Adı-Soyadı	Yasin Hakkı KOCAMAN
Sözleşme No.	TR2016/DG/04/A1-02/0550
Başvuru Yaptığı Sektör (Kamu-Üniversite-Özel Sektör)	Kamu
Başvuru Esnasında Bağlı Bulunulan Kurum	Sermaye Piyasası Kurulu
Başvuru Esnasında Bağlı Bulunulan Kurumdaki Unvan	Bilişim Uzmanı
Çalışma Alanı (AB Müktesebat Başlığı)	Mali Hizmetler
Öğrenim Görülen Ülke	Hollanda
Öğrenim Görülen Şehir	Amsterdam
Programın Öğretim Dili	İngilizce
Üniversite	Amsterdam Üniversitesi
Fakülte	İktisat ve İşletme Fakültesi
Bölüm	Uluslararası Finans
Program Adı	MSc. Uluslararası Finans : Finans ve Teknoloji Branşı
Programın Başlangıç/Bitiş Tarihleri (PDS belgesindeki tarihler)	02/09/2019 - 31/08/2020
Öğrenim Süresi (ay)	12
Tez/Araştırma Çalışmasının Başlığı	Finansal Yeniden Bildirme Raporlarının Makine Öğrenmesi ile Tespit Edilmesi
Danışmanının Adı/Soyadı	Torsten Jochem
Danışmanının E-posta Adres/leri	T.Jochem@uva.nl

Name/Surname	Yasin Hakkı KOCAMAN
Contract No.	TR2016/DG/04/A1-02/0550
Sector as of the Application Date (Public Sector-University-Private Sector)	Public Sector
Institution as of the Application Date	Capital Markets Board of Turkey
Title as of the Application Date	IT Specialist
Field of Study (i.e. EU Acquis Chapter)	Financial Services
Country of Host Institution	Netherlands
City of Host Institution	Amsterdam
Language of the Programme	English
Host Institution	Amsterdam University
Faculty	Amsterdam Business School
Department	International Finance
Name of the Programme	Master in International Finance (MIF) : Finance and Technology Track
Start/End Dates of the Programme (as in the PDS)	02/09/2019 - 31/08/2020
Duration of the Programme (months)	12
Title of the Dissertation/ Research Study	Restatement Detection Using Machine Learning
Name of the Advisor	Torsten Jochem
E-mail/s of the Advisor	T.Jochem@uva.nl

THESIS APPROVAL PAGE

I have examined the dissertation entitled

RESTATEMENT DETECTION USING MACHINE LEARNING

presented by

YASİN HAKKI KOCAMAN

and hereby certify that it is worthy of acceptance.

15/01/2021

Torsten JOCHEM

UNIVERSITY OF AMSTERDAM

AMSTERDAM BUSINESS SCHOOL

Table of Contents

ÖZET

Bu çalışma makine öğrenmesi yöntemlerini kullanarak 2010 yılı ile 2019 yılları arasında gerçekleştirilen finansal yeniden bildirim raporlarını tespit etmeyi hedeflemektedir. Bu amaca yönelik olarak, yeniden bildirim raporlarının hileli işlemden farklı oluşu ve rapor tipleri göz önünde bulundurularak birden çok farklı kaynak kullanılarak yeni bir veri seti oluşturulmuştur. Söz konusu veri setinin tümüne ve alt setlerine makine öğrenmesi algoritması uygulanarak sonuçlar elde edilmiştir. Daha önce gerçekleştirilmiş benzer çalışmalardan farklı olarak, firmalar tarafından bildirilen finansal veriler kullanılmamış, dar veri setinin tespit modeline etkisi incelenmiştir. Çalışma, öncelikle düzenleyici otoriteler olmak üzere, firmalar, denetçiler ve nihai olarak yatırımcılar için faydalı olma potansiyeline sahiptir.

Anahtar kelimeler: Makine Öğrenmesi, Finansal Yeniden Bildirim Raporları, Hileli İşlem Tespiti

ABSTRACT

This study aims to utilize machine learning methods to detect restatements between 2010 and 2019. To that end, a new database is created from three different data sources. As restatement is different than fraud, the effects of restatements are taken into account, as well as the type of the restatement, which could be a minor revision or major change. Different types of machine learning algorithms are applied to the full dataset and multiple subsets with varying results. Lastly, compared to the previous similar studies, firm stated financial variables are not included and effect of minimal data use is investigated in creating a detection model. The study has potential to be useful to various stakeholders in finance; primarily regulators, firms, auditors and ultimately investors.

Key words: Machine Learning, Restatement, Fraud Detection

UNIVERSITY of AMSTERDAM

AMSTERDAM BUSINESS SCHOOL

Master in International Finance

Master Thesis

Restatement Detection Using Machine Learning

Yasin Hakkı KOCAMAN

MIF Student

12945846

Under Supervision of

Prof. Dr. Torsten JOCHEM

January 2021

0. Statement of Originality

This document is written by Student Yasin Hakkı Kocaman who declares to take full responsibility for the contents of this document.

I declare that the text and the work presented in this document are original and that no sources other than those mentioned in the text and its references have been used in creating it.

The Faculty of Economics and Business is responsible solely for the supervision of completion of the work, not for the contents.

1. Introduction

Machine learning is becoming ubiquitous in financial research in recent years with developments in both technology and data mining methods. As more and more data becomes available for researchers and practitioners, novel methods are developed and difference in approaches is investigated.

There are various insights to be gained from machine learning methods when applied to different datasets, even helping us uncover some relationships that is not readily apparent or quite easily investigated. Leveraging the power of pattern recognition via machine learning could help us explore some new models to investigate new theories. These insights gained from machine learning could be of interest to a wide array of parties, including academics, firms, traders as well as regulatory bodies. Which, in turn, could be incorporated to the decision making process, such as risk based audit by regulators.

