk, & Brodley, 1998).

The main goal of machine learning is to devise a learning algorithm that could figure out how to perform tasks by generalising information from examples. The performance measurement of a machine learning algorithm is based on its accuracy of data set classifications (Bradford et al., 1998). However, some types of misclassifications may be worse than others, hence, a cost matrix (sometimes called a performance or confusion matrix) can be used to represent the differing cost of each type of misclassification. Each row in this matrix is used to represent

the predicted label and each column corresponds to the actual labelling (Elkan, 2001b; Nguyen, Doncescu, & Siegel, 2016; Moattar & Fadaei Noghani, 2016). Machine learning classifiers are built to maximise classification accuracy. To incorporate cost sensitivity into learning, adjustments must be made to the learning process. However, the adjustment needs to consider the cost of misclassification (Margineantu, 2002). One method of incorporating cost sensitivity into learning is to use a classifier such as the BNC which can provide class probability estimates and also compute the expected cost of each label.

The BNC employed in this study takes into account the unmodified data set and it also counts all the attributes from every instance as normal. Yet, rather than predicting the label with the highest probability, the label with the lowest cost is selected. This is done by getting all label probabilities and determining the expected cost of each possible labelling. The BNC is based on the Bayes theorem and it assumes class conditional independence whereby the effect of an attribute value on a given class is independent of the values of the other attributes. This is to simplify the computations involved. In fact, this is the reason why the BNC is coined as "naïve". The tool can be trained very efficiently in a supervised machine learning setting (Diware et al., 2016). It often works much better in complex real-world situations and it has the advantage of requiring only a small amount of training data to estimate the parameters (means and variances of the variables) which are necessary for classification. Because independent variables are assumed, only the variances of the variables for each class, rather than the entire covariance matrix, need to be determined (Doreswamy & Nagaraju, 2010).

Hence, the complete possible states that resulted from testing for manipulation existence (M) in all cases (i.e. either a manipulated company or non-manipulated company) are presented by the confusion matrices in Tables 3 and 4 respectively (Actual Versus Predicted Statuses are presented in the tables where Manipulator status is denoted by label M, and Non-Manipulator status is denoted by label NM). Two test data sets were used for testing. One test data set was based on M-Score classifications and one test data set was based on the manual auditors' methods' results. In these sets, the M-Score predicted six manipulated cases, 32 non-manipulated cases and 15 missing cases, due to incomplete data. However, the manual auditors' methods' results predicted 18 cases of manipulators and 20 cases of non-manipulators. Fifteen cases (which were missing in the test set of the Beneish model) were excluded to match the same cases of the testing set, for more precision. Thus, the

Table 3: Confusion Matrices-Naive Bayes Classifier:
Beneish M-Score

	Predicted		Actual
	M	NM	Label
M-Score	2	4	M
	1	31	NM

Note: Error rate 13.16%.

same companies were tested with the Beneish model and the manual auditors' methods. In this regard, each confusion matrix covers the total number of 38 cases in the 2008-2009 test data sets excluding the missing cases (class labels). These results, as prescribed, are based on the application of a computer software tool.

As shown in Table 3 above, the BNC can correctly classify two out of six actual manipulated companies and 31 out of 32 actual non-manipulated companies with an error rate of 13.16 per cent. The result indicates that in every 13.16 per cent of the cases for which a company committed an earnings manipulation, the network predicted the mis-classification. Further, this finding also demonstrates that the BNC has a classification rate of 86.84 per cent for revealing the true classification of a company as manipulating its earnings. This was based on the financial data of the M-Score. However, based on the assumption that restatement is only due to manipulation, this result was compared to the manual auditors' methods (actual manipulation) in EMD which show a 39.47 per cent error rate (misclassification rate) as shown in Table 4 below.

In this regard, the error classification rate of earnings manipulation when using the M-Score is three times less than the error produced when using the traditional approach. This new finding is promising

Table 4: Confusion Matrices-Naive Bayes Classifier:
Manual Auditors' Methods

	Predicted		Actual
	M	NM	Label
Manual Auditor's	3	15	M
Methods	0	20	NM

Note: Error rate 39.47%.

for accountants, auditors, financial experts and other professionals who evaluate through the manual auditors' methods or other scientific models. When an auditor runs an audit assignment, the auditor would usually use the basic ratio analysis, audit checklists and questionnaires to identify the possibility of accounting manipulations. Thus, by applying the BNC complemented with mathematical models in the EMD framework, the auditing process is strengthened.

