Predicting Audit Quality Using Machine Learning Algorithms

Abstract

Audit quality has always been the focus of audit research, especially since the passage of the Sarbanes-Oxley Act in 2002. Much research has been done to measure and predict audit quality, and the existing predictive models commonly use regression. By contrast, this paper uses various supervised learning algorithms to predict audit quality, which is proxied by restatements, the best measure of audit quality that is publicly available (Aobdia, 2015). Using 14,028 firm-year observations from 2008 to 2016 in the United States and ten different supervised learning algorithms, the research mainly shows that Random Forest algorithm can predict audit quality more accurately than logistic regression, and that audit-related variables are better than financial variables in predicting audit quality. The results of this paper can provide regulators, investors, and other stakeholders a more effective tool than the traditional logistic regression to assess and predict audit quality, thus better protecting the benefit of the general public and ensuring the healthy functioning of the capital market.

Key words:

Audit Quality, Machine Learning Algorithms, Restatements

1.Introduction

Audit quality has been a major topic in audit research published over the last two decades, especially since the Enron scandal and the passage of the Sarbanes-Oxley Act (M. DeFond and Zhang, 2014). These research mainly focus on finding the causal relationships between audit quality and other variables of interest using regression models (Becker, Defond, Jiambalvo, and Subramanyam, 1998; Deis and Giroux, 1992; Eshleman and Guo, 2014; Francis and Yu, 2009; Ghosh, 2005; Lennox, Wu, and Zhang, 2014). Though it is hard to measure audit quality because the amount of assurance that auditors provide is unobservable (M. DeFond and Zhang, 2014), various proxies have been used to infer audit quality, such as financial statement restatements, going concern opinions, and abnormal accruals etc. (M. DeFond and Zhang, 2014). Among all the proxies for audit quality, restatement and whether the issuer meets/beats the zero earnings threshold are shown to be the best publicly available measures in terms of their predictive ability to the PCAOB Part 1 Findings, which is an accurate measure of audit process quality (Aobdia, 2015).

Many factors have been identified to affect audit quality, such as abnormal audit fees (Blankley, Hurtt, and MacGregor, 2012), auditor industry specialization (Romanus, Maher, and Fleming, 2008), auditor changes (Romanus et al., 2008), brand name of the auditor (Eshleman and Guo, 2014), and auditor size (Francis and Yu, 2009). To model and predict audit quality, the mainstream research uses linear regression (Francis and Yu, 2009) or logistic regression (Francis and Yu, 2009; Lennox et al., 2014), depending on whether the dependent variable, the proxy for audit quality, is continuous or discrete. However, if the main purpose is to make predictions, some machine learning algorithms may perform better than regressions. With all the current knowledge of what can affect audit quality, machine learning algorithms can be constructed for the use of regulators, investors, and other stakeholders to assess and predict audit quality more accurately, thus better protecting the benefit of the general public and ensuring the efficient functioning of the capital market.

In the machine learning domain, regression is a subset of supervised learning, in which the algorithms learn from the available examples with known "labels" (Alpaydin, 2014). Besides regression, other common supervised learning algorithms are Artificial Neural Networks (ANN), Decision Tree (DT), Naïve Bayes (NB), and Support Vector Machine (SVM) etc. Supervised learning algorithms have been very successful in performing prediction tasks such as image/voice recognition, email classification, fraud detection, and bankruptcy prediction etc. (Alpaydin, 2014). However, until now, no published research has been done to predict audit quality using these supervised learning algorithms. Therefore, this paper aims to fill that gap by using multiple supervised learning algorithms to model and predict audit quality, which is proxied by financial statement restatement, the best measure of audit quality that is publicly available (Aobdia, 2015).

This paper addresses four research questions: 1) how accurately can machine learning algorithms predict audit quality and which algorithms work the best? 2) which variables are the most predictive of audit quality? 3) which group of variables are more predictive of audit quality, the audit-related variables or the financial variables? 4) are the predictive abilities of the two groups of variables complementary or supplementary? Answering the above four research questions can provide a clear guidance to the regulators, investors, and other stakeholders on which algorithms and variables to choose to best predict audit quality. The results of this research show that 1) compared to regressions, machine learning algorithms, especially Random Forest, perform better in predicting audit quality; 2) the six most predictive variables are: auditor market share, client's total assets, auditor portfolio share, audit fee, auditor size, and the brand name of the auditor; 3) compared to financial variables, audit-related variables perform better in predicting audit quality; and 4) the predictive ability of the algorithms is the highest when both financial variables and audit-related variables are included in the independent variables, indicating that the two groups of variables complement each other in predicting audit quality.

This research contributes to the audit literature in three aspects: 1) this paper pioneers in constructing machine learning algorithms to predict audit quality, and provides evidence that Random Forest is more accurate in predicting audit quality than regressions; 2) this research identifies six most predictive variables of audit quality, five of which are audit related variables, providing new evidence to the previous audit quality research; 3) the results of this paper provide regulators, investors, and other stakeholders more powerful tools to assess and predict audit quality.

2.Literature Review and Research Questions

For a long time, audit quality is defined as "the market-assessed joint probability that a given auditor will both discover a breach in the client's accounting system and report the breach" (DeAngelo, 1981; M. DeFond and Zhang, 2014). However, this definition understates the benefit of high audit quality by restricting the auditor's role to the simple detection and reporting of "black and white" GAAP violations (M. DeFond and Zhang, 2014). DeFond and Zhang (2014) argue that high quality auditors should consider not only the compliance of the clients with GAAP, but also "how faithfully the financial statement reflect the firm's underlying economics". Besides, audit quality is a component of Financial Report Quality (FRQ), which also depends on the client's financial reporting system and innate characteristics (M. DeFond and Zhang, 2014). Thus, to reflect the higher level of benefit of high audit quality and the close relationship between audit quality and FRQ, DeFond and Zhang (2014) define high audit quality as the "greater assurance that the financial statements faithfully reflect the firm's underlying economics, conditioned on its financial reporting systems and innate characteristics", not just making sure the client's mechanical compliance with accounting standards.

2.1 Measurements of Audit Quality

Audit quality is hard to measure because the amount of assurance that auditors provide is unobservable (M. DeFond and Zhang, 2014). However, there are multiple proxies from which to infer audit quality, and these proxies can be classified into either the inputs or outputs of the audit process (M. DeFond and Zhang, 2014). The output-based audit quality measures include material misstatements (e.g. restatements and Accounting and Auditing Enforcement Releases (AAERs)), auditor communication (e.g. going concern opinions), financial reporting quality characteristics (e.g. discretionary accruals and meet/beat earnings targets), and perception-based measures (e.g. earnings response coefficients and cost of capital); the input-based audit quality measures include auditor characteristics (e.g. auditor size and auditor industry specialization), and auditor-client contracting features (e.g. audit fee) (M. DeFond and Zhang, 2014). The output-based audit quality proxies directly reflect the FRQ of the client (M. DeFond and Zhang, 2014). Thus it is important to disentangle the effect of audit quality from that of the client's financial reporting system and innate characteristics (M. DeFond and Zhang, 2014). Besides, auditors are responsible to obtain only "reasonable", but not absolute, assurance that material misstatements are detected, due to the nature of audit evidence and the characteristics of fraud (PCAOB, 2017). Therefore, when using output-based audit quality measures, the factors that auditors cannot control should be considered.

While there is no consensus on which audit quality measures are the best because each has its own strengths and weakness depending on the research setting (M. DeFond and Zhang, 2014), Aobdia (2015) finds that restatements and whether the client meets or beats the zero earnings threshold can better predict Part I Findings, which is an accurate measure of audit quality derived from audit deficiencies of individual engagements identified during the PCAOB inspections process, than others. Compared to the measurement of whether the issuer meets/beats the zero earnings and other proxies such as accrual-based metrics, restatement reflects the actual audit quality being delivered, thus it is a relatively strong evidence of poor audit quality (M. DeFond and Zhang, 2014). Moreover, restatement is a very direct and egregious measure of audit quality (M. L. DeFond and Francis, 2005; M. DeFond and Zhang, 2014; Romanus et al., 2008) because it indicates that the auditor mistakenly issued an unqualified opinion on materially misstated financial statements (M. DeFond and Zhang, 2014). Besides its directness and egregiousness, its dichotomized value is highly consentaneous and convenient for the purpose of making predictions. Therefore, restatement is chosen as the proxy for audit quality in this paper, in which the focus is on assessing and predicting audit quality. However, since SEC only examines one third of the public companies' financial statements, there may be some "false negatives" (will be discussed in section three) existing in the audit engagements that are not examined. Furthermore, since the restatement reflects the existence of material misstatements in the financial statements, it cannot capture the subtle audit quality variation (M. DeFond and Zhang, 2014). Moreover, the instances of restatements are relatively rare compared to the whole sample, which will result in an imbalanced dataset. To address the data imbalance and the "false negative" issues, some techniques are deployed in this research (discussed in section three).

2.2 Factors that Reflect/Affect Audit Quality

There are many factors that have been identified to reflect/affect audit quality, such as abnormal audit fees (Blankley et al., 2012), auditor industry specialization, auditor changes (Romanus et al., 2008), brand name of auditor (Eshleman and Guo, 2014), and auditor size (Francis and Yu, 2009) etc. Since this research measures audit quality by restatement, which is an output-based proxy constrained by client's financial reporting system and innate characteristics, it is important to control for the client's innate risks to disentangle their effects from those of audit quality on restatement (M. DeFond and Zhang, 2014). Thirty-six factors that have been shown from the previous literature to significantly affect/reflect audit quality or the client's innate risks are listed in Table 1 in Appendix 1. The variables that are directly related to the characteristics of the audit engagement are defined in this paper as the "audit related variables". And the rest are designated as "financial variables" because they are essentially the financial indicators of the client.

2.3 Machine Learning and Supervised Learning Algorithms

Machine learning is a subset of Artificial Intelligence (AI). At its core, machine learning is "programming computers to optimize a performance criterion using example data or past experience" (Alpaydin, 2014). By "learning" from example data or past experience, the algorithms will automatically extract the hidden knowledge of performing certain tasks that humans cannot find explicit solutions, such as pattern recognition in images and videos, classifying spam emails from legitimate ones, and predicting fraudulent behaviors (Alpaydin, 2014). There are generally three types of machine learning algorithms: supervised learning, unsupervised learning, and semi-supervised learning. In supervised learning, the algorithms are trained and tested on example or past data with “labels” (Alpaydin, 2014), for example, whether or not an email is spam, whether or not a fraudulent behavior has happened, and whether or not a voice recording comes from Bob, etc. Common supervised learning algorithms are Naïve Bayesian (NB), Bayesian Belief Network (BBN), Artificial Neural Networks (ANN), Decision Trees (DT), Support Vector Machines (SVM), Random Tree (RT) and Random Forest (RF) etc. Supervised learning algorithms are mainly used for classification/prediction tasks, and they have been used to predict economic events such as frauds and bankruptcy (Cecchini, Aytug, Koehler, and Pathak, 2010; Chen, Huang, and Kuo, 2009; Dimmock and Gerken, 2012). Different with supervised learning whose aim is “to learn a mapping from the input to an output whose correct values are provided by a supervisor”, unsupervised learning is trained and applied on unlabeled data and it focuses on finding the regularities/patterns from the input data (Alpaydin, 2014). One common method in unsupervised learning is clustering where the aim is to find clusters or grouping of input. For example, in customer segmentation, customers with similar attributes are clustered in the same group so that different services can be provided to different customer groups (Alpaydin, 2014). Semi-supervised learning falls between supervised and unsupervised learning and it is trained on a combination of labeled and unlabeled data (Castle, 2018a). Semi-supervised learning is used when labeling massive amounts of data is time-consuming and expensive, and it is commonly used in webpage classification, speech recognition, and genetic sequencing (Castle, 2018b).