With the technological advancements in the recent years, vast amounts of data, referred to as big data which generally has the generally accepted characteristics of volume, variety, veracity, value and velocity, made exploration of complex pattern easier. This so called *new oil* enables growing datasets to be scrutinized to benefit academic research uncovering new understandings. Secondly, there is a potential benefit to regulators, as this is a relatively inexpensive source of information that could be incorporated to early warning systems or risk based oversight mechanisms. Furthermore, firms and traders could use this as a part of decision making process, to explore new profit avenues. Lastly, any interested party could leverage the power of big data with machine learning to answer interesting and sometimes yet to be asked questions.

There are various financial data sources which can be used for financial analysis. One such source is restatements, which are filing of changes to previously filed independent audit reports. These can range from small to significantly large and could have either a positive or negative effect to the bottom-line of the companies. As, they can be result of weak audits, clerical errors or even fraud, it is possible that analysis of restatements could shed light to not immediately available underpinnings

In this paper, restatement analysis is conducted, utilizing machine learning methodologies. Various machine learning methods including basic and gradient boosted regression are applied to the restatement data that is created for this study from 3 different data sources to recognize patterns and detect anomalies. This will shed light into this area of research and could be especially useful to the regulators to incorporate into their early warning systems or risk based audits. The central question is whether there is a substantial relationship between selected variables and likelihood of restatements, followed with the investigation into the predictive power of various machine learning methodologies within the context of data at hand.

While there exists various fraud related research in economics, this study differs in a few ways from the prior literature. Firstly, this study does not depend on the previously established datasets, but creates its own through various different data vendors/streams and uses different dataset. Secondly this study investigates effect of restatement type, comparing "little r" and "Big R" as well as taking into consideration how the restatement effects firm value to bridge the estimation to fraud related research: using the effect of restatement to gauge whether a restatement is a result of intentional misstatement or unintentional error. Lastly, this study does not just utilize one method of machine learning but compares different types of machine learning algorithms with the created dataset, comparing their effectiveness.

The paper is structured as follows: next chapter gives background information on machine learning related economics research in the context of fraud in general followed with restatement related studies. In institutional background chapter, restatement research is discussed in the context of various stakeholders, primarily regulators. Methodology and Data chapter gives a cursory introduction to machine learning and discusses specifics of the data. Results chapter gives details about the experimentation and the outcomes. Finally, last chapter gives an overview and debates some shortcomings and future work.

2. Literature Review

Machine learning has been utilized extensively in many areas of research. Within the context of financial research, it is used to both confirm theoretical underpinnings and find new patterns in data, benefiting various stakeholders including regulators. One of the datasets that is being investigated is related to restatements; which are financial statements that have been disclosed in various SEC filings. They are regarded as an indicator of audit and reporting quality (Hayes and Boritz, 2019)

While most of the research utilizing machine learning are focused on the fraud related research, restatements are not as extensively investigated: Even though misstatements are different from fraud and classified as irregularities, they are not necessarily a result of fraud and can be related to accounting/clerical/SEC and generally a result of weak audit procedures (Lobo and Zhao, 2013)

Some restatements also have multiple underlying reasons, out of the generally reported types of accounting rule application failure, financial fraud, irregularities and misrepresentations, errors in accounting and clerical applications, SEC involvement (such as comment letter that triggered the restatement; or formal or informal SEC inquiry into the circumstances surrounding the restatement) or any other issues with particular importance. As such, they lack some qualities fraud-type data inherently possesses.

Restatements could be type of either[1] immaterial revisions (little r) or of material concern (Big R) to the financial report quality, and either could be effecting the previous report as good (i.e. improves financial standing) or bad, differing from (most) fraud. Tan and Young (2014) show through univariate testing that "little r" firms do not significantly differ from non-revising firms in their sample between 2009 and 2012, also showing the effect to the report does not change their findings, hence providing experimental proof to the idea that "Big R" firms should be given more importance in restatement-related machine learning

[1] There are also out-of-period adjustments to take note of, which are errors in previous financial statements that do not significantly affect and would not require restatement of either type.

research. “Big R” statements are also more extensively investigated in various studies, such as Hribar and Jenkins (2004), in which negative effects of restatements are shown.

Another point of interest is whether restatements are result of intentional misstatement (i.e. fraud-like) or correction of unintentional error. These could be attributed to management integrity or auditor independence. However, while most of the data does not provide a distinction between two, as it is not outright clear what type the restatement falls under. The language used in restatements have been analysed via machine learning methodologies (Hennes et al, 2008) and shown that there is a distinction between unintentional error and intentional misstatement in the selection of words in the announcements.

Within the literature there is no clear-cut definition of fraud. Karpoff, Koewster, Lee and Martin (2017) found that there is a differing view on fraud in different databases, and this is another reason why some classifications on intentional misstatement proportion differs. (Hayes and Boritz, 2019). While Hennes et al. (2008) claim that nearly one quarter of restatements are fraud and irregularity, Plumlee and Yohn (2010) claim that only 3% of restatements they analysed classified as manipulation. As such, data quality is part of a broader issue within the context of fraud research.

While restatements are not necessarily fraud as discussed above, there are many similarities to the fraud research and are included in the fraud related financial research. Also, most of the studies utilize similar or same machine learning methods to the datasets. Bao, Ke, Li, Yu and Zhang (2019) is a recent study on accounting fraud detection using machine learning, in which ensemble learning method is investigated and some novel performance evaluation methods are introduced. The study also opts to use raw financial data instead of long-established financial ratios to better utilize the power of machine learning, shedding pre-conceptions about the interaction of different accounting data.

