Index	Title	Author
1	Prediction Error, Noise, and Bias of Auditors' Going Concern Opinions and the Role of Machine Learning	Chanyuan (Abigail) Zhang, Yu Gu, Miklos A. Vasarhelyi
2	Using Supervised Learning Algorithms to Predict Discontinued Operations in Nonprofit Institutions	Chengzhang Wu and Richard B. Dull
3	Continuous Student Performance Monitoring with SWAM	Fangbing Xiong and Hussein Issa
4	Outlier Detection Analysis: an Integrated Application to the Financial Data	Hanxin Hu, Qing Huang, Hanchi Gu, Kathy Wei, Alex Kogan, and Miklos Vasarhelyi
5	Business News Headlines and the Prophetic Vision of Bankruptcies: An Application of Natural Language Processing	Deniz Appelbaum, Huijue (Kelly) Duan, Hanxin Hu, Ting (Sophia) Sun
6	Text Visual Analysis in Auditing: Data Analytics for Journal Entries Testing	Heejae (Erica) Lee, Lu Zhang, Qi Liu, Miklos Vasarhelyi
7	GASB Post Implementation Review: A Process of Unstructured Data Collection	Huaxia Li, Kathy Wei, Kevin Moffitt, Miklos Vasarhelyi
8	Blockchain-Enabled Continuous Audit: Implementation of Blockchain - Enabled Smart Contract with the Integration of Business Process Management	Jumi Kim, Maurício Vasconcellos Leão Lyrio, Jun Dai, Miklos A. Vasarhelyi
9	Blockchain for Improving Management and Transparency in Public Procurement: A Proposal for Smart Contracts Model for the Executive Branch of the State of Santa Catarina, Brazil	Maurício Vasconcellos Leão Lyrio, Jun Dai, Jumi Kim, Miklos A. Vasarhelyi
10	Discuss the Financial Distress of Local Governments: Using Machine Learning to Predict the Possibility of Bankruptcy	Ruanjia Liu, Huaxia Li, and Kyunghee Yoon
11	A Predictive Analytical Approach to the Evaluation of Internal Controls	Huijue (Kelly) Duan, Miklos Vasarhelyi, Mauricio Codesso
12	Using Machine Learning to Detect and Predict Restatements - The Enigma of Unlabeled Positives	Lanxin Jiang, Miklos Vasarhelyi , Chanyuan (Abigail) Zhang
13	Do Auditors Respond to Cues that Provide Indications of A Heightened Risk of Material Misstatement for Fair Value Measurement?	Helen Brown-Liburd, Dereck Barr-Pulliam, Stephani Mason, Meehyun Kim
14	Expected Loan Loss Provisioning Using a Machine Learning Approach	Nichole Li, Alexander Kogan
15	Continuous Monitoring with Interactive Data Visualization: An Application to Healthcare Payroll Process	Guangyue Zhang, Hilal Atasoy, Miklos Vasarhelyi
16	Audit with Machine Learning: Applying an Unsupervised Algorithm on Journal Entries of an Australian Bank	Danyang (Kathy) Wei, Soohyun Cho, Miklos A. Vasarhelyi
17	Channel Stuffing Detection Model: Using Processing Mining	Huijue Duan, Jumi Kim, Qing Huang, Miklos A. Vasarhelyi
18	Cybersecurity Risk Disclosure During the COVID-19 Pandemic	Qingman Wu, Won Gyun No
19	Measuring Text Information of Green House Gas (GHG) Emission	Dong Gil Kim and Won Gyun No
20	Impact of Business Analytics on Managerial Accounting	Wenru Wang, Miklos A. Vasarhelyi
21	Impact of the Remote Working Environment on Cybersecurity Risk in Organizations	Hongmin (William) Du, Won Gyun No
22	Predicting the Discontinuity of Non-Profit Organizations Using the Machine Learning Approach	Xinxin Wang , Heejae Lee, and Richard Dull
23	Audit 4.0-based ESG Assurance: An Example of Using Satellite Images on GHG Emissions	Yu Gu, Jun Dai, Miklos A. Vasarhelyi

Prediction Error, Noise, and Bias of Auditors' Going Concern Opinions and the Role of Machine Learning

Chanyuan (Abigail) Zhang, Yu Gu, Miklos A. Vasarhelyi

Introduction

- This research studies the error, bias, and noise of auditors' going concern opinion (GCO) in predicting firm default and we examine the role of machine learning (ML) in impacting the above prediction features.
- We find that advanced machine learning models can significantly reduce the error, bias, and noise of default firms compared to GCO, consistent with the theory in Kahneman et al. (2021). Following Kahneman et al. (2021), we also explore the value of diversity in improving prediction quality. To that end, we construct four "artificial auditors" representing Big4 auditors and we find that the consensus from these artificial auditors can significantly reduce the prediction error compared to GCO, which is issued by one auditor.
- Our study adds to the accounting literature by examining the quality of GCO and the mechanism through which ML improves default prediction from the angle of prediction error, bias, and noise.