2.4 Development of Research Questions

In this research, since the audit quality is measured by restatement which has “labels” (i.e. whether or not the financial statement got restated), the supervised learning algorithms should be used. Previous research on audit quality or restatement use regressions because their goal is to find causal relationships between audit quality/restatement and other variables of interest (Aier, Comprix, Gunlock, and Lee, 2005; Becker et al., 1998; Deis and Giroux, 1992; Eshleman and Guo, 2014; Francis and Yu, 2009; Ghosh, 2005; Kinney, Palmrose, and Scholz, 2004; Lennox et al., 2014; Plumlee and Yohn, 2010; Schmidt and Wilkins, 2013). However, if making predictions is the main purpose, many supervised learning algorithms other than regressions can be utilized. Though these algorithms have been very successful in predicting economic events such as fraud and bankruptcy, it is not clear whether they can be used to accurately predict audit quality and which variables should be included in the algorithm to achieve the best predictive ability. Thus, the first two research questions this research aims to address are:

RQ1: How accurately can supervised learning algorithms predict audit quality and which algorithms work the best?

RQ2: What factors are the most predictive of audit quality using supervised learning algorithms?

Restatement is an output-based audit quality proxy which is constrained by the firm's financial reporting system and its innate characteristics (M. DeFond and Zhang, 2014). Besides, audit quality is not independent of the firm's financial reporting system and its innate characteristics, because firm managers are likely to choose the quality of the financial reporting systems for an expected audit quality delivered by the auditor and that the auditors consider the quality of the firm's financial reporting system and its innate characteristics in selecting clients and in the audit planning process (M. DeFond and Zhang, 2014). Thus, to mitigate bias, this research includes both audit related variables and financial variables in the independent variables. In a similar research, Dutta, Dutta, and Raahemi (2017) predicts restatement using supervised learning algorithms. However, they do not treat restatement as a proxy for audit quality and few audit related variables are included: among their 116 independent variables, only two variables ("Big four auditor" and "Auditor Identification") are audit related, and the rest are all financial variables. Though Dutta, Dutta, and Raahemi (2017) and this research both use restatement as the dependent variable, this paper uses restatement as the proxy for audit quality, thus includes audit related variables as the major part of the independent variables. Therefore, based on Dutta, Dutta, and Raahemi (2017), other questions of interest are:

RQ3: Which group of variables are more predictive of audit quality using supervised learning algorithms, the audit related variables or the financial variables?

RQ4: Are the predictability of the two groups of variables complementary or supplementary using supervised learning algorithms?

3. Empirical Implementation

3.1 Data Collection

In this paper, the audit related data and the restatement data come from the Audit Analytics database and the financial data come from COMPUSTAT. The time period of the sample spans from 2008 to 2016. This research chooses 2008 as the starting point because it is post-SOX and post-financial crisis. In this paper, the instances of restatements in 10-Ks due to accounting errors and fraud are used. The details of how the restated instances are generated for this research are provided in Appendix 2.

Thirty independent variables are collected and calculated based on Table 1 in Appendix 1. The other six variables are not included because they are not publicly available. Table 2 shown below lists the thirty independent variables and Table 3 in Appendix 3 provides the details of how each variable is calculated. The shaded variables are audit related variables. There are sixteen audit-related variables. The three accrual variables (TotalNetAccruals, AbnormalAccruals, and AbsAbnAcc) are regarded as both audit-related and financial variables because they are not only the indicators of audit quality (M. DeFond and Zhang, 2014), but also the financial indicators of the client. In further analysis, the accrual variables are excluded from the audit-related variables and the main results still hold.

Table 2

Independent Variables	**Description**	**Independent Variables**	**Description**
LAF	Logarithm of audit fees	LTA	Logarithm of end of year total assets
DLAF	Difference of Log(Audit Fees)	LEV1	Capital Structure
Auditor Size	Measure of practice office size based on aggregated client audit fees of a practice office in a specific fiscal year.	LEV2	Capital Structure
Big4	Whether or not the auditor is Big4	FREEC	Demand for external financing
Auditor Change	Whether or not the client changed auditor	SALESGROWTH	One-year growth rate of a firm's sales revenue
INFLUENCE	Ratio of a specific client's audit fees relative to the aggregated audit fees generated by the practice office that audits the client	OCF	Operating cash flows deflated by lagged total assets
TENURE	Measure the familiarity between the auditor and the client.	BANKRUPTCY	The Altman Z-score, which is a measure of the probability of bankruptcy, with a lower value indicating greater financial distress.
GC	Going concern opinion	BMratio	Book to Market ratio
PRIORGC	Previous going concern opinion	SMALL_PROFIT	Whether or not the client has small profits
AuditorMarketShare	Auditor market share	SMALL_INCREASE	Whether or not the client has slight profits
AuditorPortfolioShare	Auditor portfolio share	FIN	Financing
WeightedMarketShare	Weighted auditor market share based on client sales	ACC	Change in noncash working capital plus change in noncurrent operating assets plus change in net financial assets, scaled by total assets (Richardson et al. 2002)

Specialist	Whether or not the auditor is considered as a specialist	EXANTE	1 if firm's free cash flow is <-0.1, and 0 otherwise where free cash flow is net income less accruals divided by average of last three years capital expenditures
TotalNetAccurals	Total net accruals	EPSGWTH	Growth in EPS
AbnormalAccurals	Abnormal accurals	materialweakness	Internal control indicator disclosed by the client in its SOX 302 disclosure
AbsAbnAcc	Absolute value of abnormal accruals		

The general steps of constructing the sample dataset are as follows: 1) merge the restatement data, the audit-related variables, and the financial variables by CIK code and fiscal year; 2) choose the peers of the restated firm-year observations by matching the restated instances with the non-restated ones with the same SIC code and fiscal year (Cecchini et al., 2010); 3) delete observations with missing values; and 4) keep the observations from 2008 to 2016. The final sample has a total of 14,028 firm-year observations from 2008 to 2016, with the restated instances counting for 7.6% of the whole sample. The details of how the sample data was filtered are provided in Table 4.

Table 4

	# firm-year observations
Financail data merged with restatement from 2002 to 2016	245299
Less: missing value in DLAF	-145312
Less: missing value in GC	-26104
Less: missing value in FE	-623
Less: missing value in SALESGROWTH	-14057
Less: missing value in Zscore	-16555
Less: missing value in LEV1	-40
Less: missing value in FREEC	-50
Less: missing value in AbsAbnAcc	-432
Less: missing value in materialweakness	-2241
Less: missing value in BM ratio	-73

Financial data merged with restatement from 2002 to 2016 (no missing values)	39812
Less: non-restatement observations whose SIC and FE never appear in those of restatement instances	-21296
Matched sample from 2002 to 2016	18516
Less: observations from 2002 to 2007	-4488
Matched sample from 2008 to 2016	14028

3.2 Descriptive Statistics and Pearson Correlation

The percentages of restated instances in each year in the sample dataset are listed in Table 5, and the trend of the percentages over the sample years is plotted in Figure 1. The figure shows that, in this sample, the percentage of restated instances has increased steadily since 2008 and started to decline from 2013, indicating that the regulators were being more strict about the financial report quality since the financial crisis and that the audit quality and financial report quality have been improved later on.

Figure 1

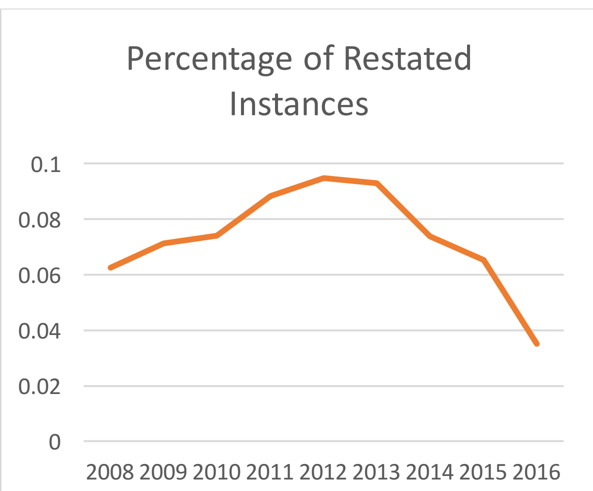


Table 5

Percentages of Restated Instances			
FE	**NumFirm**	**Num_Res**	**Res_Rate**
2008	1568	98	0.0625
2009	1461	104	0.0712
2010	1757	130	0.0740
2011	1698	150	0.0883
2012	1922	182	0.0947
2013	1777	165	0.0929
2014	1623	120	0.0739
2015	1425	93	0.0653
2016	797	28	0.0351

The descriptive statistics for the whole sample are provided in Table 6. The outliers are kept in the sample because the supervised learning algorithms used in this research are not sensitive to outliers

(Alpaydin, 2014); and deleting outliers may even cause loss of useful information that are useful for efficient classification. In the additional analysis, the winsorized data are used to perform the same analysis, and the main results still hold. From the descriptive statistics, about 10% of the observations disclosed material weakness, most (85.3%) of the auditors have been with the client for at least three years, about 12% of the sample observations received Going Concern Opinions, and more than half (62.1%) percent of the sample observations were audited by Big4 firms.

The Pearson correlation matrix is provided in Table 7. LAF (Log of Audit Fees) is statistically significantly correlated with most of the financial variables. However, only the correlation between LAF and LTA (Log of Total Asset) is economically significant (the correlation is 0.9017). This might be due to the fact that more audit effort is generally expended on larger firms, resulting in higher audit fees. Although some other audit variables are significantly correlated with some of the financial variables, the correlation coefficients are small enough to be ignored.