Most of the accounting related fraud research, however uses accounting ratios such as Dechow, Ge, Larson and Sloan (2011) which is generally regarded as the most comprehensive fraud prediction, in which financial characteristics of misstating firms are analysed and a model to predict misstatements is developed after creating a database from the AAER releases between 1982 and 2005 utilizing logistic regression for detection model development. The ratios used are grouped as accruals quality related variables, performance variables, nonfinancial variables, off-balance-sheet variables and market-related incentives which are calculated against a binary indicator variable misstatement firm-years.

Another benchmark study is conducted by Cecchini, Aytug, Koehler and Pathak (2010) which not only uses financial ratios, but also lets the machine learning methods to determine new ratios via the support vector machines for fraud detection.

With regards to the restatement focus, a recent study by Bertomeu, Cheynel, Floyd and Pan (2019) utilizes restatement data along with Accounting and Auditing Enforcement Releases (AAERs) issued by the SEC to detect accounting misstatements. The study focuses on gradient boosted regression tree method, developed by Friedman (2001) and Hastie, Tibshirani and Friedman (2001) to detect and interpret patterns present in ongoing accounting misstatements. The variables used include accounting, capital markets, governance, and auditing datasets, showing that while accounting variables do not have much explanatory power by themselves, coupling audit and market variables provides better prediction scores.

While it is shown that there is a relationship between issuance of restatements with financial ratios (Dechow, Ge, Larson and Sloan, 2011), there is no exact variable (either as a ratio or a raw value) that definitely points to restatement. As restatements encompasses both intentional misstatement and unintentional error, they can’t be handled exactly like a fraud research either. Another problem with restatement research is rarity of events compared to *normal* cases.

The previous studies show that the restatement can be correctly classified if various accounting variables are investigated and predictor power of the models are increased by adding some audit and governance related variables. Also, it is shown that there is a relationship between restatements and selected features of the study, namely: short interest, trading volume, price spread and auditor change.

It has been shown that there is a difference in power of prediction between different machine learning methodologies. Bao, Ke, Li, Yu and Zhang (2019) shows that new methods such as ensemble learning outperforms logistic regression benchmark model by Dechow et al.’s

(2011) and support-vector-machine of Cecchini et al. (2010) Another hindrance in front of reliable prediction, as posited by previous studies numerously, is data imbalance. In fraud related research, detection in which the number of fraud cases are very little compared to the non-fraud cases necessitates new methods such as creating different subsets to alleviate issues stemming from data imbalance.

Previous fraud and restatement studies used primarily SEC's Accounting and Auditing Enforcement Releases (AAERs) provided by the University of California-Berkeley Center for Financial Reporting and Management (CFRM). The other popular databases are the Government Accountability Office's (GAO) earnings restatement database and the Stanford Securities Class Action Clearinghouse (SCAC) database of securities class action lawsuits filed since the passage of the Private Securities (Bao, Ke, Li, Yu and Zhang, 2019). This study uses the Audit Analytics' (AA) earnings restatement database, which is not utilized as extensively as AAER or other databases. The study does not depend on the AA database solely, but also incorporates CRSP and Compustat databases to include price spread and trade volume from the former and short interest from the latter. Selected variables also reflect another important point; they do not rely on the financial reporting of the firms, but are from the market: Short interest, Trading Volume, Price Spread and Auditor Change. Secondly, this study investigates the effect of little r and Big R, as the distinction has repercussions unique to the restatements as not all restatements are result of a fraud. Similar to above point, the distinction between intentional misstatement or unintentional error is not clear-cut, and hence the effect of restatement is taken into consideration when conducting experiments to include only adverse effecting restatement as one of the test groups. Finally, this study utilizes different methods of machine learning to compare their predictive power.

3. Institutional background

Financial restatements have been a major concern for the regulators, investors and market participants and existing literature suggests that data mining techniques have been widely used to detect financial fraudulent activities by the practitioners and regulators. (Dutta, Dutta and Raahemi, 2017) In particular, regulators and policymakers can pay a close attention to the suspected firms and investors can take actions in advance to reduce their investment risks.

While machine learning is no panacea for detection, as there is still a big gap between detection rates by various algorithms and financial data, it still provides a meaningful pointer into the regulatory action, either preventive or detective. In ideal conditions, there is a bottleneck for regulators that are defined by resources at hand and machine learning could help extend endeavours beyond current limits. The results of detection models would help audit planning decisions (Perols, Bowen, Zimmermann and Samba, 2015) and potential investigations (Walter, 2013)

The insights gained would be of immediate use, especially to decide what to include or exclude in the detection models to identify restatements. The machine learning methods are still in need of reputable and trustworthy source of data to feed the algorithms, but that is also becoming easier and or cheaper to acquire with the digitalization of the information. Still, it is likely that there exists some wisdom not tapped into, in the type of tribal/institutional knowledge that relies on human part, which could not be easily digitized, if ever. Still, the amount that is available should be more than enough to create early detection and prevention.

While it is impractical for regulators or corporate monitors to investigate all predicted cases of fraud, given the limited resources available to fight such fraud (Ernst & Young, 2010), the predicted likelihood of the fraud provides an important and simple decision rule for regulatory functions on the basis of suspicion (Bao, Ke, Li, Yu and Zhang, 2019)

As the restatements could be result of unintentional errors or intentional misstatements, the former generally necessitates sanctions by the regulatory agencies while the latter is generally an issue of audit quality. Martin (2013) also states that restatement disclosures could be indicative of financial reporting and audit quality. Hence while the focus is generally on regulatory oversight, these methods can also be utilized by auditors as a secondary confirmation to their findings and ultimately to prevent reputation loss as some of restatements are attributed to auditor change. Regulators also respond differently to these two distinct situations (Hayes and Boritz, 2019)

It is also important for the shareholders, as it is known that restatements are mostly wealth reducing events eventually hurting firm reputation. Hence detection provides benefit to the investors ultimately.