Data and Sample

- There are 15,973 firm-year observations spanning from 2003 to 2015 in our dataset, comparable to Gutierrez et al. (2020). GCO and Bankruptcy data comes from Audit Analytics, and default data comes from Gutierrez et al. (2020). There are 1345 GCO issuance, 194 following-year bankruptcy, and 269 following-year defaults in our dataset, comparable with Gutierrez et al. (2020).
- Like what prior literature shows, our sample shows that of the firms that received GCO in year t, less than 10% of them went bankrupt in year t+1 (in our case, 8% from 107/ (107 + 1238)), and for firms that went bankrupt in year t+1, about half received GCO in year t (in our case, 55% from 107/ (107+87)). If we consider the broader definition of default, still about 10% of GCO receivers defaulted in year t+1 (in our case, 10% from 137/ (137+1208)), and for firms that defaulted in year t+1, about half received GCO in year t (in our case, 51% from 137/ (132+137)).
- Collectively, either we measure firms' discontinuity of business by bankruptcy or default, auditors' GCO has limited ability of prediction in terms of low precision and low recall. In other words, GCO is high in both false positive and false negative errors.

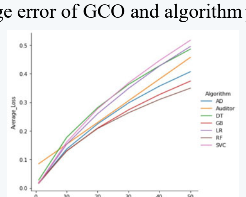
Research Design

- 1) we focus on the prediction error of GCO and algorithm predictions with the consideration of the cost imbalance between false positive and false negative errors;
- 2) we dive into the noise and bias of GCO and examine the mechanism through which learning models impact default prediction compared to GCO;
- 3) we introduce diversity in default prediction by creating four "artificial auditors" representing the Big4 auditors and examine if the consensus of "artificial auditors" can produce better predictions compared to GCO.

Revisit Using Learning Models to Predict Firm Default

- We adopt a rolling-window design to conduct the machine learning experiment. There are six learning algorithms used in this experiment: Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), Support Vector Machine Classifier (SVC), Gradient Boosting (GB), and AdaBoost (AB). The default parameters of each algorithm are used.
- We observe that tree-based ensemble models, especially GB and RF, consistently produce predictions that are more accurate, more precise, and more complete compared to predictions of GCO.

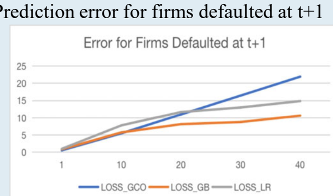
Fig.1 Average error of GCO and algorithm predictions



Error, Noise, and Bias of GCO and Predictions from Learning Models

- We examine the prediction error, noise, and bias of GCO issued to our sample firms. Next, we compare the prediction error, noise, and bias of learning models' predictions to those of GCO.
- We can observe from Figure 2 that when there is a cost imbalance, GB produces less prediction error compared to GCO. Although GB and LR generate predictions with higher error (noise) compared to GCO when there is a high imbalance of cost, such increased error is caused by more false positive mistakes on surviving firms, and they are less costly compared to false negative mistakes for default firms.

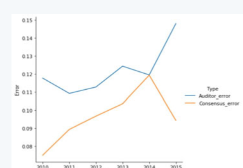
Fig.2 Prediction error for firms defaulted at t+1



Error, Bias, and Noise of Models of Big4 Auditors

- We follow Kahneman et al. (2021) to explore whether introducing diversity, the desired variance, to the GCO issuance process can reduce the prediction error of GCO. To do so, we create four "artificial auditors" representing the Big4 auditors and examine if the consensus of "artificial auditors" can produce better predictions compared to GCO.
- Figure 3 demonstrates that from 2010 to 2015, the consensus error is always less than auditor error, consistent with Kahneman et al. (2021)'s theory that introducing diversity in the decision-making process can improve the quality of prediction.

Fig.3 Prediction error for audit error and consensus error



Contribution

- Our study pioneers in examining the prediction error, bias, and noise of auditors' GCO and how learning models of default prediction performs in these aspects compared to auditors' GCO. The initial evidence shows that learning models, especially complicated and tree-based models can significantly reduce the noise of default prediction and produce less false negative errors that have costly consequences.
- This study adds evidence to the accounting literature about learning model's relative advantage compared to existing audit procedures in predicting firm default and can spark future research to further examine related issues.