The performance matrix for the test of manipulation state (M) which is the interest of quality test was used as a measurement of the performance evaluation in the two testing scenarios (M-Score and Manual Auditors Methods). Sensitivity (type I error) and specificity (type II error) of the BNC were used to evaluate the classification under M-Score versus manual audit methods. This is illustrated in Table 5.

Table 5: Performance Matrix: Beneish M-Score vs Manual Auditors' Methods

	Sensitivity (%)	Specificity (%)
M-Score	33.33	96.88
Manual Auditor's Methods	16.67	100

From this study, it appears that sensitivity is the probability that will positively give a case with the condition to be a manipulator. It indicates how the test will be true positive (M) in the setting of the EMD (i.e. 33.33 per cent) for M-Score and manual auditors' methods (16.67 per cent). Specificity indicates how a test will be true negative (NM) in companies without having to manipulate their earnings (i.e. 96.88 per cent) for the M-Score and for specified audit methods (100 per cent).

Since the focus of this study is EMD, more indication towards sensitivity which shows more prediction of true manipulation in validating financial data of the M-Score model over the manual auditors' methods is noted. After constructing the BNC using the financial ratios of the Beneish model and after learning the BNC Network, the BNC classifies the cases as manipulated cases (six such cases were correctly predicted i.e. 33.33 per cent) using M-Score financial data. This shows the best sensitivity of the BNC.

In terms of classification accuracy (86.84 per cent), the outcome of this study precedes the findings of Herawati (2015) and Kara, Ugurlu, and Korpi (2015). It also conforms to the results of Findik and Ozturk

(2016) in approximate classification by the level of 90 per cent. The results noted from this study support the capability of the Beneish model for EMD.

Therefore, the introduction of a machine learning layer using the BNC under the Beneish model can result in a higher classification accuracy of the data than when using the manual auditors' methods. This outcome supports the methodology approach presented in Figure 1 which shows how machine-learning techniques could assist auditors prior to making decisions about financial statements, particularly in the framework of EMD. Adding a new layer of machine learning technique, after using mathematical models such as the Beneish model and comparing this with manual audit methods, will strongly enhance the decision making within the framework of EMD.

5.   Conclusions

The EMD issue has raised many concerns in previous research with variation noted in solutions and limitations. Recently, a new approach in EMD such as supervised machine learning has been introduced. The tool has shown its capability in improving the efficiency of systems and the designs of machines especially in the classification process.

In this study, the supervised machine learning technique was used as a classification tool to evaluate and compare the advanced mathematical model, "the Beneish model" with the "Typical Manual Auditors' Methods". This study was conducted using audited financial data over four consecutive years from 2006 to 2009. Unlike prior studies, the BNC was introduced for the purpose of evaluating both the Beneish model and the auditors' manual methods. Previous research has shown that the BNC is an optimal classifier and it best fits the EMD classification problem highlighted. The results drawn from this study reveal that the M-Score generated through the Beneish model produced better sensitivity as compared to the Manual Auditors' Methods in assessing financial data for the purpose of detecting earnings manipulation. The BNC had correctly classified 86.84 per cent of the manipulated companies under the M-Score assessment while under the manual auditors' methods it had accurately categorised only 60.53 per cent.

Since prior studies have identified manipulating companies from the list of enforcement actions for fraud and restatement, the identification of fraud cases by authorities and governments appear to have been less effective. Furthermore, the Tax Compliance Departments or

the Certified Auditing Agencies may have selection biases in some fraud cases thus, detection is low. Moreover, the criteria used to assess and evaluate EMD are diverse, making the process less efficient. Thus, using the Beneish model would help to overcome this issue.

The aim of this study is to enhance the processes of earnings manipulation detection by creating a preliminary tool which, in this case, is based on machine learning, to assess any typical manual audit behaviour against any advanced mathematical, statistical, or scientific model. The study also aims to demonstrate to what extent the BNC can be applied on advanced mathematical model such as the Beneish model to reshape modern auditing framework.