Machine learning either should confirm already established economic theories, or should provide results that are interpretable. As such, while some insights gained could be in line with the theories and proven facts of the academic studies, the explanation of the results that are arrived via machine learning methods sometimes could not be easily attributed to the current theories. This topic is discussed deeply under the name *explainable AI*, and much debate is going on to turn black-box structure of machine learning and artificial intelligence. This is even more critical, if unexpected results are to be interpreted correctly and insights to be gained accurately.

Another shortcoming of machine learning in the context of fraud related research is Type I vs. Type II detection, whereas Type I is defined as classification of fraud firms as non-fraud and Type II is defined as classification of non-fraud firms as fraud firms. (Beneish, 1999) As such, for regulatory purposes, Type I errors are costlier compared to Type II errors and models should be selected to compensate this inherent situation in detection methods. It is also established that resource-constrained regulators value minimal false positives (Cecchini, Aytug, Koehler and Pathak, 2010)

Similar to doping in sports, situation resembles a dove-hawk problem, as both the data is dynamic, always changing according to the advancements in detection, fraud will also change the methods to avoid detection, which becomes more pronounced problem when coupled with *regulatory lag*. There are shifts in market, regulatory challenges, changes in both regulation and focus, and finally various crises; economic, technological, health-related, natural etc. that effect both the data and effectiveness of methods.

No machine learning method could be single source for investigation effort, but it would definitely be useful if applied with the expertise and judgement to alleviate scarce investigative resources. It is believed that more research into machine learning in the context of detection would benefit all of the market participants; regulators, auditors, firms and ultimately investors.

4. Data and Methodology

Anomaly detection is a vital task, with various application in numerous domains including finance and law enforcement. Interest in machine learning in social sciences is not without reason and it is apparent that advents in the domain provided better results with the help of vast amount of data availability. Machine learning methods used for this kind of tasks can be classified under different categories such as supervised vs. unsupervised machine learning. Problems are data quality, scalability, effectiveness etc. are some aspects of the real world applications that necessitate further scrutiny. (Akoglu and Faloutsos, 2013)

The specific branch of rare event discovery is called anomaly (or abnormality) detection, under which fraud detection can be listed. As with many similar algorithms, outlier handling, correlations, correct display of data which lies on multi-dimensions, scale and dynamic nature are also of importance.

While machine learning encompasses a broad range of methodologies ranging from basic regression to neural networks, the approach used in the context of fraud (and restatement) research falls under the category of binary classification problem.

However, with regards to the restatement, the cases also include non-fraud cases, as misstatement can be caused by different reasons. As such, there is a distinction between intentional manipulations (prominently financial fraud) and unintentional errors (audit/clerical/accounting mistakes).

The data needed for the experimentation is restatement data, structured as firm-years in which the restatement start date taken as the year value. If multiple restatements happen in the same year, they are regarded as a single instance. For the study, all of the financial restatement data between 2000-2020 is collected from the Audit Analytics database, including firm details (ticker, name, gvkey etc.), type of restatement (accounting, fraud, clerical error or other issues), date details (beginning date, ending date), auditor details (beginning auditor, ending auditor, restatement auditor), effect of the restatement (improves or worsen financial statement), little r – Big R distinction (DATE_OF_8K_402, blank for revision (little r) restatements).

The financial information is gathered from both Compustat database and CRSP database and collated with the restatement data. For the variables, firm reported financial data is excluded to investigate whether lack of data from the firm would be reasonable for a restatement detection system. As such, included variables are all market variables; short interest, price spread, trade volume and lastly auditor change.

The first variable included is short interest, which is number of shares short sold in proportion to the total number of shares outstanding, which displays the anticipation of the investors. As such, it is a likely indicator of misstatement anticipation. (Boehmer, Huszár, Jordan, 2009)

Foremost indicator of a firm value is its stock price, which is also a very public information with fast alignment to market news. As such, the change is reflective of market sentiment and is included in the study as a variable.

In a similar vein to the price spread, trade volume is an indication of investor interest and is a public value. Trade volume is also indicative of fraud in some cases (citation needed), and is included in the study as a restatement detection feature.

Both of the above values are collected as monthly values and turned into yearly average values.

Auditor change is included. As presented in Keune and Johnstone (2012); given that auditors are able to detect even relatively small misstatements good audits should issue restatements. Therefore, auditor change is of importance in issuance of restatements and is included in the restatement the variables.

In the study, relationship between the variables listed and chance of restatement is investigated. While theoretical background confirms the relationship with the variables selected, it is not clear whether they can be used to detect restatements. As such, various machine learning algorithms is tested.

Distinguishing restatement (or fraud) from no restatement case is a classification problem, based on the variables used in the right hand side of the regression equation. As such the function

$$f(x) = R^{v} \rightarrow \{0,1\}$$

is such that features (R) is mapped to either class 0 (no restatement) or class 1 (restatement). This is similar to a fraud detection functions and also a subclass of multivariate classifications.

While previous studies use the same predictors that were found to be significant (Perols, 2011; Lin et al., 2003), in this study, financial variables are not included and as such has far less amount of variables/predictors in the structure.