RUTGERS
Rutgers Business School
Newark and New Brunswick

Using Supervised Learning Algorithms to Predict Discontinued Operations in Nonprofit Institutions

Chengzhang Wu and Richard B. Dull

Introduction

- Nonprofit organizations play an important role in economy worldwide. The operation of those organizations primarily rely on the donation from donors. However, the discontinued operation of nonprofit organizations due to financial reasons have brought the problem of unbalanced economy resource allocation.
- The prediction of bankruptcy or dissolution of for-profit companies has been widely studied using different data analytic methodologies, and various machine learning approaches have been evaluated and demonstrated effective. However, the prediction of non-profits discontinued operation is still rarely studied.
- This study attempts to do such prediction comparing the performance of Logistic Regression, Decision Tree, Random Forest, Multilayer Perceptron, Support Vector Machine and Bayes Net. The overall effectiveness of different prediction performance will be assessed.

Research Questions

- RQ1: Among all the algorithms compared in this research, which algorithm performs more effectively?
- RQ2: In the group of predictors used in this study, which set of variables can be used to get better prediction results?
- RQ3: Does cost-sensitive learning algorithm improve prediction quality?

Predicting Factors

- The predictors used in the current research are based on two prior literature.
- First, Howard P. Tuckman and Chang (1991) identifies four criteria used to indicate the financial vulnerability of nonprofits. These criteria include inadequate equity balances, revenue concentration, administrative costs and operating margins. These four financial variables are used to predict discontinued operations of nonprofit organizations. These four predictors are identified as financial vulnerability (FV) set in this research.
- The second source of predictors are derived from William J. Ritchie and Kolodinsky (2003) research. In that study, 16 financial performance measurement ratios used for nonprofits are classified into four categories. These four categories are Fiscal Performance (FP), Fundraising Efficiency (FE), Public Support (PS) and Investment Performance and Concentration (IPC).

Data Collection

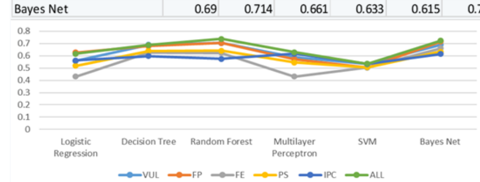
- This study uses a Form 990 database, which contains all the e-filings from 2011 to 2018. SQL is used to a query in the Form 990 data to list the all the unique records that indicated termination of operation and the year of termination.
- The number of nonprofits that discontinued operation in Form 990 accounts only 1% of the entire dataset. This algorithm is not able to capture enough information on the imbalanced dataset. Therefore, it is necessary to adjust the original dataset to be balanced to be suitable for training the model.
- The study uses under sampling technique to address the imbalanced training set. In the training set, 831 records that indicate discontinued operation in electronically filed Form 990. The other 831 records that do not indicate discontinued operation are randomly selected. The final training set used in this study includes 1662 records.

Evaluation Metrics

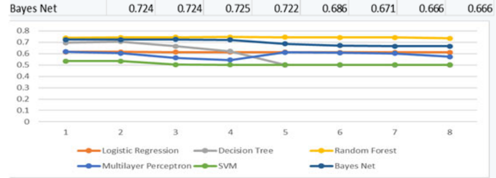
- When the binary classification prediction model is applied on test set, four categories of results could come out: True Positive, False Positive, True Negative and False Negative. The following confusion matrix depicts these four outcomes.
- Based on the matrix, there are several measurements can be used to evaluate the performance of each algorithm (listed in the following table). Considering the imbalanced nature of the population, this study uses AUC to evaluate how effective is the trained model perform on test set. The closer to1, the better the performance.

Results

Variables	VUL	FP	FE	PS	IPC	ALL
Logistic Regression	0.559	0.627	0.431	0.518	0.563	0.616
Decision Tree	0.689	0.679	0.626	0.64	0.598	0.688
Random Forest	0.705	0.705	0.624	0.641	0.574	0.739
Multilayer Perceptron	0.587	0.572	0.431	0.545	0.617	0.63
SVM	0.534	0.507	0.504	0.507	0.533	0.535
Bayes Net	0.69	0.714	0.661	0.633	0.615	0.724



Cost	1	2	3	5	10	15	20	25
Logistic Regression	0.616	0.615	0.614	0.613	0.613	0.612	0.612	0.612
Decision Tree	0.697	0.706	0.665	0.621	0.5	0.5	0.5	0.5
Random Forest	0.739	0.742	0.745	0.747	0.744	0.743	0.743	0.735
Multilayer Perceptro	0.617	0.604	0.563	0.544	0.611	0.607	0.603	0.573
SVM	0.535	0.535	0.503	0.502	0.501	0.501	0.501	0.501
Bayes Net	0.724	0.724	0.725	0.722	0.686	0.671	0.666	0.666



Conclusion and Discussion

- The highest AUC value is achieved under ALL group using Random Forest. This indicates that Among all the six algorithms, Random Forest generates the better performance.
- All the five sets of financial ratios together is more effective to predict nonprofits discontinued operation.
- The performance of different algorithms stays relatively stable with the increment of positive to false negative cost ratio. Overall, the prediction of nonprofits discontinued operation is not sensitive to the change of positive to false negative cost ratio.