From the study conducted, it can be concluded that traditional audit methods, which are based on simple methods of basic ratio analysis, sampling, vouching of every transaction, manual verification of accounts, and non-technological incorporation, need to be integrated with technological advancements so as to satisfy the needs of the new economy. The identification of material misstatements of financial statements is a critical step in the audit field today because fraud identification and prevention is important. However, fraud identification and prevention is impeded by the complexity of the transactions involved and the growth of global economies in the new era of technology. Therefore, modern audit methods complemented with supervised classification and machine learning can assist the auditing analytical procedures within the EMD context. Incorporating advanced mathematical models such as the Beneish model can provide better results. The traditional method is disadvantaged by the limited analytical capabilities of auditors in analysing big financial data and their limited testing abilities for all business cases. Nonetheless, the mathematical procedure using the Beneish model lacks the qualitative featured data of manual auditing. The Beneish model can build an advanced model through expectation maximisation or any algorithm. This will definitely strengthen the Beneish model through machine learning which can help to detect earnings manipulations, regardless of the qualitative audit factors. At this level, incorporating both the Beneish model and machine learning technique is a step forward for the auditing field.

This study contributes valuable knowledge for potential users of financial statements such as shareholders, financial analysts, auditors, accountants, government tax controllers, financial forensic investigators and academic researchers. It constitutes a preliminary study on machine-learning approach for auditing. It broadens the scope of using the

Beneish model for identifying earnings manipulators and in assisting the decision-making process. It can indicate manipulations under the application of machine learning tools. Furthermore, the comparison noted in the power of classification in EMD using a machine learning technique such as the BNC will encourage auditors or other related investigative parties to search for tools and techniques which can be embedded into their continuous auditing procedure. This can certainly enable them to have a better understanding of the attributes of earnings manipulation and machine learning in detecting fraud.

Despite the contributions of this study, the research is limited by the single industry analysis, single machine learning technique, single mathematical model, and the small sample size. These limitations may restrict the generalisation of the results to other industries. Therefore, for future contributions, an expansion of data in terms of other industries should be examined with the application of other manipulation detection models and machine learning techniques to test for the classification of earnings manipulation.

References

Ahmed, T., & Naima, J. (2016). Detection and analysis of probable earnings manipulation by firms in a developing country. *Asian Journal of Business and Accounting, 9*(1), 59-81.

Anto, S., & Chandramathi, S. (2011). Supervised machine learning approaches for medical data set classification: A review. *International Journal of Computer Science & Technology, 2*(4), 234-240.

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. doi: 10.2308/aud.2000.19.1.169

Beneish, M.D. (1997). Detecting GAAP violation: Implications for assessing earnings management among firms with extreme financial performance. *Journal of Accounting and Public Policy, 16*(3), 271-309. doi: 10.1016/s0278-4254(97)00023-9

Beneish, M.D. (1999). The detection of earnings manipulation. *Financial Analysts Journal, 55*(5), 24-36. doi: 10.2469/faj.v55.n5.2296

Beneish, M.D., Lee, C.M.C., & Nichols, D.C. (2013). Earning manipulation and expected returns. *Financial Analysts Journal, 69*(2), 57-82. doi: 10.2469/faj.v69.n2.1

Bradford, J.P., Kunz, C., Kohavi, R., Brunk, C., & Brodley, C.E. (1998). Pruning decision trees with misclassification costs. *Proceedings of the European Conference on Machine Learning* (pp. 131-136). Berlin: Springer. doi: 10.1007/bfb0026682

Braz, R., Amir, E., & Roth, D. (2008). A survey of first-order probabilistic models. In D.E. Holmes & L.C. Jain (Eds.), *Innovations in Bayesian Networks: Theory and applications. Studies in computational intelligence, 156* (pp. 289-317). Berlin: Springer. doi: 10.1007/978-3-540-85066-3_12

Cleary, R., & Thibodeau, J.C. (2005). Applying digital analysis using Benford's Law to detect fraud: The dangers of type 1 errors. *Auditing: A Journal of Practice & Theory, 24*(1), 77-81. doi: 10.2308/aud.2005.24.1.77

D'Amico, E., & Mafrolla, E. (2013). The importance of earnings management detection models to identify fraud: A case from Italian listed firms. *Journal of Modern Accounting and Auditing, 9*(1), 68-75.

De Angelo, L. (1986). Accounting numbers as market valuation substitutes: A study of management buyouts of public shareholders. *The Accounting Review, 61*(3), 400-420.