Table 6

Summary statistics of the whole dataset (2008-2016)					
Variable	**Obs**	**Mean**	**Std.**	**Min**	**Max**
Restatement	14,028	0.076	0.265	0	1
LAF	14,028	13.388	1.467	8.006	18.001
DLAF	14,028	0.031	0.308	-2.179	4.309
LTA	14,028	5.488	2.725	-6.908	12.906
LEV1	14,028	1.044	29.349	0.000	3172.479
LEV2	14,028	4.619	224.738	0.000	25968.970
FREEC	14,028	-0.328	4.877	-266.000	33.400
materialwe~s	14,028	0.103	0.304	0	1
AuditorSize	14,028	19.273	3.364	8.987	21.932
INFLUENCE	14,028	0.052	0.157	0.0000125	1
TENURE	14,028	0.853	0.354	0	1
SALESGROWTH	14,028	1.757	81.857	-9.286	9326.500
OCF	14,028	-0.262	4.766	-264.000	33.406
ZScore	14,028	-22.452	1246.852	-67719.000	112927.100
BMratio	14,028	-6.685	752.541	-89098.640	91.474

TotalNetAc~s	14,028	361.263	3985.768	-57027.000	179488.000
AbnormalAc~s	14,028	-1.469	45.508	-1448.237	4174.311
AbsAbnAcc	14,028	3.206	45.419	0.002	4174.311
SMALL_PROFIT	14,028	0.203	0.402	0	1
SMALL_INCR~E	14,028	0.041	0.198	0	1
GC	14,028	0.117	0.322	0	1
PRIOGC	14,028	0.146	0.353	0	1
Specialist	14,028	0.987	0.113	0	1
WeightedMa~e	14,028	0.003	0.006	1.58E-07	0.3557009
AuditorPor~e	14,028	0.074	0.169	0.0001192	1
AuditorMar~e	14,028	0.194	0.197	0.0000125	1
FIN	14,028	0.163	1.661	-0.0006973	116.7
EXANTE	14,028	0.602	0.489	0	1
EPSGrowth	14,028	0.457	0.498	0	1
AuditorCha~e	14,028	0.062	0.241	0	1
Big4	14,028	0.621	0.485	0	1

Table 7

Pearson Correlation Matrix for the whole sample

	Restat~t	LAF	DLAF	LTA	LEV1	LEV2	FREEC	materi~s	Audit~ze	INFLUE~E	TENURE	SALESG~H	OCF	ZScore	BMratio
Restatement	1														
LAF	0.1111*	1													
DLAF	0.0298*	0.0391*	1												
LTA	0.1014*	0.9017*	0.014	1											
LEV1	0.0066	-0.0501*	-0.0018	-0.0890*	1										
LEV2	0.0002	-0.0357*	-0.0069	-0.0546*	0.1930*	1									
FREEC	0.0033	0.1261*	-0.0126	0.1698*	-0.0554*	-0.0306*	1								
materialwe~s	0.0371*	-0.2724*	0.0975*	-0.3292*	0.0220*	0.0111	-0.0969*	1							
AuditorSize	0.0794*	0.7857*	-0.0263*	0.7159*	-0.0499*	-0.0323*	0.1126*	-0.3073*	1						
INFLUENCE	-0.0176*	-0.4004*	0.0444*	-0.3485*	0.0171*	0.0065	-0.0404*	0.1456*	-0.6439*	1					
TENURE	0.0273*	0.2976*	-0.0351*	0.2827*	-0.0123	-0.0021	0.0548*	-0.1980*	0.3298*	-0.1293*	1				
SALESGROWTH	-0.0006	-0.0195*	0.0803*	-0.0163	-0.0002	-0.0004	-0.0129	0.0119	-0.0206*	0.0054	-0.0250*	1			
OCF	0.0029	0.1269*	-0.0082	0.1738*	-0.0569*	-0.0314*	0.9986*	-0.0977*	0.1128*	-0.0398*	0.0536*	-0.0111	1		
ZScore	-0.0094	0.0378*	0.0278*	0.0835*	-0.1834*	-0.5177*	0.0813*	-0.0198*	0.0318*	0.0009	-0.003	0.0003	0.0832*	1	
BMratio	0.0023	0.0186*	0.004	0.0227*	-0.0083	-0.9755*	0.0015	0.002	0.0156	0.0001	-0.0024	0.0002	0.0016	0.4587*	1
TotalNetAc~s	-0.0068	0.1463*	0.0611*	0.1622*	-0.0022	0.0109	0.0075	-0.0274*	0.0603*	-0.0260*	0.0291*	-0.0013	0.0085	-0.0039	-0.0118
AbnormalAc~s	-0.0037	-0.0272*	0.0248*	-0.0177*	0.0083	0.5136*	-0.0181*	-0.0071	-0.0214*	0.0033	0.0003	0.0278*	-0.0153	-0.2293*	-0.5248*
AbsAbnAcc	-0.0009	-0.0499*	0.0180*	-0.0560*	0.0288*	0.5252*	-0.1399*	0.0217*	-0.0432*	0.0097	-0.007	0.0273*	-0.1381*	-0.2728*	-0.5250*
SMALL_PROFIT	0.0710*	0.2050*	-0.0023	0.2477*	-0.0136	-0.0092	0.0368*	-0.0813*	0.1586*	-0.0718*	0.0720*	-0.0097	0.0368*	0.0103	0.0049
SMALL_INCR~E	0.0477*	0.0628*	-0.0102	0.0738*	-0.0054	-0.0037	0.0146	-0.0107	0.0448*	-0.0153	0.0124	-0.0041	0.0145	0.0042	0.0021
GC	-0.0387*	-0.4543*	0.0032	-0.5476*	0.0770*	0.0500*	-0.1657*	0.3561*	-0.4400*	0.2176*	-0.1975*	0.0096	-0.1685*	-0.0800*	-0.0258*
PRIOGC	-0.0397*	-0.4893*	0.0433*	-0.5546*	0.0678*	0.0442*	-0.1529*	0.3566*	-0.4953*	0.2522*	-0.2465*	0.0374*	-0.1537*	-0.0707*	-0.0222*
Specialist	0.0091	0.1859*	0.0376*	0.1717*	-0.0101	-0.0800*	0.0685*	-0.1048*	0.1105*	0.0239*	0.0669*	-0.0028	0.0704*	0.0601*	0.0744*
WeightedMa~e	-0.0202*	0.2337*	0.0147	0.2195*	-0.0125	-0.009	0.0269*	-0.0777*	0.2039*	0.0099	0.0897*	-0.0048	0.0269*	0.01	0.0047
AuditorPor~e	-0.0384*	-0.4319*	0.0424*	-0.3731*	0.0151	0.0049	-0.0421*	0.1478*	-0.6726*	0.9439*	-0.1320*	0.0095	-0.0407*	0.0007	0.0012
AuditorMar~e	0.1546*	0.6137*	0.0024	0.5630*	-0.0267*	-0.0178*	0.0638*	-0.1967*	0.6581*	-0.3071*	0.2588*	-0.0133	0.0630*	0.0203*	0.0087
FIN	0.003	-0.0492*	0.0203*	-0.0840*	0.4928*	0.1102*	-0.1491*	0.0403*	-0.0382*	0.0117	-0.0142	0.0002	-0.1509*	-0.1136*	0.0003
EXANTE	-0.0243*	-0.0501*	0.1067*	-0.0183*	0.0021	0.0062	0.0042	0.0332*	-0.0403*	0.0144	-0.0229*	-0.0078	0.0073	-0.0105	-0.0071
EPSGrowth	-0.0296*	-0.0037	0.0240*	0.0173*	0.0142	0.0115	-0.0122	-0.0006	-0.0068	0.0148	-0.0073	-0.0071	-0.0123	-0.0009	-0.0097
AuditorCha~e	-0.0123	-0.2094*	-0.0201*	-0.1938*	-0.0016	-0.0026	-0.0132	0.1540*	-0.2287*	0.0861*	-0.6188*	0.0325*	-0.0126	0.0240*	0.0015
Big4	0.0945*	0.7130*	-0.0171*	0.6515*	-0.0352*	-0.0232*	0.0832*	-0.2626*	0.8900*	-0.4129*	0.3461*	-0.0172*	0.0828*	0.0264*	0.0114

Pearson Correlation Matrix for the whole sample (Cont.)

	TotalN~s	Abnorm~s	AbsAbn~c	SMALL_~T	SMALL_~E	GC	PRIOGC	Specia~t	Weight~e	A~oShare	A~tShare	FIN	EXANTE	EPSGro~h	Audit~ge	Big4
TotalNetAc~s	1															
AbnormalAc~s	0.0077	1														
AbsAbnAcc	0.0033	0.8324*	1													
SMALL_PROFIT	0.0291*	-0.0074	-0.0118	1												
SMALL_INCR~E	0.0035	-0.0031	-0.0048	0.4099*	1											
GC	-0.0369*	0.0278*	0.0642*	-0.1687*	-0.0652*	1										
PRIOGC	-0.0394*	0.0292*	0.0591*	-0.1706*	-0.0630*	0.6840*	1									
Specialist	0.0094	-0.0907*	-0.1031*	0.0167*	0.0109	-0.1135*	-0.1084*	1								
WeightedMa~e	0.0858*	-0.005	-0.0111	0.0450*	-0.0012	-0.1075*	-0.1234*	0.0592*	1							
AuditorPor~e	-0.0261*	0.0026	0.008	-0.0836*	-0.0222*	0.2279*	0.2606*	0.0362*	0.0545*	1						
AuditorMar~e	0.0862*	-0.012	-0.0240*	0.1421*	0.0298*	-0.2698*	-0.3005*	0.1109*	0.4210*	-0.3318*	1					
FIN	-0.0001	0.0407*	0.0428*	-0.015	-0.0063	0.0910*	0.0873*	-0.002	-0.0077	0.0168*	-0.0115	1				
EXANTE	0.1475*	0.0027	-0.0153	-0.0299*	-0.0453*	0.0488*	0.0753*	0.0206*	0.0308*	0.0236*	-0.0309*	0.0188*	1			
EPSGrowth	0.0417*	0.0174*	0.0126	-0.1443*	-0.1447*	0.0170*	0.0007	-0.0105	0.0186*	0.012	0.0152	0.0021	0.0651*	1		
AuditorCha~e	-0.0173*	-0.0065	0.0014	-0.0475*	-0.0171*	0.1470*	0.1658*	-0.0442*	-0.0652*	0.0888*	-0.1744*	0.0064	0.0046	0.0049	1	
Big4	0.0644*	-0.0153	-0.0314*	0.1470*	0.0368*	-0.3469*	-0.3958*	0.1151*	0.2423*	-0.4558*	0.7054*	-0.0277*	-0.0304*	0.0148	-0.2404*	1

Note: the correlations that are significant at 5% level are starred. The shaded areas show that the audit related variables and the financial variables that are statistically significantly correlated with each other.