Rutgers Business School
Newark and New Brunswick

Continuous Student Performance Monitoring with SWAM

Fangbing Xiong and Hussein Issa

SWAM System and BYOC courses

"BYOC: Emerging Topics in Business" is a course build on the concept of School With A Million courses (SWAM), which is a new online learning initiative of CAR Lab at Rutgers Business School.

SWAM is a platform that offers students many different modules, which cover a variety of topics. Students could take a 1-credit worth module and focus on a particular Emphasis, or take multiple credits covering different modules for different Emphasis. The important part of this course is that they can Build Your Own Courses (BYOC).

The purpose of this new learning approach is to offer them the freedom to choose the learning topics based on their interests, flexibly earn required education credits, and take the courses whenever and wherever you want. The instructor closely monitors the students learning progress and provides timely feedback and assistance when needed.

Design It Yourself in Accounting and Auditing

- **Description:** This course allows students to design this course by allowing them to select modules of interest. More specifically, students can choose any five modules from the set of all eligible modules across all BYOC courses. A comprehensive list of potential modules can be found at the end of this syllabus.
- **Required Module:** None
- **Recommended Modules** (select five):
  - Introduction to Audit Automation
  - Basics of Continuous Auditing and Continuous Monitoring
  - Introduction to Process Mining
  - Basics of Artificial Intelligence in Accounting and Audit
  - Analyses of Exceptions and Anomalies
  - Emerging Audit Evidence
  - Introduction to Blockchain and Smart Contracts
  - Introduction to Crypto Currencies
  - Basics of Cybersecurity

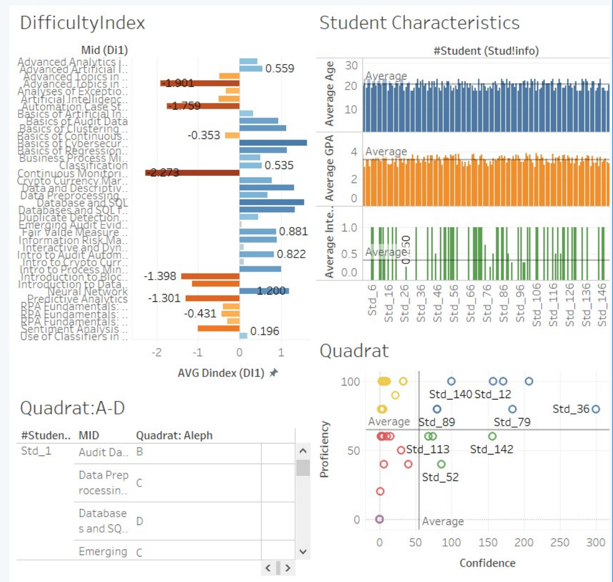
Faculty Dashboard

Realtime analysis

- Course content difficulty level
- Class performance
- Section performance
- Student performance in a specific module
- Student performance in a specific course
- Student performance across all courses
- Student progress over time
- Analysis based on students demographics

Automated early warning system

The system can automatically send an early reminder to students who haven't started their work by the deadline or perform poorly compared to their peers. Robotic Process Automation (RPA) would power the process.



Student Dashboard

Realtime Analysis for students

- Performance by module
- Performance in a specific course
- Progress over time
- Comparison to class performance

Prescriptive analysis and recommendations

Students get feedback on their study process from the system as a prescriptive analysis. This may help them to adjust their study schedule and to improve their performance.

The recommendation system would give tips for selecting the customized modules for each individual.

Components

LENA

AI-based learning diagnostic of current performance analysis, and prescriptive analytics about future learning goals.

Reco / BotViser

The recommender system provides AI-based intelligent curricula recommendations

Building Courses

Register students and select customized modules

Content Delivery

Model provides dynamic learning experience with self-paced approach.