Dikmen, B., & Kukkocaoglu, G. (2010). The detection of earnings manipulation: The three-phase cutting plane algorithm using mathematical programming. *Journal of Forecasting, 29*(5), 442-466. doi: 10.1002/for.1138

Diware, K., Borhade, A., & Ringe, S. (2016). A holistic study of top data mining algorithms. *International Journal of Advanced Research in Computer Science and Software Engineering, 6*(2), 397-404.

Do, C.B., & Batzoglou, S. (2008). What is the expectation maximization algorithm? *Nature Biotechnology, 26*(8), 897-899. doi: 10.1038/nbt1406

Domingos, P., & Pazzani, M. (1997). Beyond independence: Conditions for the optimality of the simple Bayesian classifier. *Machine Learning, 29*(2), 103-130. doi: 10.1023/a:1007413511361

Doreswamy, K.S., & Nagaraju S. (2010). Naive Bayesian classifier for knowledge discovery from materials informatics. *Proceedings of the International Conference on Information and Communication Technology (ICICT)* (pp. 228-233). Maduari, Tamilnadu: India.

Drabkova, Z. (2016). Models of detection of manipulated financial statements as part of the internal control system of the entity. *ACRN Oxford Journal of Finance and Risk Perspectives Special Issue of Finance Risk and Accounting Perspectives, 5*(1), 227-235.

Durtschi, C., Hillison, W., & Pacini, C. (2004). The effective use of Benford's Law to assist in detecting fraud in accounting data. *Journal of Forensic Accounting, 5*(1), 17-34.

Elkan, C. (2001a). Magical thinking in data mining. *Proceedings of the Seventh International Conference on Knowledge Discovery and Data Mining (KDD)* (pp. 426-431). San Francisco, CA: USA.

Elkan, C. (2001b). The foundations of cost-sensitive learning. *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)* (pp. 973-978). Seattle, WA: USA.

Findik, H., & Ozturk, E. (2016). Measurement of financial information manipulation with the help of Beneish model: A research on BIST manufacturing industry. *Journal of Business Research Turk, 8*(1), 483-499.

Green, B.P., & Choi, J.H. (1997). Assessing the risk of management fraud through neural network technology. *Auditing: A Journal of Practice & Theory 16*(1), 14-28.

Gupta, R., & Gill, N. (2012). Prevention and detection of financial statement fraud: An implementation of data mining framework. *International Journal of Advanced Computer Science and Applications, 3*(8), 150-156.

Hariri, O., & Brewer, L. (2004). If Collin Powel had commanded Enron: The hidden foundation of leadership. *Business Strategy Review, 15*(2), 37-45. doi: 10.1111/j.0955-6419.2004.00312.x

Healy, P.M. (1985). The effect of bonus schemes on accounting decisions. *Journal of Accounting and Economics, 7*(1), 85-107. doi: 10.1016/0165-4101(85)90029-1

Heckerman, D. (1998). A tutorial on learning with Bayesian Networks. In M.I. Jordan (Ed.), *Learning in graphical models* (pp. 301-354). Dordrecht: Springer. doi: 10.1007/978-94-011-5014-9_11

Hematfar, M., & Hemmati, M. (2013). A comparison of risk-based and traditional auditing and their effect on the quality of audit reports. *International Research Journal of Applied and Basic Sciences, 4*(8), 2088-2091.

Herawati, N. (2015). Application of Beneish M-Score models and data mining to detect financial fraud. *Procedia - Social and Behavioral Sciences, 211*, 924-930. doi: 10.1016/j.sbspro.2015.11.122

Hogan, C.E., Rezaee, Z., Riley, R.A., & Velury, U.K. (2008). Financial statement fraud: Insights from the academic literature. *Auditing: A Journal of Practice & Theory, 27*(2), 231-252. doi: 10.2308/aud.2008.27.2.231

Jones, J.J. (1991). Earnings management during import relief investigations. *Journal of Accounting Research, 29*(2), 193-228. doi: 10.2307/2491047

Kamal, M., Saleh, M., & Ahmad, A. (2014). Detecting earnings management manipulation by Malaysian public listed companies: The reliability of Beneish M-Score Model. *Proceedings of Malaysia Indonesia International Conference on Economics, Management and Accounting (MIICEMA)* (pp. 67-81). Bangi, Selangor, Malaysia. Available at http://eprints.unsri.ac.id/5126/1/Prosiding_MIICEMA_Malaysia.pdf (accessed on 5 April 2016).