3.3 Research Design

In supervised learning, the algorithms need to be first trained on training dataset and then tested on testing dataset. The rule-of-thumb of splitting is to choose around 75% of the sample as the training data and the rest as the testing data (Alpaydin, 2014). This research sets out the data from 2008 to 2013 as the training data (count for 72.3% of the whole sample) and those from 2014 to 2016 as the testing data (count for 27.7% of the whole dataset), because the purpose of this research is to make future predictions. Table 8 summarizes the training and testing dataset.

Table 8

Original Dataset	*Dataset Period*	*No. of Instances*	*No. of Restatement Instances*	*Percentage of Restatement*
Whole	2008-2016	14028	1070	7.6%
Training	2008-2013	10183	829	8.1%
Testing	2014-2016	3845	241	6.3%

3.3.1 Preprocess Training Data

Since the restatement instances only count for 7.6% of the entire sample, the dataset is very imbalanced. Data imbalance might be problematic because the algorithms might not be able to capture enough information from the limited restated instances and therefore may blindly predict every future instance as non-restated. However, if the algorithms are trained on the training dataset that has almost the same distribution with the entire dataset (called natural or stratified distribution), the data imbalance may not be a problem (Fawcett, 2016). Imbalanced data are very common in the real world and many methods have been studied to deal with them (Fawcett, 2016). Generally speaking, data imbalance can be mitigated by either 1)creating a balanced dataset from the existing imbalanced data; or 2)adjusting the algorithms to make them more sensitive to the rare classes; or

3)constructing algorithms that can perform well on imbalanced data (Fawcett, 2016). But all the manipulations to address the data imbalance issue should be done in the training procedure, and the algorithms should always be tested on the original imbalanced testing data (Fawcett, 2016). Most studies use matching technique, especially Propensity Score Matching (PSM), to create a balanced sample (Abbott, Parker, and Peters, 2004; Aier et al., 2005; Kinney et al., 2004; Romanus et al., 2008). Dutta et al. (2017) use Synthetic Minority Oversampling Technique (SMOTE) to create a balanced dataset by generating synthetic restated instances. However, Perols (2011) argues that the imbalanced data can be kept when using classification algorithms because the goal of using classification algorithms is to establish algorithms and predictors are useful in predicting the outcome.

Since there is no consensus on whether it is necessary to deal with the data imbalance issue, this research train the algorithms using both the original imbalanced training dataset and the synthetic balanced training dataset generated by SMOTE, and then test them on the original imbalanced testing dataset. In the balanced training dataset generated by SMOTE, the synthetic instance is created by first taking the vector between the current data point and one of its k nearest neighbors, and then multiplying this vector by a random number between 0 and 1 (Dutta et al., 2017). The SMOTE filter in WEKA, a commonly used machine learning software, is used to generate the synthetic balanced training data. The summary of this balanced training data is listed in Table 9.

Table 9

Balanced Dataset	*Dataset Period*	*No. of Instances*	*No. of Restatement Instances*	*Percentage of Restatement*
Training	2008-2013	18638	9284	49.8%

When training the algorithms, either audit-related variables, or financial variables, or both are included as independent variables, because the third and fourth research questions of this research want to find out which group of variables has better predictive ability and whether they complement each other in predicting audit quality.

3.3.2 Train and Test Algorithms

The main supervised learning algorithms used in this paper are: Naïve Bayesian (NB), Bayesian Belief Network (BBN), Artificial Neural Network (ANN), Support Vector Machine (SVM), Decision Tree (DT), Random Tree (RT) and Random Forest (RF). Other advanced algorithms such as Bagging, Stacking and AdaBoost are also used. The algorithm names and their corresponding choices in WEKA are listed in Table 10, and some brief introduction of the major algorithms used are provided below.

Table 10

Algorithm Name	**Choice in WEKA**
Naïve Bayesian	WEKA>bayes>NaiveBayes
Bayesian Belief Network	WEKA>bayes>BayesNet
Artificial Neural Network	WEKA>functions>MultilayerPerceptron
Support Vector Machine	WEKA>functions>SMO
Decision Tree	WEKA>trees>J48
Random Tree	WEKA>trees>RandomTree
Random Forest	WEKA>trees>RandomForest
Bagging	WEKA>meta>Bagging
Stacking	WEKA>meta>Stacking
AdaBoost	WEKA>meta>AdaBoostM1

Naïve Bayesian and Bayesian Network

Naïve Bayesian (NB) classifier is based on Bayes' theorem with the "naïve" assumption that every pair of features are independent (Zhang, 2004). Given a class variable y and a dependent feature vector $x_1$through x_n, NB theorem state the following relationship (Zhang, 2004):

$$P(y|x_1, \dots, x_n) = \frac{P(y)P(x_1, \dots, x_n|y)}{P(x_1, \dots, x_n)}$$

The naïve independence assumption is:

$$P(x_i|y, x_1, \dots, x_{i-1}, x_{1i+1}, \dots, x_n) = P(x_i|y)$$

Thus, for all i the above relationship is simplified to

$$P(y|x_1, \dots, x_n) = \frac{P(y)\prod_{i=1}^{n} P(x_i|y)}{P(x_1, \dots, x_n)}$$

Given the input, $P(x_1, \dots, x_n)$ is constant. Therefore, the classification rule can be derived as follows:

$$P(y|x_1, \dots, x_n) \propto P(y)\prod_{i=1}^{n} P(x_i|y)$$

$$\hat{y} = \arg\max_{y} P(y)\prod_{i=1}^{n} P(x_i|y)$$

Bayesian networks use this Bayes' rule for probabilistic inference (Murphy, 1998).

Artificial Neural Network

Artificial neural network models take their inspiration from the functioning of the brain, and the backpropagation is used to train the neural network for a variety of applications (Alpaydin, 2014). When ANN is used for classification, the perceptron is the basic processing element which converts the inputs it receives into outputs as a function of a weighted sum of the inputs. For example, $y = \frac{1}{1+\exp\left[-\boldsymbol{W}^T\boldsymbol{X}\right]}$, where **x** is a vector of inputs, **w** is a vector of weights, and y is the output. The weights **w** need to be "learned" through backpropagation till the errors are minimized.

Support Vector Machine

A support vector machine determines a hyperplane in the feature space that best separates positive from negative examples, and "a feature space results from mapping the observable attributes to properties that might better relate to the problem at hand" (Cecchini et al., 2010).

Decision Tree and Random Tree

A decision tree is "a hierarchical model for supervised learning whereby the local region is identified in a sequence of recursive splits in a smaller number of steps", and it is composed of internal decision nodes and terminal leaves (Alpaydin, 2014). The goal of using decision tree is "to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features" (scikit-learn, n.d.).

Random Forest

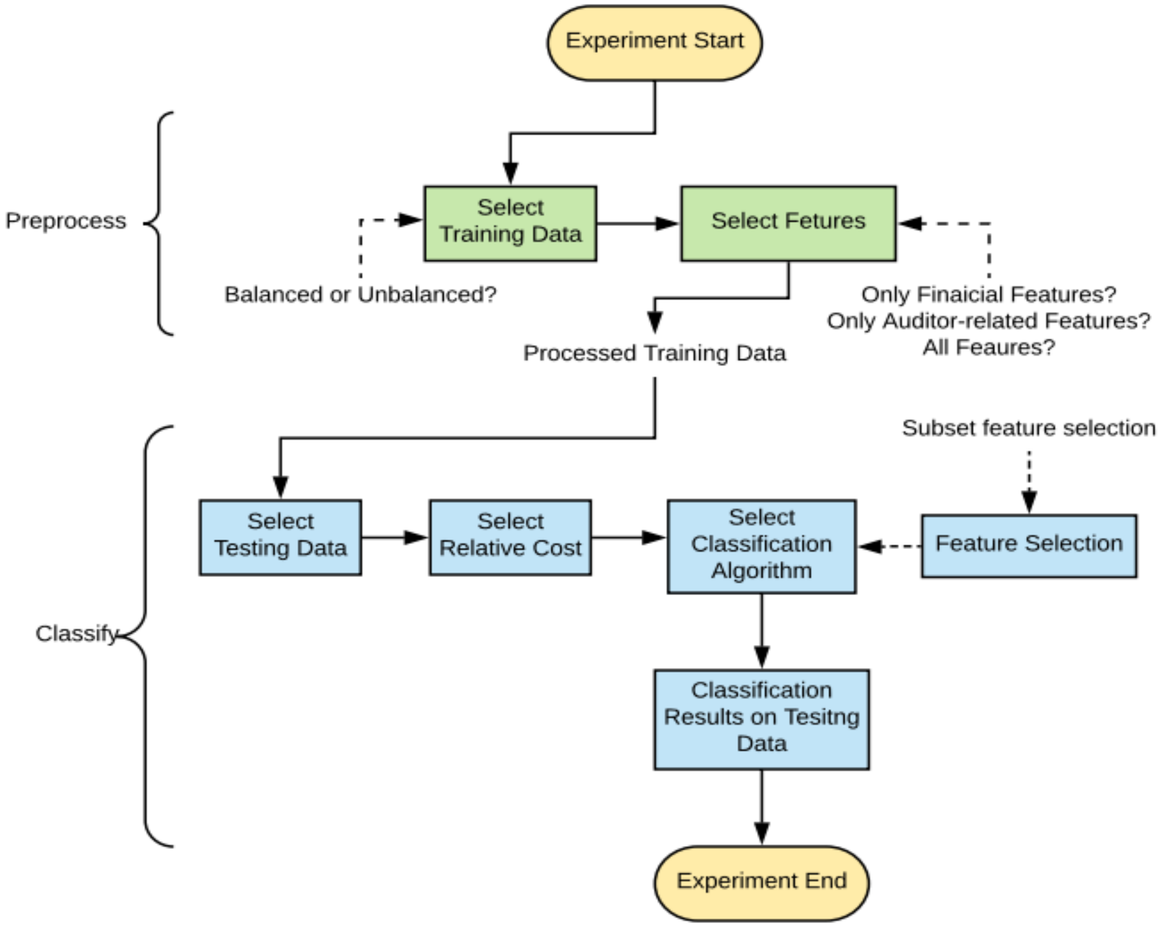
Random forest builds multiple decision trees and merges them together to get a more accurate and stable prediction (Donges, 2018). Random forests searches for the best feature among a random subset of features, and this results in a wide diversity that generally results in a better model (Donges, 2018).

In classifying instances into restated or non-restated, there are 4 outcomes: 1) the actual restated instances are correctly classified as restated (True Positive); 2) the actual restated instances are wrongly classified as non-restated (False Negative); 3) the actual non-restated instances are correctly classified as non-restated (True Negative); and 4) the actual non-restated instances are wrongly classified as restated (False Positive). In this particular context, false negative is more serious than false positive, so the cost of false negative should be higher than that of false positive. The relative cost of false negative to that of false positive is called misclassification cost. To identify under which level of misclassification cost do the algorithms work the best, the misclassification cost is set from 1 to 100 (Cecchini et al., 2010). In WEKA, the CostSensitiveClassifier can be used for this purpose.