Assessments

Automated evaluation on individual selected modules

Difficulty Index

The assumption is that neither one single student nor some students got a low score indicates that the module material is challenging. However, when many students who got high scores in other modules but they could not perform well in the particular module, we say the module is difficult. Thus, average and variance is more reasonable for difficulty evaluation. A negative SI with a large absolute value should imply a high complicated level and vise versa.

$$\text{Difficulty Index} = \Sigma \frac{SI\ of\ each\ student\ who\ choose\ module\ A}{\#student\ choose\ module\ A}$$

$$\text{Student Index }(for\ one\ student) = sign(Score\ of\ Module\ A - Average\ scores\ of\ five\ modules) \times \ln[(Score\ of\ Module\ A - Average\ scores\ of\ five\ module)]^2$$

Quadrat: Proficiency V.S. Confidence

Target: Evaluate students and identify students who need help.

Proficiency = Average scores of two attempts of a quiz

Evaluate the accuracy of students' answers.

Confidence = Time spent/Average scores of two attempts of a quiz

Evaluate the students' confidence in the answers they select.

	High proficiency	Low proficiency
High confidence	Ready for more advanced material	Misunderstanding
Low confidence	Need more practice	Under prepared

RUTGERS
Rutgers Business School
Newark and New Brunswick

Outlier Detection Analysis: an Integrated Application to the Financial Data

Hanxin Hu, Qing Huang, Hanchi Gu, Kathy Wei,
Alex Kogan, and Miklos Vasarhelyi

Definition

Outlier detection methods have been universally utilized to identify abnormal data points in datasets. There appear three generic definitions of outliers existing in prior literature:

(1)Hawkins (1980) considers an observation as an outlier if it significantly deviates from other observations leading to the suspicion that it was generated by a different mechanism;

(2)Analogous to the definition offered by Hawkins (1980), Barnett and Lewis (1984) also argue that outliers refer to the data points that are markedly distinguished from other samples in data.

(3)Johnson and Wichern (2002) defines an observation as an outlier if it is inconsistent with other observations remaining in data;

Literature

- Our examination of extant literatures shows that there seem to be limited utilization of outlier detection methods in the accounting and auditing domain (i.e., fraud detection, loan application processing, insurance fraud filtering).

- However, the successful application of these algorithms in related domains (e.g., intrusion detection and activity monitoring) demonstrate their capabilities of automatically and effectively identifying exceptional entries in datasets, manifesting the promise of successful employment of well-grounded outlier detection methods in the accounting and auditing domains.

Structure

- We identify the popular categories of algorithms by means of extensive literature search utilizing various keywords, such as “outlier detection survey”, “outlier detection literature review”, “outlier detection methods review”, etc.
- The most influential and popular surveys are selected to identify theoretically well-developed and widely-used categories of outlier detection approaches.
- This study provides overview descriptions for each popular algorithm including basic information, validation process, and utilization of real and synthetic datasets. I
- We summarize some common tools and packages that can be used to perform the tasks of outlier detection, as well as demonstrate the usage of the reviewed methods on the example of one specific credit card dataset from Taiwan.
- Finally, we discuss certain characteristics of accounting and auditing data, and pinpoint caveats often encountered when accounting researchers/ auditors attempt to apply outlier detection approaches in their domains.

Summary of Popular Categories in Literature and Surveys

	1	2	3	4	5	6	7	8	9
Chandola et al. (2009)	X	X	X	X					
Hodge and Austin (2004b)	X	X		X		X		X	
Patcha and Park (2007)	X	X		X		X		X	
Markou and Singh (2003)	X	X		X					
Pimentel et al. (2014)	X	X						X	
Akoglu et al. (2015)							X		
Zhang et al. (2010)	X	X		X		X			
Ahmed et al. (2016)	X			X					
Zimek et al. (2012)	X	X	X						
Wang et al. (2019)	X	X	X	X	X	X	X	X	X
Aggarwal (2017)	X	X	X	X	X	X	X	X	X
Domingues et al. (2018b)	X	X			X	X			
Emmott et al. (2015)		X	X		X				
Campos et al. (2016)		X	X						
Goldstein and Uchida (2016)	X	X	X	X					
Falcão et al. (2019)	X	X	X	X	X	X			
Chalapathy and Chawla (2019)					X	X		X	
Bulusu et al. (2020)	X				X	X		X	
Pang et al. (2021)					X	X		X	

1. Statistic-based 2. Distance-based 3. Density-based 4. Clustering-based 5. Ensemble-based 6. Learning-based 7. Graph-based 8. Network 9. Data Stream

RUTGERS
Rutgers Business School
Newark and New Brunswick

Business News Headlines and the Prophetic Vision of Bankruptcies: An Application of Natural Language Processing