In most of the machine learning algorithms, the sample is divided into two: a training set in which the algorithm is trained and a validation sample used to evaluate the prediction ability of the model. In this study the division is done randomly, and look-ahead bias is ignored.

Linear regression

In simple linear regression, relationship between the independent variables and dependent variable is modelled as a linear equation:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots$$

Logistic regression

Logistic regression is used for events fitting to a logistic curve (instead of a linear equation) The logistic function takes any value and returns categorical output, that is fitting for classification problems. As the problem is a detection problem of binary nature (fraud/non-fraud), binary logistic regression is used. The formula is as follows:

$$f(x) = \begin{cases} 1, & if \ \Pr(x) \geq 0.5 \\ 0, & otherwise \end{cases}$$ where Pr(x)= $\frac{1}{1+e^{-\beta x}}$

Gradient Boosted Regression Trees

The gradient boosted regression tree is utilized by Bertomeu, Cheynel, Floyd and Pan (2019) in their study and is used in this study as well. GBRT is done by dividing the outcomes into regions (trees) and then iteratively fitting to improve accuracy (boosting).

As building blocks of GBRT is out of scope for this study, details are left out, reviews of which can be found in Friedman (2001), Guelman (2012), Zhang and Haghani (2015) and Kleinberg, Lakkaraju, Leskovec, Ludwig and Mullainathan (2017).

5. Descriptive Statistics

The data used is from Audit Analytics database, updated in 15th of November, 2020. The database tracks both immaterial revisions "little r" and material 8K 4.02 non-reliance restatements made since August of 2004 and while the restatement years goes as far back as 1984, most of the restatements are made in the recent years, with a trend towards decrease in the numbers. For the study, the restatements between year 2010 and 2019 is used. While restatement is a lagging type of data, most of the restatement filed happens within two years. The following table (Table 1) displays the restatement numbers that totals to 17029, which are issued for the years between 2000 and 2019, last 20 years:

Table 1 Restatement distribution by year

Year	**# of Restatements**	**Percentage**	**Year**	**# of Restatements**	**Percentage**
2000	924	5.43%	**2010**	892	5.24%
2001	1090	6.40%	**2011**	849	4.99%
2002	1139	6.69%	**2012**	835	4.90%
2003	1249	7.33%	**2013**	805	4.73%
2004	1354	7.95%	**2014**	723	4.25%
2005	1251	7.35%	**2015**	642	3.77%
2006	1095	6.43%	**2016**	471	2.77%
2007	938	5.51%	**2017**	471	2.77%
2008	870	5.11%	**2018**	355	2.08%

2009	925	5.43%	**2019**	151	0.89%
Total	17029	100.00%			

This table displays the percentages of restatements between year 2000 and 2019. A clear trend downwards in the number of cases can be observed. The lag in restatement is prominent in year 2019 as not all of the restatements are issues as of data date of 15.11.2020

The restatement types are divided into four groups, most of the cases being classified as accounting type restatement. Fraud accounts to less than 1% of the total cases between 2000 and 2019, as listed in Table 2.

Table 2 Restatement distribution by type

Type	Cases	Percentage
Accounting	16298	95.71%
Fraud	160	0.94%
Clerical	520	3.05%
Other	51	0.30%
Total	17029	100.00%

This table displays the percentages of restatements types between year 2000 and 2019. Most of the cases are of accounting variety, amounting more than 95% of the cases.

The effect of restatement is generally of that of favourable, and 85% percent of the cases improves the financial statement that is restated. As such, when adverse effect is thought more akin to the fraud cases for machine learning purposes, the data imbalance problem is overemphasized. Table 3 displays the distribution.

Table 3 Restatement distribution by effect

Effect	Cases	Percentage
Improves	14475	85.00%
Adverse	2552	14.99%
Neither	2	0.01%
Total	17029	100.00%

This table displays the percentages of restatements effect to financial statements between year 2000 and 2019. Most of the cases improves previous statements, which leaves even less data for adverse effect to analyse in the context of fraud.

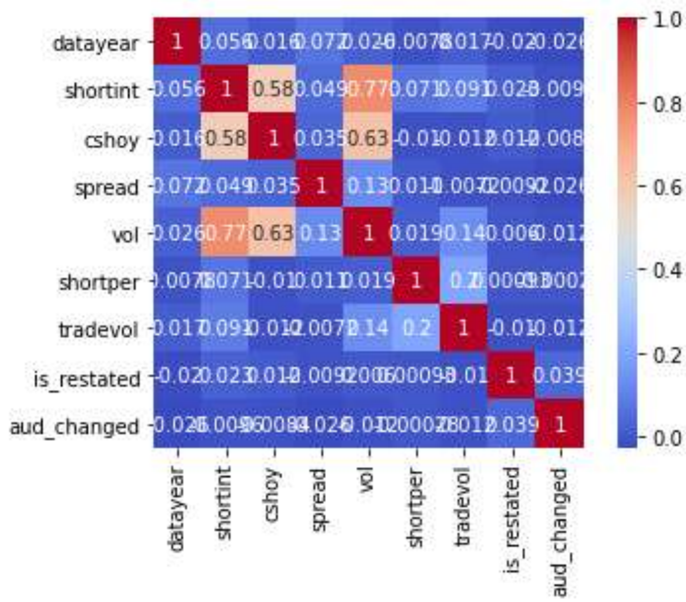
The distinction between little r and Big R is done via inclusion of date variable in date of 8K 402, which is left blank in case of little r adjustment to the financial statements. While there is a trend of increased usage of little r, the percentage distribution amount does not change much from year to year as presented in Table 4.