Subset feature selection can be used to remove less significant or redundant attributes to help build parsimonious models (Dutta et al., 2017), and this research compares the performance of the algorithms with and without subset feature selection. The feature selection can be realized using the WEKA function AttributeSelectedClassifier. The evaluator chosen is CfsSubsetEval, which evaluates the worth of a subset of attributes by considering the individual predictive ability of each feature along with the degree of redundancy between them. The searching method “bi-directional” is chosen.

After the algorithms are trained on the training data, the trained algorithms will be tested on the original imbalanced testing data. Figure 2 illustrates the whole procedures from training to testing performed using WEKA.

Figure 2



3.3.3 Evaluate the Performance

As has been discussed above, there might be four outcomes when the trained algorithms classify instances in the testing data: true positive, false negative, true negative, and false positive. A confusion matrix (Figure 3) is a matrix that summarizes all the outcomes.

Figure 3

Confusion Matrix for Two Classes (Excerpted from Alpaydin, 2014)

	Predicted class		
True class	Positive	Negative	Total number
Positive	True Positive (TP)	False Negative (FN)	p
Negative	False Positive (FP)	True Negative (TN)	n
Total number	p’	n’	N

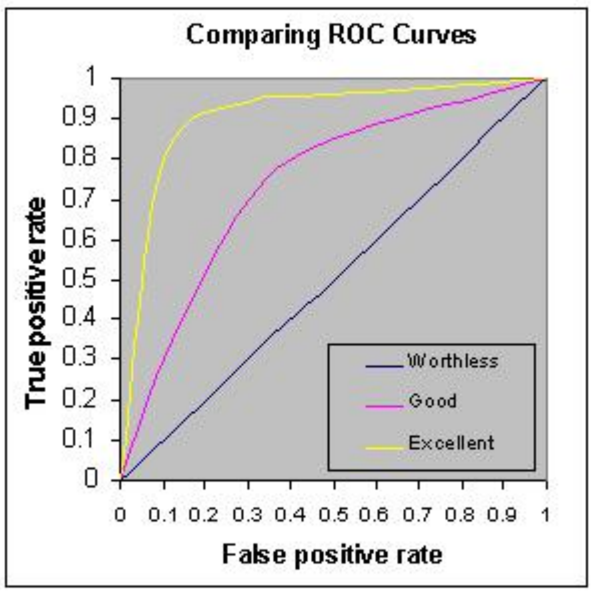
There are several indexes (listed in Table 11) that can be used to evaluate the performance of the algorithms. Due to the imbalance of the testing data, this research uses the Recall, Specificity, and the Area Under Curve (AUC) to evaluate the performance of the algorithms, because these indexes indicate how effectively the algorithms correctly classify each instance into its actual class. The closer these three indexes are to 1, the better the performance.

Table 11

Performance Measures Used in Two-class Problems (Excerpted from Alpaydin, 2014)	
Name	Formula
Error Rate	(FP+FN)/N
Accuracy Rate	(TP+TN)/N=1-error rate
TP-Rate	TP/p
FP-Rate	FP/n
Precision	TP/p’
Recall	TP/p=TP-Rate
Sensitivity	TP/p=TP-Rate
Specificity	TN/n=1-FP-Rate

The index of AUC is briefly introduced as follows. For different values of Θ, which is the threshold above which an instance will be classified as positive, the pairs of Recall and FP-Rate form the Receiver Operating Characteristics (ROC) curve (Figure 4), and the area under the ROC curve is AUC (Alpaydin, 2014). The ideal ROC always has a FP-Rate of 0 and a TP-Rate of 1, so the ideal AUC is 1. Therefore, the closer the AUC is to 1, the better the performance of the algorithm. Following the machine learning literature (Alpaydin, 2014; Cecchini et al., 2010), AUC is used in this research to compare the performance of different algorithms.

Figure 4. ROC Curves (Excerpted from Tape, n.d.)



4. Results

4.1 Imbalanced Training Data

When the algorithms are trained on the original imbalanced training data, they only perform well when the relative cost of "False Negative" and "False Positive" is 10. Among all the algorithms, Random Forest, Bagging with Random Forest, AdaBoost with Random Forest, and Stacking with Random Forest outperform the others. This might be because Random Forest is not sensitive to the imbalanced data. The testing results when the imbalanced training data are used and when the misclassification cost is 10 are listed in Table 12.

Table 12

Testing results using original imbalanced training data at the misclassification cost of 10

Input		Algorithm	Recall	Specificity	AUC	Accuracy
All Variables	Without Feature Selection	Random Forest	0.734	0.712	0.723	0.713
		Bagging with Random Forest	0.759	0.69	0.725	0.694
		AdaBoost with Random Forest	0.722	0.722	0.722	0.722
		Stacking with Random Forest	0.701	0.757	0.729	0.753
	With Feature Selection	Random Forest	0.544	0.766	0.655	0.752
		Bagging with Random Forest	0.556	0.734	0.645	0.723
		AdaBoost with Random Forest	0.544	0.765	0.654	0.751
		Stacking with Random Forest	0.515	0.764	0.639	0.748
Financial Variables	Without Feature Selection	Random Forest	0.639	0.638	0.638	0.638
		Bagging with Random Forest	0.714	0.607	0.66	0.613
		AdaBoost with Random Forest	0.025	0.997	0.511	0.936
		Stacking with Random Forest	0.606	0.63	0.618	0.629
	With Feature Selection	Random Forest	0.556	0.658	0.607	0.651
		Bagging with Random Forest	0.631	0.626	0.628	0.626
		AdaBoost with Random Forest	0.593	0.651	0.622	0.648
		Stacking with Random Forest	0.494	0.698	0.596	0.685
Audit Related Variables	Without Feature Selection	Random Forest	0.627	0.738	0.682	0.73
		Bagging with Random Forest	0.656	0.721	0.688	0.717
		AdaBoost with Random Forest	0.631	0.71	0.671	0.705
		Stacking with Random Forest	0.614	0.759	0.687	0.75
	With Feature Selection	Random Forest	0.556	0.737	0.646	0.726
		Bagging with Random Forest	0.573	0.707	0.64	0.698
		AdaBoost with Random Forest	0.116	0.974	0.545	0.92
		Stacking with Random Forest	0.531	0.715	0.623	0.703

Note:
There are 17 Non-Audit related features: LTA, LEV1, LEV2, FREEC, materialweakness, SALESGROWTH, OCF, ZScore, BNratio, TotalNetAccurals, AbnormalAccurals, AbsAbnAcc, SMALL_PROFIT, SMALL_INCREASE, FIN, EXANTE, and EPSGrowth
There are 16 Audit related features: LAF, DLAF, AuditorSize, INFLUENCE, TENURE, TotalNetAccurals, AbnormalAccurals, AbsAbnAcc, GC, PRIOGC, Specialist, WeightedMarketValue, AuditorPortfolioShare, AuditorMarketShare, AuditorChange, and Big4
The selected features from "All Features" are LTA, SMALL_INCREASE, AuditorPortfolioShare, AuditorMarketShare; The selected features from "Non Audit-related Features" are LTA, SALESGROWTH, OCF,

Zscore, SMALL_PROFIT, FIN; The selected features from "Audit-related Features" are LAF, AuditorPortfolioShare, AuditorMarketShare.

Without feature selection, each algorithm performs better when trained by audit-related variables than by financial variables, and the highest performance is achieved when it is trained by all variables. The same hold true when subset feature selection is performed. Thus, the predictive ability of audit-related variables is greater than that of financial variables, and these two groups of variables complement each other in predicting audit quality. When all the variables are used as inputs, each algorithm has a higher AUC value without subset feature selection, and the same hold true when either the audit related variables or the financial variables are used as inputs. This may be because dropping a subset of input variables causes loss of information that are useful for the classification.

To see which variables are the most predictive of audit quality, the variables are ranked in terms of their predictive ability using the evaluator in WEKA called GainRatioAttributeEval, which evaluates the worth of an attribute by measuring the gain ratio with respect to the class. This ranking is listed in Table 13 and the shaded variables are audit related variables.

Table 13

Ranking of predictability using Random Forest (with original training dataset, relative cost=10)

Rank	Variable	Rank	Variable
1	AuditorMarketShare	16	TotalNetAccruals
2	LTA	17	LEV2
3	AuditorPortfolioShare	18	AbsAbnAcc
4	LAF	19	GC
5	AuditorSize	20	AbnormalAccruals
6	Big4	21	BMratio
7	OCF	22	PRIOGC
8	FREEC	23	TENURE
9	ZScore	24	EXANTE

10	INFLUENCE	25	EPSGrowth
11	SMALL_INCREASE	26	materialweakness
12	SMALL_PROFIT	27	DLAF
13	SALESGROWTH	28	Specialist
14	FIN	29	AuditorChange
15	LEV1	30	WeightedMarketValue

Among the six most predictive variables listed in Table 13 there are five audit related variables: the market share of the auditor, the portfolio share of the auditor, log of audit fees, size of the auditor, and the brand name of the auditor. This probably explains why the algorithms perform better when they are trained by only audit-related variables than by only financial variables. Except LAF (log of audit fees), all the other most predictive audit related variables barely have economically significant correlation with the financial variables, indicating that these audit related variables provide unique information in predicting audit quality. To further prove this, the Random Forest algorithm is trained using only AuditorMarketShare and AuditorPortfolioShare variables as inputs with the misclassification cost set to 10, and the algorithm still performs well to some degree: the recall for restatement is 0.481, the recall for non-restatement is 0.734, and the AUC is 0.607.

Till now, the four research questions raised in this research can be answered as follows: 1) Random Forest algorithm works the best in predicting audit quality and can achieve an AUC value of 0.723 when trained on all variables without feature selection; 2) the most predictive variables are: the market share of the auditor, the log of client's total assets, the portfolio share of the auditor, log of audit fees, size of the auditor, and the brand name of the auditor; 3) audit-related variables have better predictive ability than financial variables; and 4) audit-related variables and financial variables complement each other in predicting audit quality.

4.2 Balanced Training Data

When the algorithms are trained on the synthetic balanced data, those that are sensitive to imbalanced data start to perform decently, for example, Bayesian Belief Network, Artificial Neural Network, and Support Vector Machine. The details of the testing results are listed in Table 14, Table 15, and Table 16. When all variables are included as inputs (Table 14), the algorithms generally don't work well if the subset feature selection is performed. Without feature selection, SVM achieves an AUC of 0.654 regardless of the level of misclassification cost; Bayesian Network works the best when the misclassification cost is 1; MultilayerPerceptron and Random Forest perform the best when the relative cost is 5; and J48 performs the best when the misclassification cost is 10. The highest AUC value (0.696) is achieved when the Random Forest is used at the misclassification cost of 5.