Kara, E., Ugurlu, M., & Korpi, M. (2015). Using Beneish model in identifying accounting manipulation: An empirical study in BIST manufacturing industry sector. *Journal of Accounting, Finance and Auditing Studies, 1*(1), 21-39.

Kighir, A., Omar, N., & Mohamed, N. (2014). Earnings management detection modeling: A methodological review. *World Journal of Social Sciences, 4*(1), 18-32.

Kirkos, E., Spathis, C., & Manolopoulos, Y. (2007). Data mining techniques for the detection of fraudulent financial statements. *Expert Systems with Applications, 32*(4), 995-1003.

Koskivaara, E. (2004). Artificial neural networks in analytical review procedures. *Managerial Auditing Journal, 19*(2), 191-223. doi: 10.1016/j.eswa.2006.02.016

Kotsiantis, S.B. (2007). Supervised machine learning: A review of classification techniques. *Informatica, 31*(3), 249-268.

Kuenkaikaew, S., & Vasarhelyi, M.A. (2013). The predictive audit framework. *The International Journal of Digital Accounting Research, 13*, 37-71. doi: 10.4192/1577-8517-v13_2

Lee, T.H., & Ali, A. (2008). The evolution of auditing: An analysis of the historical development. *Journal of Modern Accounting & Auditing, 4*(12), 1-8.

Lin, J.W., Hwang, M.I., & Becker, J.D. (2003). A fuzzy neural network for assessing the risk of fraudulent financial reporting. *Managerial Auditing Journal, 18*(8), 657-665. doi: 10.1108/02686900310495151

Mahama, M. (2015). Detecting corporate fraud and financial distress using the Altman and Beneish models: The case of Enron Corp. *International Journal of Economics, Commerce and Management, 3*(1), 1-18.

Margineantu, D.D. (2002). Class probability estimation and cost-sensitive classification decisions. Machine Learning. *Proceedings of the 13th European Conference on Machine Learning (ECML)* (pp. 270-281). Helsinki: Finland. doi: 10.1007/3-540-36755-1_23

Miller, J.E. (2009). The development of the Miller Ratio (MR): A tool to detect for the possibility of earnings management (EM). *Journal of Business & Economics Research 7*(1), 79-90. doi: 10.19030/jber.v7i1.2252

Moattar, M.H., & Fadaei Noghani, F. (2016). Ensemble classification and extended feature selection for credit card fraud detection. *Journal of AI and Data Mining*. Available at http://jad.shahroodut.ac.ir/article_788_4aca5dbb9d67e9211cee82aa9367abc0.pdf (accessed on 22 October 2016).

Moorthy, M.K., Mohamed, A.S.Z., Gopalan, M., & San, L.H. (2011). The impact of information technology on internal auditing. *African Journal of Business Management, 5*(9), 3523-3539.

Nedelcu, B. (2013). Business Intelligence Systems. *Database Systems Journal*, 4(4), 12-20.

Nguyen, T., Doncescu, A., & Siegel, P. (2016). Performance comparison of ADTree and Naive Bayes algorithms for spam filtering. *World Academy of Science, Engineering and Technology, International Journal of Mathematical, Computational, Physical, Electrical and Computer Engineering, 10*(5), 265-270.

Nigrini, M., & Mittermaier, L.J. (1997). The use of Benford's Law as an aid in analytical procedures. *Auditing: A Journal of Practice & Theory, 16*(2), 52-67.

Nwoye, U.J., Okoye, E.I., & Oraka, A.O. (2013). Beneish model as effective complement to the application of SAS No. 99 in the conduct of audit in Nigeria. *Management and Administrative Sciences Review, 2*(6), 640-655.

Omar, N., Koya, R.K., Sanusi, Z.M., & Shafie, N.A. (2014). Financial statement fraud: A case examination using Beneish model and ratio analysis. *International Journal of Trade, Economics and Finance, 5*(2), 184-186. doi: 10.7763/ijtef.2014.v5.367

Paolone, F., & Magazzino, C. (2014). Earnings manipulation among the main industrial sectors: Evidence from Italy. *Economia Aziendale Online, 5*(4), 253-261.