Table 4 Restatement distribution by revision type

Revision	Cases	Percentage
Big R	5519	32.41%
little r	11510	67.59%
Total	17029	100.00%

This table displays the percentages of restatements revisions between year 2000 and 2019. One third of the cases are of Big R variety, which are revision of material concern (contrast to little r, which are immaterial revisions).

Figure 1 Correlation Matrix



The relationships between the variables, in the correlation is shown in Figure 1, correlation matrix. There is a clear correlation between short interest (shortint) and trade volume (vol), which is expected as some percentage of all trades in a certain stock is short selling (unless prohibited). The same can be said for the common shares outstanding (cshoy) as with more stock more trades are conducted. While tradevol variable, which is found by dividing volume by common shares outstanding, could have a different explanatory power as it presents percentage of trades. Lastly, as expected from the relationship of trade volume with both common shares outstanding and short interest, there is a correlation between common shares outstanding and short interest.

6. Experiment Results

The original dataset was created using restatement data between 2000 and 2019, but only the data from 2010 onwards are selected for machine learning parts to alleviate problems stemming from how little r is treated by firms due to regulation changes and some other minor effects. However, this division could be used to create two separate training/test subsets in further studies, especially to account for look ahead biases, instead of random splits, which is used in this study.

From 2010 onwards a total of 63037 non-restatement and 1731 restatement cases are counted across 10348 different firms. Summary statistics for the dataset is as follows in Table 5:

Table 5 Summary statistics of dataset between years 2010-2019

Dep. Variable:	is_restated	R-squared (uncentered):	0.006
Model:	OLS	Adj. R-squared (uncentered):	0.006
Method:	Least Squares	F-statistic:	102.8
Date:	Sat, 12 Dec 2021	Prob (F-statistic):	1.91e-87
Time:	00:01:35	Log-Likelihood:	25602.
No. Observations:	64768	AIC:	-5.120e+04
Df Residuals:	64764	BIC:	-5.116e+04
Df Model:	4		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
shortper	4.522e-05	4.64e-06	9.746	0.000	3.61e-05	5.43e-05
spread	0.0011	0.001	1.071	0.284	-0.001	0.003
tradevol	7.707e-05	5.79e-05	1.330	0.183	-3.65e-05	0.000
aud_changed	0.0638	0.004	16.035	0.000	0.056	0.072

==

Omnibus:	69708.208	Durbin-Watson:	1.949
Prob(Omnibus):	0.000	Jarque-Bera (JB):	3144785.858
Skew:	5.813	Prob(JB):	0.00
Kurtosis:	35.096	Cond. No.	956.

==

In the feature engineering parts, the duplicate values are removed and data is winsorized to minimize outliers' effect. For the yearly values, an average of 12 months is taken as the final value for each firm-year case. Infinite values, division-by-zero errors and nan values are treated, creating a balanced dataset.

Multiple experimentations are conducted with different machine learning methods using various hyperparameters to uncover relationship between select variables and restatement issuance. To account for restatement types and effects, as well as data imbalance problem, different subsets are created and experimented on, detailed in the following parts:

6.1. Full dataset detection

Using the whole dataset (2010-2019), division into training and test subsets, to %75 to %25 of the whole set is administered respectively. The result is a regression score of 0.003 and the model was not able to classify any of the frauds correctly. Changing hyper parameters (division sizes and random seed value which decides how the data is divided) does not change the outcome.

The resulting formula was;

$y = 0.025 + 0.00002 * x1 + -0.00135 * x2 + -0.00021 x3 + 0.05024 x4$

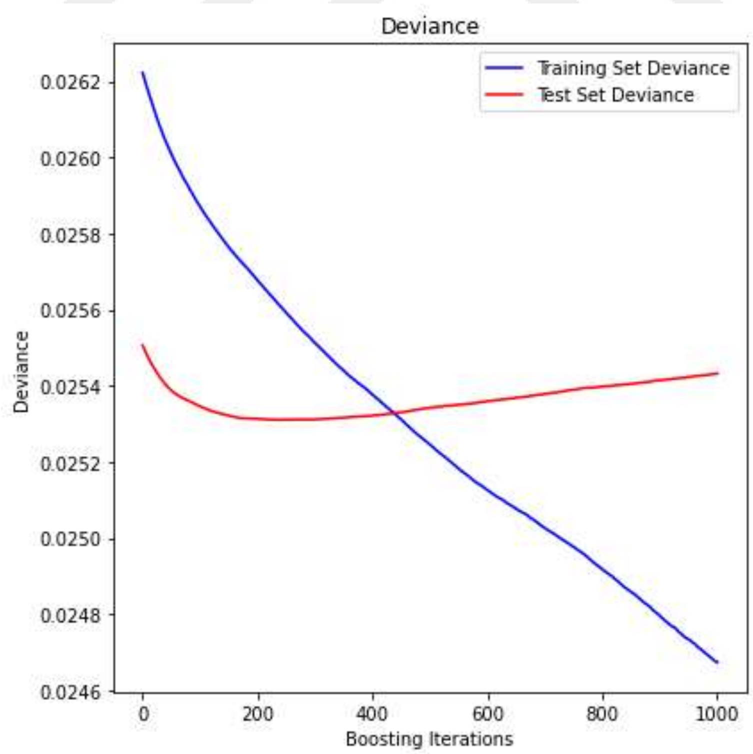
x1, x2, x3 and x4 being short interest (percentage), price spread, trading volume (percentage) and auditor change (binary) respectively

Using cross validation technique (cv=10) to the same dataset, the average CV score is found to be -0.002, which reveals the same inconclusive results as above.

Using K fold method of cross validation, throughout multiple iterations, the correct restatement detection and classification stays at 0.