Deniz Appelbaum, Huijue (Kelly) Duan, Hanxin Hu, Ting (Sophia) Sun

Motivations

- Bankruptcy prediction involves the early identification of companies that are experiencing a high level of extreme financial distress with a great probability of bankruptcy.
- Auditors are required to issue a Going Concern Opinion whose variables may be regarded as a predictor for bankruptcy. Their ability to read early signals of financial distress and subsequent bankruptcy would be invaluable.
- Although the auditor has access to early unaudited financial statements, other interested stakeholders can only access previously audited financial information which may not be relevant regarding current financial conditions due to reporting lag.
- News information is accessible to all stakeholders, it is available as a constant stream of "big textual data" in print, online, emails, social media, and blogs. There are many sources of news that report business events, market conditions, and other relevant information.
- News information possesses the ability to impact the auditors' and stakeholders' bankruptcy prediction.

Methodology

- Natural Language Processing Analysis Tools
  - TextBlob
  - Flair
  - VADER
- Machine Learning
  - RUSBoost
  - Balanced Random Forest

Contributions

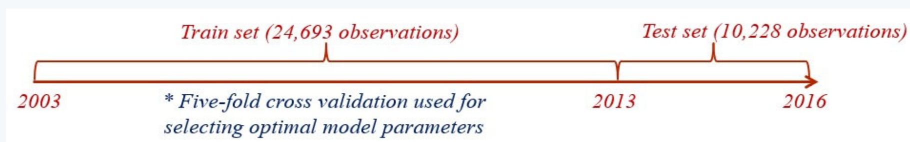
- This is the first study using New Headline to predict bankruptcy
  - News headlines potentially serve as unique indicators of sentiment and information as they are concisely worded and convey the sentiment and tone of the subsequent article.
  - News headlines are efficient and effective sources of information. It contains succinct phrases and less repetition of words, which significantly reduce the impact of irrelevant words for natural language processing.
- This study is relevant to a growing accounting literature that attempts to improve the ability for auditors and other stakeholders to predict bankruptcy by examining the textual data from different sources.

Data

- 34,921 firm-year observations from 2003-2016
  - 66 observations filed for bankruptcy
- 33 linguistic features
  - Frequency of company mentions
  - Frequency of bankruptcy-specific topics mentions
  - Sentiment features
- 20 research-based financial market variables
  - Big 4, size, aturn, curr, lev, roa, sg, mb, cash, ocf, lagloss, delay, weak, new_equity, New_debt, tenure, firstyear, litg, clev, sg_deficiency

Empirical Design

- Employ both machine learning and logistic regression
  - Machine learning approach:
    - Test the prediction accuracy and incremental prediction power of linguistic features
    - All 53 independent variables included as predictors



  - Logistic regression:
    - Examine the explanatory power and additional interpretability of linguistic features
    - Based on thee entire sample set, incorporate all independent variables as well as year and industry fixed effects

Results

- 34 qualitative features derived from published news and MD&A sections are selected as the most predictive to bankruptcy occurrence in the future;
- Relative to the financial independent variables, this set of features has statistically significant predictive power to the bankruptcy occurrence in the subsequent one/two years.

Conclusions

- The objective is to develop a bankruptcy prediction model based on a sample of Reuters news headlines whose words are analyzed using state of the art text mining and machine learning techniques.
- The empirical results suggest a significant incremental contribution from the linguistic content of news headlines and MD&A reports in the prediction explanatory accuracy of bankruptcy projections as early as two years in advance.
- There are two limitations of this study
  - Whether the headline content indeed accurately reflects the content of the article text in both factual information and sentiment perspectives
  - Including other and/or additional raw data items than the financial independent variables included in this study might provide different results.

RUTGERS
Rutgers Business School
Newark and New Brunswick

Text Visual Analysis in Auditing: Data Analytics for Journal Entries Testing

Heejae (Erica) Lee, Lu Zhang, Qi Liu, Miklos Vasarhelyi

Introduction

- Due to increased volume and complexity of business transactions in the organizations, identifying erroneous or abnormal transactions or journal entries become more challenging for audit professions. Business transaction data in the accounting system includes not only numeric values of the journal entries and categorical variables including account type and recorded date/time, but also textual data such as descriptions or comments. Understanding descriptions or other type of textual data of business transactions is also important since this information captures the nature of transactions in qualitative manner.
- While text analysis can provide new information about the company's business transactions, performing the analysis and interpreting the results require computer language skills and expertise in text mining techniques. Visual data mining, the integration of data mining technique and data visualization, can make data analysis more effective by involving human in the data exploration process (Keim, 2002).
- In this study, we propose Text Visual Analysis (TVA) approach for auditing. We argue that integration of text analysis and data visualization can improve the efficiency of audit data analytics for textual data in the organization's accounting information system for journal entry testing. The proposed method can be used for identifying unusual journal entries based on textual data of the transactions including descriptions and comments. This paper can contribute to both textual analysis and data visualization literature in accounting research by introducing the integration of two different domains. In addition, this study can provide guidance for audit practitioners and standard setters on how to apply textual analysis in audit tasks.