Table 14

Testing results using balanced training data (All variables)							
Input		Misclassification Cost	Algorithm	Recall	Specificity	AUC	Accuracy
All Variables	Without Feature Selection	1	BaysianNet	0.589	0.57	0.617	0.571
			MultilayerPerceptron	0.149	0.948	0.549	0.898
			SVM	0.577	0.73	0.654	0.721
			J48	0.178	0.935	0.556	0.887
			Random Tree	0.162	0.887	0.524	0.841
			Random Forest	0.058	0.988	0.523	0.93
		5	BaysianNet	0.651	0.507	0.579	0.516
			MultilayerPerceptron	0.469	0.738	0.603	0.721
			SVM	0.577	0.73	0.654	0.721
			J48	0.216	0.93	0.573	0.885
			Random Forest	0.73	0.662	0.696	0.666
		10	BaysianNet	0.68	0.486	0.583	0.498
			MultilayerPerceptron	0.689	0.386	0.538	0.405
			SVM	0.577	0.73	0.654	0.721
			J48	0.353	0.861	0.607	0.829
			Random Tree	0.162	0.887	0.524	0.841
			Random Forest	0.876	0.439	0.657	0.466
		20	SVM	0.577	0.73	0.654	0.721
			J48	0.481	0.725	0.603	0.71
			Random Tree	0.162	0.887	0.524	0.841
		30	SVM	0.577	0.73	0.654	0.721
			J48	0.647	0.526	0.587	0.534
	With Feature Selection	1	J48	0.187	0.784	0.485	0.747
			Random Tree	0.303	0.725	0.514	0.699
			Random Forest	0.22	0.776	0.498	0.741

Note:

The selected subset features from "All Features" are materialweakness, AuditorSize, TENURE, SMALL_PROFIT, SMALL_INCREASE, Specialist, FIN, EXANTE, EPSGrowth

Table 15

<table>
<tr><th colspan="7">Testing results using balanced training data (Financial variables)</th></tr>
<tr><th colspan="2">Input</th><th>Misclassification Cost</th><th>Algorithm</th><th>Recall</th><th>Specificity</th><th>AUC</th><th>Accuracy</th></tr>
<tr><td rowspan="18">Financial variables</td><td rowspan="18">Without Feature Selection</td><td rowspan="4">1</td><td>BaysianNet</td><td>0.266</td><td>0.848</td><td>0.557</td><td>0.812</td></tr>
<tr><td>SVM</td><td>0.656</td><td>0.583</td><td>0.619</td><td>0.588</td></tr>
<tr><td>Random Tree</td><td>0.17</td><td>0.883</td><td>0.526</td><td>0.838</td></tr>
<tr><td>Random Forest</td><td>0.021</td><td>0.991</td><td>0.506</td><td>0.93</td></tr>
<tr><td rowspan="6">5</td><td>BaysianNet</td><td>0.469</td><td>0.721</td><td>0.595</td><td>0.705</td></tr>
<tr><td>SVM</td><td>0.656</td><td>0.583</td><td>0.619</td><td>0.588</td></tr>
<tr><td>MultilayerPerceptron</td><td>0.548</td><td>0.625</td><td>0.586</td><td>0.62</td></tr>
<tr><td>J48</td><td>0.199</td><td>0.88</td><td>0.54</td><td>0.837</td></tr>
<tr><td>Random Tree</td><td>0.17</td><td>0.883</td><td>0.526</td><td>0.838</td></tr>
<tr><td>Random Forest</td><td>0.502</td><td>0.722</td><td>0.612</td><td>0.708</td></tr>
<tr><td rowspan="6">10</td><td>BaysianNet</td><td>0.539</td><td>0.667</td><td>0.603</td><td>0.659</td></tr>
<tr><td>SVM</td><td>0.656</td><td>0.583</td><td>0.619</td><td>0.588</td></tr>
<tr><td>MultilayerPerceptron</td><td>0.693</td><td>0.435</td><td>0.564</td><td>0.451</td></tr>
<tr><td>J48</td><td>0.44</td><td>0.699</td><td>0.57</td><td>0.683</td></tr>
<tr><td>Random Tree</td><td>0.17</td><td>0.883</td><td>0.526</td><td>0.838</td></tr>
<tr><td>Random Forest</td><td>0.718</td><td>0.506</td><td>0.612</td><td>0.519</td></tr>
<tr><td rowspan="2">15</td><td>J48</td><td>0.598</td><td>0.584</td><td>0.591</td><td>0.585</td></tr>
<tr><td>SVM</td><td>0.253</td><td>0.82</td><td>0.537</td><td>0.784</td></tr>
</table>

Note:
There are 17 Non Audit-related features: LTA, LEV1, LEV2, FREEC, materialweakness, SALESGROWTH, OCF, ZScore, BNratio, TotalNetAccurals, AbnormalAccurals, AbsAbnAcc, SMALL_PROFIT, SMALL_INCREASE, FIN, EXANTE, and EPSGrowth
The selected features from "Non Audit-related Features" are materialweakness, SMALL_PROFIT, FIN, EXANTE, and EPSGrowth.

When only financial variables are included as inputs (Table 15), the performance is good enough only when no subset feature selection is performed. Bayesian Network works the best when the misclassification cost is 10; SVM works well when the misclassification cost is within 10; and Random Forest works well when the misclassification cost is between 5 and 10. The highest AUC value (0.619) is achieved when SVM is used when the misclassification cost is within 10.

Table 16

Testing results using balanced training data (Audit related variables)							
Input		Misclassification Cost	Algorithm	Recall	Specificiy	AUC	Accuracy
Audit related variables	Without Feature Selection	1	Bayesian Network	0.689	0.413	0.551	0.43
			Naïve Bayesian	0.813	0.253	0.533	0.288
			MultilayerPerceptron	0.527	0.774	0.650	0.758
			J48	0.469	0.688	0.579	0.674
			Random Tree	0.39	0.778	0.584	0.754
			Random Forest	0.44	0.81	0.625	0.787
			SVM	0.743	0.409	0.576	0.430
	With Feature Selection	1	Bayesian Network	0.506	0.454	0.48	0.457
			Naïve Bayesian	0.751	0.314	0.533	0.342
			MultilayerPerceptron	0.689	0.42	0.555	0.437
			J48	0.452	0.512	0.482	0.508
			Random Tree	0.515	0.428	0.471	0.434
			Random Forest	0.519	0.421	0.47	0.427
			SVM	0.801	0.306	0.553	0.337

Note:
There are 16 Audit-related features: LAF, DLAF, AuditorSize, INFLUENCE, TENURE, TotalNetAccurals, AbnormalAccurals, AbsAbnAcc, GC, PRIOGC, Specialist, WeightedMarketValue, AuditorPortfolioShare, AuditorMarketShare, AuditorChange, and Big4

When only audit related variables are included as inputs (Table 16), the algorithms perform better when no subset feature selection is performed, and the performance is decent only when the misclassification cost is 1. The highest AUC value (0.758) is achieved when MultilayerPerceptron is used, next comes RandomForest (AUC of 0.625).

4.3 Summary of Results

No matter whether the algorithms are trained by the original imbalanced data or the synthetic balanced data, the performance is better when no subset feature selection is performed. Table 17 summarizes the results from imbalanced and balanced data without feature selection.

Table 17

Summary of overall performance (without subset feature selection)						
Input		Misclassification Cost	Algorithm	Recall	Specificity	AUC
Balanced Training Data	All Variables	1	Random Forest	0.058	0.988	0.523
			BaysianNet	0.589	0.57	0.617
		5	Random Forest	0.73	0.662	0.696
		10	Random Forest	0.876	0.439	0.657
			SVM	0.577	0.73	0.654
	Financial Variables	1	SVM	0.656	0.583	0.619
		5	SVM	0.656	0.583	0.619
			Random Forest	0.502	0.722	0.612
		10	Random Forest	0.718	0.506	0.612
			BaysianNet	0.539	0.667	0.603
		15	J48	0.598	0.584	0.591
	Audit Related Variables	1	MultilayerPerceptron	0.527	0.774	0.65
			Random Forest	0.44	0.81	0.625
Unbalanced Training Data	All Variables	10	Random Forest	0.734	0.712	0.723
	Financial Variables		Random Forest	0.639	0.638	0.638
	Audit Related Variables		Random Forest	0.627	0.738	0.682

Note:
The selected features from "All Features" are materialweakness, AuditorSize, TENURE, SMALL_PROFIT, SMALL_INCREASE, Specialist, FIN, EXANTE, EPSGrowth
The selected features from "Non Audit-related Features" are materialweakness, SMALL_PROFIT, FIN, EXANTE, and EPSGrowth.

When the algorithms are trained on balanced training data, the highest value of AUC when all variables are included is 0.696, that when only financial variables are included is 0.619, and that when only audit related variables are included is 0.650. This coincides with the result generated from the original unbalanced training data: the audit-related variables have better predictive ability of audit quality than financial variables, and the combination of the two groups achieves the best performance, indicating that audit related variables and financial variables complement each other in predicting audit quality.

When all variables are included in inputs and when the misclassification cost is set to 10, Random Forest has higher value of AUC when it is trained on imbalanced data than on balanced data, and the same holds true when only financial variables or only audit-related variables are used as inputs. When all variables are used, the algorithm that has the best performance when trained on the synthetic balanced data underperforms the best algorithm trained on the imbalanced data, and the same holds true when either only audit-related variables or only financial variables are used. For example, when balanced training data and only financial variables are used, the highest value of AUC (0.619) is achieved by SVM, which is still lower than the highest value of AUC (0.638) achieved when unbalanced data and only financial variables are used. These results indicate that in this particular context of this research, the data imbalance is not an issue, and the performance of the algorithms trained on imbalanced data are even better than that of algorithms trained on the synthetic balanced data. Since the original imbalanced data is an authentic reflection of the real world, the results generated from the original data are also more reliable than those generated from the synthetic data.

Till now, the conclusions derived from the above analysis can be summarized as follows: 1) Supervised learning algorithms can be used to accurately predict audit quality, especially when Random Forest is applied on the original data without feature selection; 2) the most predictive variables are: the market share of the auditor, the log of client's total assets, the portfolio share of the auditor, log of audit fees, size of the auditor, and the brand name of the auditor; 3) audit related variables have better predictability than financial variables; and 4) audit related variables and financial variables complement each other in predicting audit quality.

5. Further Analysis

5.1 Compare Random Forest with Logistic regression

Since Random Forest performs extremely well in the previous analysis, it will be compared with logistic regression here to see which can better predict audit quality. Table 18 lists the performance indicators of the two algorithms under each condition.

The results show that Random Forest outperforms Logistic regression in each scenario in terms of AUC value, showing the superior performance of Random Forest in this particular context. One explanation for the poorer performance of Logistic Regression is that it is sensitive to the outliers in the dataset, however, the Random Forest algorithm is robust to outliers.