Using gradient boosted regression method and performing 1000 boosting stages, the mean squared error on test set results with 0.0254, revealing not much difference between training and test subsets. However, even though the MSE is close to zero, detection of fraud is low, as, like most of the fraud detection models, there is a clear imbalance between non-restatement cases and restatement cases. The model classifies overrepresented non-restatement cases and classifies accordingly which results with low MSE value. This problem is addressed in the next part.

Figure 2 GBRT results with full dataset



This figure displays the loss function for each iteration, comparing test and training subsets. While the training test deviance decreases as the number of iterations increase as expected, overtraining the model with more than 200 iterations show diminishing return for detection on test subset.

6.2. Randomized undersampling detection

One of the greatest challenges in fraud detection is related to fraud observation rarity (Perols, Bowen, Zimmermann and Samba, 2015), and to overcome this problem several methods are employed in literature.

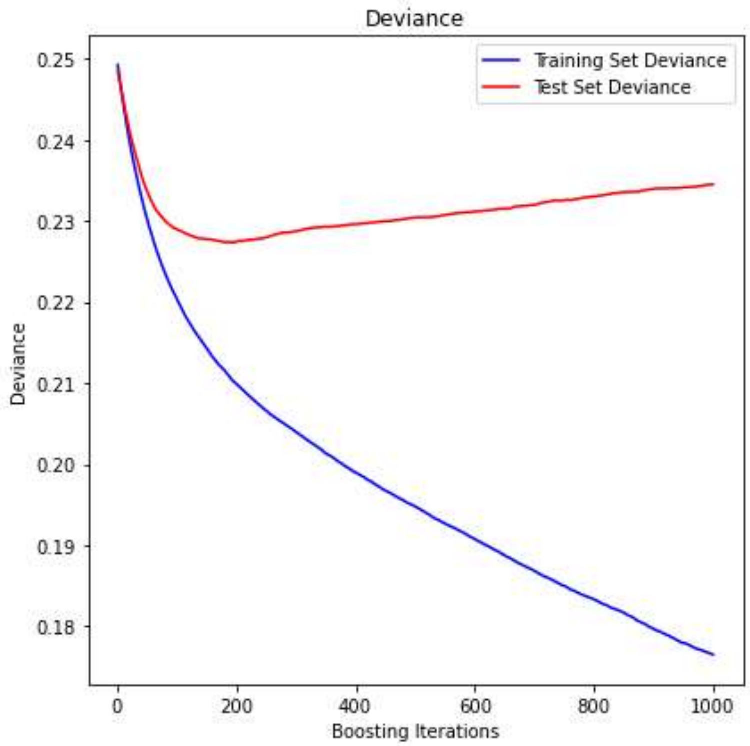
In this study, multi-subset observation undersampling (OU) is utilized, addressing imbalance issue by creating multiple subsets which include all restatement cases and equal number of non-restatement cases selected at random. One of the results is as follows:

$$y = 0.481 + 0.00106 * x1 + -0.05337 * x2 + -0.00705\ x3 + 0.16394\ x4$$

with a score of regression of 0.024 and correct classification of 58% (488/828) and correct restatement classification of 51% (237/463) on an example with low CV score. As the subset is random, while each iteration differs from each other, the mean value reverts to the 50%, not much different than chance.

Gradient boosted regression method provides similar results and performing 1000 boosting stages, the mean stays flat as there is limited number of cases and not much variation is possible after 200 iterations, which goes on to show that GBRT is feasible for larger datasets.

Figure 3 GBRT results with undersampling subset



This figure displays the loss function for each iteration, comparing test and training subsets. Compared to the previous results, loss function has increased, with test set showing similar curve.

6.3. Adverse effect subset detection

As discussed in previous chapters, restatements differ from the fraud by the way of not always affecting financial statement negatively. To investigate this particular effect, only the restatements that affect financial restatements adversely are included in the detection model. No discernible detection is reached with this subset configuration as well. The formula is as follows:

y = 0.018 + 0.00012 * x1 + -0.00148 * x2 + -0.00023 x3 + 0.03650 x4 while the CV is very close zero in most of the iterations, averaging zero ultimately. As for GBRT, the same problems persist.

6.4. Big R subset detection

Similar to the previous part, not all restatements are filed under the same type, a distinction is made that divides restatement into two; immaterial revisions – little r and material restatement with material concern – Big R. As some studies suggest little r firms do not significantly differ from non-revising firms, greater scrutiny could focus on Big R only. As such, a new subset is created with Big R firms.

With this configuration, the scarcity of restatement makes imbalance problem even more pronounced, as Big R amounts to less than 30% of the all restatements, resulting with a formula of

$$y = 0.004 + 0.00001 * x1 + 0.00042 * x2 + -0.00004\ x3 + 0.01487\ x4$$

with correct classification for most of the cases (0.9957) and zero correct restatement classification. Cross validation does not provide better detection rates as well.

7. Conclusion

In this paper, restatement analysis is conducted, using multiple machine learning algorithms including basic and gradient boosted regression, applied to the dataset created for the study from multiple data sources. As the restatement is different from financial fraud, different types of restatement is taken into consideration; little r and Big R, and each case is evaluated individually. For detection, firm provided financial data is not used and only market variables are utilized in creation of the models, including spread, trade volume, auditor change and short interest.

Firstly, background information on machine learning related economics research in the context of fraud detection in general and restatements in particular is discussed. Institutional background in this research are is examined. Following theoretical debates, methodology, data and results chapters gives details about machine learning, data structure, experimentation and the outcomes.

The results suggest that with the current structure of the data, parameters and hyper-parameters, it is not possible to create a reliable detection model, even with different detection algorithms. However, the inconclusive results are not incompatible with the theoretical background discussed in the literature review.