Patil, T.R., & Sherekar, S.S. (2013). Performance analysis of Naive Bayes and J48 classification algorithm for data classification. *International Journal of Computer Science and Applications, 6*(2), 256-261.

Pustylnick, I. (2011). Empirical algorithm of detection of manipulation with financial statements. *Journal of Accounting, Finance and Economics, 1*(2), 54-67.

Russell S.J., & Norvig P. (1995). *Artificial intelligence: A modern approach.* New Jersey, USA: Prentice Hall.

Saad, A., Zaarour, I., Bejjani, P., & Ayache, M. (2012). Handwriting and speech prototypes of Parkinson patients: Belief Network approach. *International Journal of Computer Science, 9*(3), 499-505.

Saad, A., Zaarour, I., & Zeinedine, A. (2013). A preliminary study of the causality of freezing of gait for Parkinson's disease patients: Bayesian belief network approach. *International Journal of Computer Science Issues, 10*(3), 88-95.

Salehi, M. (2008). Evolution of accounting and auditors in Iran. *Journal of Audit Practice, 5*(4), 57-74.

Sankar, M.R., & Subramanyam, K.R. (2001). Reporting discretion and private information communication through earnings. *Journal of Accounting Research, 39*(2), 365-386. doi: 10.1111/1475-679x.00017

Sayana, S.A. (2003). Using CAATs to support IS audit. *Information Systems Control Journal, 1*, 1-3.

Sharma, A., & Panigrahi, K.P. (2012). A review of financial accounting fraud detection based on data mining techniques. *International Journal of Computer Applications, 39*(1), 37-47. doi: 10.5120/4787-7016

Simpson, C.K. (2016). An examination of Enron corporation: Could the fraud have been detected sooner? *International Journal of Scientific and Research Publications, 6*(7), 64-78.

Turban, E., Aronson J.E., & Liang T.P. (2005). *Decision support systems and intelligent systems.* New Jersey, USA: Prentice Hall.

Tumi, A. (2013). An investigative study into the perceived factors precluding auditors from using CAATs and CA. *International Journal of Advanced Research in Business, 1*(3), 1-45.

Utami, I., & Nahartyo E. (2016). The impact of interactive reviews with group support system on information ambiguity. *Asian Journal of Business and Accounting, 9*(1), 105-139.

Vasile-Daniel, C. (2010). How financial auditors use CAATS and perceive ERP systems? *Annals of the University of Oradea, Economic Science Series, 19*(1), 490-495.

Verner, J. (2012). Why you need internal audit at the table. *Business Finance.* Available at http://businessfinancemag.com (accessed on 7 September 2016).

Yue, D., Wu, X., Wang, Y., Li., Y., & Chu, C.H. (2007). A review of data mining-based financial fraud detection research. *Proceedings of the International Conference on Wireless Communications: Networking and Mobile Computing (WiCOM)* (pp. 5519-5522). Shanghai: China. doi: 10.1109/wicom.2007.1352

Zaarour, I., Saad, A., Eddine, A.Z., Ayache, M., Guerin, F., Bejjani, P., & Lefebvre, D. (2015). Methodologies for the diagnosis of the main behavioral syndromes for Parkinson's disease with Bayesian Belief Networks. In Q.N. Tran, & H. Arabnia (Eds.), *Emerging trends in computational biology, bioinformatics, and systems biology algorithms and software tools* (pp. 467-485). Waltham, MA, USA: Elsevier. doi: 10.1016/b978-0-12-802508-6.00026-0Z

Zhang, H. (2004). The Association for the Advancement of Artificial Intelligence (AAAI) and the MIT Press. In V. Barr & Z. Markov (Eds.), *The Optimality of Naive Bayes*. In *Proceedings of the Seventeenth Florida Artificial Intelligence Research Society Conference* (FLAIRS) (pp. 562-567). Florida: USA. Available at http://www.cs.unb.ca/profs/hzhang/publications/FLAIRS04ZhangH.pdf (accessed on 9 April 2016).

Zhang, D., & Zhou, L. (2004). Discovering golden nuggets: Data mining in financial application. *IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews), 34*(4), 513-522. doi: 10.1109/tsmcc.2004.829279