Background

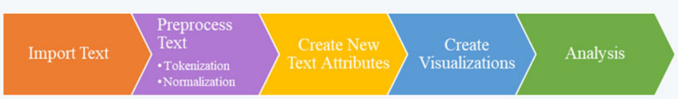
- The importance of extracting information from narrative or text has been recognized by accounting researchers for the last few decades. Several researchers reported that textual data in financial statement can be used to develop models to detect or predict negative financial outcomes such as fraud and bankruptcy (Glancy & Yadav, 2011; Gupta & Gill, 2012; Hajek & Henriques, 2017; Shirata et al., 2011). Using textual data in auditing can provide opportunities to collect new type of audit evidence.
- Examples of textual data that auditors can use for auditing include not only accounting reports, but also contracts, transcripts of conference call, emails, posts on social media, press release, etc. (Sun & Vasarhelyi, 2018; Warren, Moffitt, & Byrnes, 2015; Zhang, Stone, & Xie, 2019). Preprocessing plays an important role in text analysis since textual data are usually unstructured or semi-structured (Vijayarani et al., 2016). Many of textual data we discussed above also require preprocessing to reduce dimensionality. Tokenization and normalization are widely used preprocessing techniques in textual analysis (Sridhar, 2015; Webster & Kit, 1992).
- Introducing Big Data to audit procedures increases the task complexity by producing a large number of information cues. Visualization can play as a medium for human interaction with the data in the data analysis process. In addition, using visual programming or data mining software can facilitate data analysis since the analysis can be performed by drag-and-dropping and connecting different visual objects (Jović et al., 2014).

Case Study: Tools and Data

- **Tool**
- Orange Data Mining software was used to conduct first 3 steps of the framework. For step 4 and 5, we used Tableau Desktop software.
- **Data**
- We examine the proposed methodology by analyzing entire journal entries of the bank for a single fiscal year. Total number of transactions in the journal file was 1,438,958. We select 'Description' and 'Journal Description' as a target for text visual analysis. There are 1,547 unique 'Description' and 754,961 unique 'Journal Description' before applying the proposed methodology. 5.4% of the journal file (77,054 records) have no 'Journal Description'. We also include those transactions have no 'Journal Description' in the analysis since missing description can also be informative to the auditors.

Attribute Name	Data Type	Definition
Statement	Categorical	Type of Financial Statement
Primary	Categorical	Type of account
Account Code	Categorical	General ledger account ID
Description	String (Text)	Description of the account
Transaction Id	Categorical	Journal entry transaction ID
Net	Monetary	Net debit/credit amount posted by the journal line
Effective Date	Date/Time	Date the transaction occurred
Created Date / Time	Date/Time	Date the transaction was processed into the system
Document Type	Categorical	Type of transaction and source that was posted
User Id	Categorical	Individual or System ID that entered the transaction
Journal Description	String (Text)	Narration for the journal

Text Visual Analysis Approach



- Step 1. Import Text
- Step 2. Preprocess Text: Preprocessing is necessary to allow machine to understand the relationship between different text and human to retrieve information from the text more efficiently.
  - Tokenization: process to convert the text into basic units of the text with a meaning (tokens). (e.g, removal of stopwords, numbers, symbols, or any other unnecessary text)
  - Normalization: process to reduce the complexity of textual features. (e.g., converting into lowercase, stemming or lemmatizing)
- Step 3. Create New Text Attributes
  - Text Attribute: the features that capture the characteristic of textual data. (e.g., categories of text from text classifiers or memberships of clusters, or a single string of the preprocess text (group of tokens))
- Step 4. Create Visualization: process of creating visualizations with the original attributes and new text attributes for analytical purpose.
- Step 5. Analysis