Table 18

Compare Random Forest with Logistic Regression (with original unbalanced training data and misclassification cost of 10)						
Input		**Algorithm**	**Recall**	**Specificity**	**AUC**	**Accuracy**
All Features	Without Feature Selection	Random Forest	0.734	0.712	0.723	0.713
		Logistic	0.573	0.762	0.667	0.750
	With Feature Selection	Random Forest	0.544	0.766	0.655	0.752
		Logistic	0.058	0.995	0.527	0.937
Non-audit related variables	Without Feature Selection	Random Forest	0.639	0.638	0.638	0.638
		Logistic	0.996	0.007	0.501	0.687
	With Feature Selection	Random Forest	0.556	0.658	0.607	0.651
		Logistic	0.000	1.000	0.500	0.937
Audit related variables	Without Feature Selection	Random Forest	0.627	0.738	0.682	0.73
		Logistic	0.556	0.771	0.663	0.757
	With Feature Selection	Random Forest	0.556	0.737	0.646	0.726
		Logistic	0.473	0.673	0.573	0.66

5.2 Exclude Accrual Variables from Audit Related Variables

In the previous analysis, the accrual variables: Total Net Accruals, Abnormal Accruals, and Absolute value of accruals are treated as both the audit related variables and the financial variables. To make sure that it is not these accrual variables that are causing bias in the results, they are now excluded from the audit related variables. To be compatible with the previous analysis, the Random Forest is used again and the results are listed in Table 19.

Table 19

Summary of performance when accruals are included/excluded from audit variables (with the original unbalanced training data and misclassification cost of 10)						
Input		Algorithm	Recall	Specificity	AUC	Accuracy
With Accruals	Without Feature Selection	Random Forest	0.639	0.704	0.671	0.70
	With Feature Selection	Random Forest	0.556	0.737	0.646	0.726
Excluding Accruals	Without Feature Selection	Random Forest	0.627	0.737	0.682	0.73
	With Feature Selection	Random Forest	0.556	0.737	0.646	0.726

The results show that excluding the accrual variables from the audit-related variables can boost the performance of the Radom Forest when no subset feature selection is performed. But the performance does not change when there is subset feature selection, because the same subset of feature is chosen before and after excluding accrual variables from audit-related variables.

5.3 Using Winsorized Data

In the previous analysis, outliers are kept in both the balanced and the imbalanced datasets. To see whether the outliers are causing bias in the results, the data winsorized at 1% are used in this further analysis. The untabulated results show that the performance of Random Forest is still better when using the original unwinsorized data. This may be because Random Forest is robust to outliers and eliminating outliers causes information loss.

6. Conclusion and Discussion

This research pioneers in constructing supervised learning algorithms that are more effective than traditional regressions to predict audit quality, which is proxied by restatement, one of the best publicly available measure of audit quality (Aobdia, 2015). Using 14,028 firm-year observations from 2008 to 2016 in the United States and ten different supervised learning algorithms, the research shows that: 1) supervised learning algorithms can be used to predict audit quality accurately, especially when Random Forest is applied; 2) the variables that are most predictive of audit quality are: the market share of the auditor, the log of client's total assets, the portfolio share of the auditor, log of audit fees, size of the auditor, and the brand name of the auditor; 3) audit related variables have higher predictive ability than financial variables; and 4) audit related variables and financial variables complement each other in predicting audit quality.

One major defect of this paper and most research that predicts restatement or fraud is that the financial variables used in the model are not the original version before restatement. This can be problematic because it means that research on audit failures is based on data that has already been restated to predict future restatements. Lack of access to the original financial data is the major reason why only the updated financial data is used for this paper and most of other related research. However, the conclusions that audit-related variables alone can predict audit quality very accurately and that they can predict better than financial variables already prove the reliability of the algorithms. Future research in the area of predicting audit quality might consider including other audit-related variables into the model, such as the textual information in audit reports, especially the Critical Audit Matters (CAMs). Another area for future research might be building more sophisticated algorithms to improve the ability to predict audit quality.

In a nutshell, from the conclusion of this research, all the stakeholders who wish to predict audit quality are suggested to use Random Forest algorithm which are trained with the original imbalanced data and with both audit-related variables and financial variables.

References

Abbott, L. J., Parker, S., and Peters, G. F. (2004). Audit Committee Characteristics and Restatements. *AUDITING: A Journal of Practice and Theory*, *23*(1), 69–87. https://doi.org/10.2308/aud.2004.23.1.69

Aier, J. K., Comprix, J., Gunlock, M. T., and Lee, D. (2005). The Financial Expertise of CFOs and Accounting Restatements. *Accounting Horizons*, *19*(3), 123–135. https://doi.org/10.2308/acch.2005.19.3.123

Alpaydin, E. (2014). *Introduction to Machine Learning* (3rd ed.). The MIT Press.

Aobdia, D. (2015). The Validity of Publicly Available Measures of Audit Quality -Evidence from the PCAOB Inspection Data. *Working Paper*, (June), 1–51. https://doi.org/10.2139/ssrn.2629305

Becker, C. L., Defond, M. L., Jiambalvo, J., and Subramanyam, K. R. (1998). The Effect of Audit Quality on Earnings Management. *Contemporary Accounting Research*, *15*(1), 1–24. https://doi.org/10.1111/j.1911-3846.1998.tb00547.x

Blankley, A. I., Hurtt, D. N., and MacGregor, J. E. (2012). Abnormal Audit Fees and Restatements. *Auditing : A Journal of Practice and Theory*, *31*(1), 79–96. https://doi.org/10.2308/ajpt-10210

Castle, N. (2018a). What is Semi-Supervised Learning? Retrieved July 9, 2018, from https://www.datascience.com/blog/what-is-semi-supervised-learning

Castle, N. (2018b). What is Semi-Supervised Learning?

Cecchini, M., Aytug, H., Koehler, G. J., and Pathak, P. (2010). Detecting Management Fraud in Public Companies. *Management Science*, *56*(7), 1146–1160. https://doi.org/10.1287/mnsc.1100.1174

Chen, H. J., Huang, S. Y., and Kuo, C. L. (2009). Using the artificial neural network to predict fraud litigation: Some empirical evidence from emerging markets. *Expert Systems with Applications*, *36*(2 PART 1), 1478–1484. https://doi.org/10.1016/j.eswa.2007.11.030

DeAngelo, L. E. (1981). Auditor Size and Audit Quality. *Journal of Accounting and Economics*, *3*(May), 183–199. https://doi.org/10.1016/0165-4101(81)90002-1

Dechow, P. M., Ge, W., Larson, C. R., and Sloan, R. G. (2011). Predicting Material Accounting Misstatements. *Contemporary Accounting Research*, *28*(1), 17–82. https://doi.org/10.1111/j.1911-3846.2010.01041.x

DeFond, M. L., and Francis, J. R. (2005). Audit Research after Sarbanes-Oxley. *Auditing : A Journal of Practice and Theory*, *24*(SUPPL.), 5–30. https://doi.org/10.2308/aud.2005.24.Supplement.5

DeFond, M., and Zhang, J. (2014). A Review of Archival Auditing Research. *Journal of Accounting and Economics*, *58*(2–3), 275–326. https://doi.org/10.1016/j.jacceco.2014.09.002

Deis, D. R., and Giroux, G. A. (1992). Determinants of Audit Quality in the Public Sector. *The*

Accounting Review, *67*(3), 462–479. https://doi.org/10.2307/247972

Dimmock, S. G., and Gerken, W. C. (2012). Predicting fraud by investment managers. *Journal of Financial Economics*, *105*(1), 153–173. https://doi.org/10.1016/j.jfineco.2012.01.002

Donges, N. (2018). The Random Forest Algorithm – Towards Data Science. Retrieved July 11, 2018, from https://towardsdatascience.com/the-random-forest-algorithm-d457d499ffcd

Dutta, I., Dutta, S., and Raahemi, B. (2017). Detecting financial restatements using data mining techniques. *Expert Systems with Applications*, *90*, 374–393. https://doi.org/10.1016/j.eswa.2017.08.030

Eshleman, J. D., and Guo, P. (2014). Do Big 4 Auditors Provide Higher Audit Quality After Controlling for the Endogenous Choice of Auditor? *Auditing : A Journal of Practice and Theory*, *33*(4), 197–220. https://doi.org/10.2308/ajpt-50792

Fawcett, T. (2016). Learning from Imbalanced Classes - Silicon Valley Data Science. Retrieved July 10, 2018, from https://www.svds.com/learning-imbalanced-classes/

Francis, J. R., and Yu, M. D. (2009). Big 4 Office Size and Audit Quality. *The Accounting Review*, *84*(5), 1521–1552. https://doi.org/10.2308/accr.2009.84.5.1521

Ghosh, A. (2005). Auditor Tenure and Perception of Audit Quality. *The Accounting Review*, *80*(2), 585–612. https://doi.org/10.2308/accr.2005.80.2.585

Kinney, W. R., Palmrose, Z.-V., and Scholz, S. (2004). Auditor Independence , Non-Audit Services , and Restatements: Was the U . S . Government Right? *Journal of Accounting Research*, *42*(3), 561–588. Retrieved from http://doi.wiley.com/10.1111/1475-679X.00088

Lennox, C. S., Wu, X., and Zhang, T. (2014). Does Mandatory Rotation of Audit Partners Improve Audit Quality? *The Accounting Review*, *89*(5), 1775–1803. https://doi.org/10.2308/accr-50800

Lin, C. C., Chiu, A. A., Huang, S. Y., and Yen, D. C. (2015). Detecting the financial statement fraud: The analysis of the differences between data mining techniques and experts' judgments. *Knowledge-Based Systems*, *89*, 459–470. https://doi.org/10.1016/j.knosys.2015.08.011

Murphy, K. (1998). A Brief Introduction to Graphical Models and Bayesian Networks. Retrieved July 11, 2018, from https://www.cs.ubc.ca/~murphyk/Bayes/bnintro.html

PCAOB. (2017). Auditing Standards of the Public Company Accounting Oversight Board.