One of the possible reasons could be because how the data is structured with regards to the period length; the entries were all firm-years, which, for a reliable detection might not provide enough of a fine grain. Further test could be conducted in which shorter periods with more frequent data, such as monthly values, are used. Another possible factor related to the data itself is class imbalance problem, as the cases of restatements are far less than non-restatement cases, and similar to fraud detection, so called needle and the haystack problem is prominent in restatement detection.

Intentional exclusion of financial variables in the model, with which some level of reliable detection has been achieved in the past, could also be quoted as one of the reason with the inconclusive results, amplifying class imbalance problem with data/variable scarcity.

Lastly, while restatements are similar to fraud, not all restatements effect stock price negatively and not all restatements follow short interest and auditor change and other variables included in the detection model. The inherent characteristics of short interest also could be a contributing factor.

While this study was not conclusive with regards to the detection, with the inclusion of more frequent data higher number of variables, a substantial detection model could be created as evidenced in prior studies and would be useful to the regulators to incorporate into their early warning systems or risk based audits. With the advancement in both machine learning and big data, new predictive models will emerge. The findings of this study may help with develop new methods and what not to include/exclude and pitfalls that might cause unreliable detection. However, as new methods emerge, detection will be harder as the cover-up of fraud will also incorporate new methods.

8. References

Akoglu, L., Faloutsos, C., Anomaly, Event, and Fraud Detection in Large Network Datasets, 2013

Bao, Yang and Ke, Bin and Li, Bin and Yu, Yingri Julia and Zhang, Jie, Detecting Accounting Fraud in Publicly Traded U.S. Firms Using a Machine Learning Approach (October 30, 2019). Available at SSRN: https://ssrn.com/abstract=2670703 or http://dx.doi.org/10.2139/ssrn.2670703

Beneish, M. 1999. The detection of earnings manipulation. Financial Analysts J. 55(5) 24–36.

Bertomeu, Jeremy and Cheynel, Edwige and Floyd, Eric and Pan, Wenqiang, Using Machine Learning to Detect Misstatements (December 1, 2019). Review of Accounting Studies, Forthcoming, Available at SSRN: https://ssrn.com/abstract=3496297

Bertomeu, Jeremy and Cheynel, Edwige and Floyd, Eric and Pan, Wenqiang, Using Machine Learning to Detect Misstatements (December 1, 2019). Review of Accounting Studies, Forthcoming, Available at SSRN: https://ssrn.com/abstract=3496297

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

Friedman, Jerome H (2001) ‘Greedy function approximation: a gradient boosting machine.’ Annals of statistics pp. 1189–1232

Guelman, Leo (2012) ‘Gradient boosting trees for auto insurance loss cost modeling and prediction.’ Expert Systems with Applications 39(3), 3659–3667

Hayes, B. Louise and Boritz, Efrim, Classifying Restatements: An Application of Machine Learning and Textual Analytics (February 15, 2019). 2016 Canadian Academic Accounting Association (CAAA) Annual Conference, Available at SSRN: https://ssrn.com/abstract=2716166 or http://dx.doi.org/10.2139/ssrn.2716166

Hribar, P., Jenkins, N.T. The Effect of Accounting Restatements on Earnings Revisions and the Estimated Cost of Capital. Review of Accounting Studies 9, 337–356 (2004). https://doi.org/10.1023/B:RAST.0000028194.11371.42

Ila Dutta, Shantanu Dutta, Bijan Raahemi, Detecting Financial Restatements Using Data Mining Techniques, Expert Systems with Applications (2017), doi: 10.1016/j.eswa.2017.08.030

Kleinberg, Jon, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan (2017) 'Human decisions and machine predictions.' The quarterly journal of economics 133(1), 237–293

Lobo, Gerald J. and Zhao, Yuping, Relation between Audit Effort and Financial Report Misstatements: Evidence from Quarterly and Annual Restatements (May 14, 2013). The Accounting Review, Volume 88, Issue 4, 2013, Available at SSRN: https://ssrn.com/abstract=2265051

Mark Cecchini, Haldun Aytug, Gary J. Koehler, Praveen Pathak, (2010) Detecting Management Fraud in Public Companies. Management Science 56(7):1146-1160. https://doi.org/10.1287/mnsc.1100.1174

Martin, R.D. 2013. Audit Quality Indicators: Audit Practice Meets Audit Research. Current Issues in Auditing 7 (2): A17-A23

Perols, Johan and Bowen, Robert M. and Zimmermann, Carsten and Samba, Basamba, Finding Needles in a Haystack: Using Data Analytics to Improve Fraud Prediction (April 6, 2015). Available at SSRN: https://ssrn.com/abstract=2590588 or http://dx.doi.org/10.2139/ssrn.2590588

Plumlee, M. and T.L. Yohn. 2010. An Analysis of the Underlying Causes Attributed to Restatements. Accounting Horizons 24 (1): 41-64.

Tan, Christine E. L. and Young, Susan M., An Analysis of 'Little r' Restatements (November 1, 2014). Fordham University Schools of Business Research Paper No. 2407659, Available at SSRN: https://ssrn.com/abstract=2407659 or http://dx.doi.org/10.2139/ssrn.2407659

Zhang, M. (2013). Evaluation of Machine Learning Tools for Distinguishing Fraud from Error. Journal of Business & Economics Research (JBER), 11(9), 393-400. https://doi.org/10.19030/jber.v11i9.8067

Zhang, Yanru, and Ali Haghani (2015) ‘A gradient boosting method to improve travel time prediction.’ Transportation Research Part C: Emerging Technologies 58, 308–324