Perols, J. (2011). Financial statement fraud detection: An analysis of statistical and machine learning algorithms. *Auditing : A Journal of Practice and Theory*, *30*(2), 19–50. https://doi.org/10.2308/ajpt-50009

Plumlee, M., and Yohn, T. L. (2010). An Analysis of the Underlying Causes Attributed to Restatements. *Accounting Horizons*, *24*(1), 41–64. https://doi.org/10.2308/acch.2010.24.1.41

Romanus, R. N., Maher, J. J., and Fleming, D. M. (2008). Auditor Industry Specialization, Auditor Changes, and Accounting Restatements. *Accounting Horizons*, *22*(4), 389–413.

https://doi.org/10.2308/acch.2008.22.4.389

Schmidt, J., and Wilkins, M. S. (2013). Bringing darkness to light: The influence of auditor quality and audit committee expertise on the timeliness of financial statement restatement disclosures. *Auditing : A Journal of Practice and Theory*, *32*(1), 221–244. https://doi.org/10.2308/ajpt-50307

scikit-learn. (n.d.). 1.10. Decision Trees — scikit-learn 0.19.1 documentation. Retrieved July 11, 2018, from http://scikit-learn.org/stable/modules/tree.html

Tape, T. G. (n.d.). The Area Under an ROC Curve. Retrieved July 11, 2018, from http://gim.unmc.edu/dxtests/roc3.htm

Zhang, H. (2004). The Optimality of Naive Bayes. *Proceedings of the Seventeenth International Florida Artificial Intelligence Research Society Conference FLAIRS 2004*, *1*(2), 1–6. https://doi.org/10.1016/j.patrec.2005.12.001

Appendix 1

Table 1

Variables that reflect/affect audit quality or reflect firm's innate risk		
Variables	**Description**	**Reference**
Abnormal Audit Fees	Residuals from the audit fee model	Blankley et al., 2012
LTA	Logarithm of end of year total assets	Blankley et al., 2012
LEV	Total debt divided by total assets	Blankley et al., 2012
FREEC	Demand for external financing, measured as the sum of cash from operations less average capital expenditures scaled by lagged total assets	Blankley et al., 2012
MATWEAK	1 if the client receives a material weakness opinion in the current year or the next year, 0 otherwise; This is the internal control indicator.	Blankley et al., 2012
Auditor Size	Measure of practice office size based on aggregated client audit fees (in $ millions) of a practice office in a specific fiscal year. In the multivariate tests, log of OFFICE (denoted InOFFICE) is used as the test variable and is based on actual fees (not rounded to millions)	Francis and Yu, 2009
INFLUENCE	Ratio of a specific client's total fees (audit fees plus non-audit fees) relative to aggregate annual fees generated by the practice office which audits the client	Francis and Yu, 2009
TENURE	Dummy variable that takes the value of 1 if auditor tenure is three years or less, and 0 otherwise	Francis and Yu, 2009
SALESGROWTH	One-year growth rate of a firm's sales revenue, and the maximum value is winsorized at 2	Francis and Yu, 2009
OCF	Operating cash flows deflated by lagged total assets	Francis and Yu, 2009
BANKRUPTCY	The Altman Z-score, which is a measure of the probability of bankruptcy, with a lower value indicating greater financial distress	Francis and Yu, 2009
VOLATILITY	Client's stock volatility and is the standard deviation of 12 monthly stock returns for the current fiscal year	Francis and Yu, 2009
MB	Log of book to market ratio	Francis and Yu, 2009
ACCRUALS	Signed abnormal accruals	Francis and Yu, 2009
ABS_ACCRUALS	Absolute value of abnormal accruals derived from the performance adjusted accruals model in Equation	Francis and Yu, 2009

SMALL_PROFIT	Dummy variable, and coded as 1 if a client's net income deflated by lagged total assets is between 0 and 5 percent, and 0 otherwise	Francis and Yu, 2009
SMALL_INCREASE	Dummy variable, and coded as 1 if a client's net income deflated by lagged total assets is between 0 and 1.3 percent, and 0 otherwise	Francis and Yu, 2009
GCREPORT	Dummy variable that takes the value of 1 if a firm receives a going-concern report in a specific fiscal year, and 0 otherwise	Francis and Yu, 2009
PRIORGC	Dummy variable that takes the value of 1 if a client received a going-concern report in the previous year, and 0 otherwise. In this paper, I set PRIOGC as 1 if it received GC in the past 3 years.	Francis and Yu, 2009
NON-SPEC	1 if a firm changed from a nonspecialist to an industry specialist, and 0 otherwise	Romanus, Maher, and Fleming, 2008
SPEC-NON	1 if a firm changed from an industry specialist to 3 nonspecialisls, and 0 otherwise	Romanus, Maher, and Fleming, 2008
NO-SPECCHG	1 if a firm changed from one industry specialist to another industry specialist, and 0 otherwise	Romanus, Maher, and Fleming, 2008
MSHARE	Auditor market share: auditor's total client sales in a particular industry divided by total industry sales	Romanus, Maher, and Fleming, 2008
AUDSPEC	Weighted auditor market share based on client sales (MSHARE * PSHARE; Neal and Riley 2004)	Romanus, Maher, and Fleming, 2008
FIN	Sum of additional cash raised from issuance of long-term debt (Compustat #9), common stock (Compustat #108) and preferred stock (Compustat #111) deflated by total assets (Compustat #6)	Romanus, Maher, and Fleming, 2008
ACC	Change in noncash working capital plus change in noncurrent operating assets plus change in net financial assets, scaled by total assets (Richardson et al. 2002)	Romanus, Maher, and Fleming, 2008
EXANTE	1 if firm's free cash flow is <-0.1, and 0 otherwise where free cash flow is net income (Compustat #172) less accruals divided by average of last three years capital expenditures (Compustat #128)	Romanus, Maher, and Fleming, 2008
EPSGWTH	Number of consecutive quarters of EPS growth for two years prior to restatement;	Romanus, Maher, and Fleming, 2008
CFOEXP	CFO’s years of work experience as CFO	Aier, Comprix, Gunlock, and Lee, 2005
CFOCPA	Dummy variable equal to 1 if the CFO has a CPA accreditation, 0 otherwise	Aier, Comprix, Gunlock, and Lee, 2005
INDEP	1 if all audit committee members are independent by BRC definition, else 0	Abbott, Parker, and Peters, 2004
EXPERT	1 if audit committee includes at least 1 director with financial expertise per the BRC’s definition, else 0	Abbott, Parker, and Peters, 2004
MINMEET	1 if audit committee meets at least four times annually during the sample year, else 0	Abbott, Parker, and Peters, 2004

BLOCK	The cumulative percentage of outstanding common stock shares held by 5 percent + blockholders not affiliated with management	Abbott, Parker, and Peters, 2004
BOARDSIZE	The number of directors on the board	Abbott, Parker, and Peters, 2004
AGEPUB	The number of years the company has been publicly traded	Abbott, Parker, and Peters, 2004

Appendix 2

Restatement Instances from the Audit Analytics

The restatement data provided by the Audit Analytics have the information of "Restatement Begin Date" and "Restatement End Date". The Audit Analytics confirmed via email that the beginning and ending dates of the restatement outline periods affected by the restatement. For example, restatement with beginning and ending dates of 01/01/2015 and 12/31/2017 affected years 2015, 2016, and 2017. And "it is possible (although rare) to have only certain years within the period affected to be restated". For example, in the example above, it is possible that a cash flow restatement would affect only 2015 and 2017, but not 2016. Based on the above information, the beginning and ending dates of the restatement for each firm are converted into firm-year observations using STATA (code provided on demand).

Appendix 3

Table 3

Independent variables and their calculations		
Independent Variables	**Description**	**Calculation**
LAF	Logarithm of audit fees	Log (Audit Fees)
DLAF	Difference of Log (Audit Fees)	Log (Audit Fees)$_t$ - Log (Audit Fees)$_{t-1}$
Auditor Size	Measure of practice office size based on aggregated client audit fees of a practice office in a specific fiscal year.	Log (Aggregated Audit Fees)
Big4	Whether the auditor is Big4	1 if the auditor is Big 4, and 0 otherwise
Auditor Change	Whether the client changed auditor	1 if the client changed the auditor in the current fiscal year, and 0 otherwise

INFLUENCE	Ratio of a specific client's audit fees relative to the aggregated audit fees generated by the practice office that audits the client	Audit Fees/Aggregated Audit Fees
TENURE	Measure the familiarity between the auditor and the client	1 if the auditor has been auditing the client for at least 3 years (Francis and Yu, 2009), and 0 otherwise
GC	Going concern opinion	1 if a firm receives a going-concern report in a specific fiscal year, and 0 otherwise
PRIORGC	Previous going concern opinion	1 if a client received a going-concern report in the previous 3 years, and 0 otherwise.
AuditorMarketShare	Auditor market share	Auditor's total client sales in a particular industry divided by total industry sales in a specific fiscal year
AuditorPortfolioShare	Auditor portfolio share	An auditor's client sales in each industry divided by the auditor's firm-wide client sales in a specific fiscal year
WeightedMarketShare	Weighted auditor market share based on client sales	AuditorMarketShare* AuditorPortfolioShare
Specialist	Whether the auditor is considered as a specialist	1 if the weighted market share meets the cutoff defined in Romanus, Maher, and Fleming (2008)
TotalNetAccurals	Total net accruals	Total Net Accruals = ΔAssets-ΔLiabilities - ΔCash
AbnormalAccurals	Abnormal accruals	Residual from the performance-adjusted accruals model in Francis and Yu (2009)
AbsAbnAcc	Absolute value of abnormal accruals	Derived from the performance-adjusted accruals model in Francis and Yu (2009)
LTA	Logarithm of year-end total assets	Log (Total Assets)
LEV1	Capital Structure	Total Debt/Total Assets
LEV2	Capital Structure	Total Liability/Total Assets
FREEC	Demand for external financing	(Cash from operations -average capital expenditures)/lagged total assets
SALESGROWTH	One-year growth rate of a firm's sales revenue	(Sales Revenue$_t$ - Sales Revenue$_{t-1}$)/Sales Revenue$_{t-1}$
OCF	Operating cash flows deflated by lagged total assets	Operating Cash Flow$_t$/Total Assets$_{t-1}$
BANKRUPTCY	The Altman Z-score, a measure of the probability of bankruptcy with a lower value indicating greater financial distress.	Z-Score = 1.2A + 1.4B + 3.3C + 0.6D + 1.0E Where: A = working capital / total assets B = retained earnings / total assets C = earnings before interest and tax / total assets

		D = market value of equity / total liabilities E = sales / total assets
BMratio	Book to Market ratio	Book Value/Market Value
SMALL_PROFIT	Whether or not the client has small profits	1 if a client's net income deflated by lagged total assets is between 0 and 5%, and 0 otherwise
SMALL_INCREASE	Whether or not the client has slight profits	1 if a client's net income deflated by lagged total assets is between 0 and 1.3%, and 0 otherwise
FIN	Financing	Sum of cash raised from issuance of long-term debt, common stock and preferred stock deflated by total assets
ACC	Change in noncash working capital plus change in noncurrent operating assets plus change in net financial assets, scaled by total assets (Richardson et al. 2002)	
EXANTE	1 if firm's free cash flow is <-0.1, and 0 otherwise, where free cash flow is net income less accruals divided by average of last three years capital expenditures	
EPSGWTH	Growth in EPS	1 if EPS has grown for 2 consecutive years, and 0 otherwise
materialweakness	Internal control indicator.	1 if the client receives a material weakness opinion in the current year, and 0 otherwise