Workflow of Script of Journal Description in Orange

Examples of TVA Dashboard

Case Study

- Journal file was imported using 'CSV File Import' widget, converted as a corpus using 'Corpus' widget, and preprocessed using 'Preprocess Text' widget. The imported text was tokenized using 'tokenization' feature of 'Preprocess Text' widget. Regular expression was used as a tokenization method. We removed stop words, symbols, numbers, and date ('Filtering'), transformed tokens to lowercase ('Transformation'), and lemmatized the tokens ('Normalization'). Regular expression was used to remove symbols, numbers, and date and 'Wordnet Lemmatizer' was used for normalization. Then, new attribute called 'Cleaned Journal Description' was created by 'Python Script' widget. We joined the preprocessed tokens with white space and form new text. We found that the proposed methodology reduced the number of unique journal descriptions from 745,961 to 25,282 . For description, we used the membership of clusters as a new text attribute. Clustering method reduced the variation of descriptions from 1,547 to 737.
- After creating two textual attributes, we created data visualizations using data visualization software Tableau. The interactive feature of data visualization software allows auditors to select specific 'Primary' and examine the membership of the clusters within the given Primary. Selecting 'Cash and Cash Equivalents' from the bottom left of the dashboard showed that there were 13 different clusters of 'Description' within this type of account. We created the bar charts to show how many transactions were recorded under each 'Journal Description'. We created visualizations for both 'Cleaned' and original journal description (top middle and top right in figure 7) and sorted it based on the number of transactions (ascending order) to examine the frequency that each description was appeared.
- We found that 'c c' was the most frequently used journal description for 'Cleaned Journal Description'. However, the visualization for 'Journal Description' showed that the most frequent journal description was 'null', which means that there was no narration about the transaction. Removing date information from the journal description in preprocessing allow computers to find similar transactions. Specifically, both 'C/C FOR 27 AUG 19' and 'C/C FOR 26 NOV 18' had 'Cleaned Journal Description' of 'c c' after preprocessing. We also found that there was a transaction that have journal description of 'accrual' by examining the journal description which were not frequently used. Considering the nature of 'Cash and Cash Equivalents' accounts, this can be a coding error. Auditors can make a judgement of whether this finding is 'material', in terms of either qualitative and quantitative, or both.

Rutgers Business School
Newark and New Brunswick

GASB Post Implementation Review: A Process of Unstructured Data Collection

Huaxia Li, Kathy Wei, Kevin Moffitt, Miklos Vasarhelyi

Introduction

- The governmental accounting has a very wide range of information users, so does the public pension accounting. Moreover, responses about the pension accounting would be expected to be different according to the group to which each stakeholder belongs.
- In order to deliver more transparent status about the public pension system to information users, in 2012, the Government Accounting Standards Board (GASB) introduced new standards—No. 67 *Financial Reporting for Pension Plans* and No. 68 *Accounting and Financial Reporting for Pensions*—by amending their existing standards No. 25 and No. 27.
- However, it is still not clear about various issues such as 1) how widely the GASB's new standards are implemented, 2) if an entity adopted the new standards, whether all the detailed amendments are completely implemented, or 3) what are the results of the new implementations.
- This research starts from filling this gap between the GASB standard and its implementation by the entities. This research will review a wide collection of CAFRs from different sources by developing automated ways. Based on collected and extracted necessary information, this research will provide the full picture of the implementation status of the GASB's new standards No. 67 and No. 68.

Robotic Report Retain and Conversion

- Unlike a for-profit disclosure platform (i.e., the SEC EDGAR system), there is no central repository for Comprehensive Annual Financial Report (CAFR). Fortunately, there are some websites which have considerable CAFRs in it. Based on the identified website, we will deploy the web-scraping technology.
- There are multiple ways in web-scraping. We will mainly use Python "Scrapy" or "Beautiful soup." By using these tools, we can automatically download the multiple CAFRs from a website.
- We have used three methods for PDF conversion, including Python packages, commercial software and Java-based tools.
- Python package is a very convenient, fast and economical method, more suitable for processing machine-recognizable, well-formatted text-based pages.
- Commercial software is another method we use, which can not only handle the scanned version of the PDF file but also can extract the text from the rotated pages.
- Tabula, which is Java-based tool, is a slightly less efficient but highly accurate method. This method requires some human effort to assist in locating the tables. We mainly use this method to deal with tables from PDF files.

Robotic Report Extraction

- **Data cleaning**: After conversion, we convert all the text to lowercase, clean up the punctuation, extra spaces, extra blank lines and useless characters.
- **Page location**: Once the text data is cleaned, we only keep the pages that contain the data we need to extract. Here we use text mining techniques to locate the corresponding pages based on the keywords defined.
- **Page split**: We separate targeted pages out for subsequent checking. If there is an abnormality in a certain data, we can quickly check the corresponding page to confirm whether there is an error in the code. If a table is split into two parts and presented on two pages, we merge these two to form a complete table.
- **Data extraction**: To extract data from converted CSV and cleaned text, we need to build a dictionary of keywords based on target items, and find the location of the values based on the index of the rows and columns, or specific paragraphs. We mainly use regular expression for data extraction.

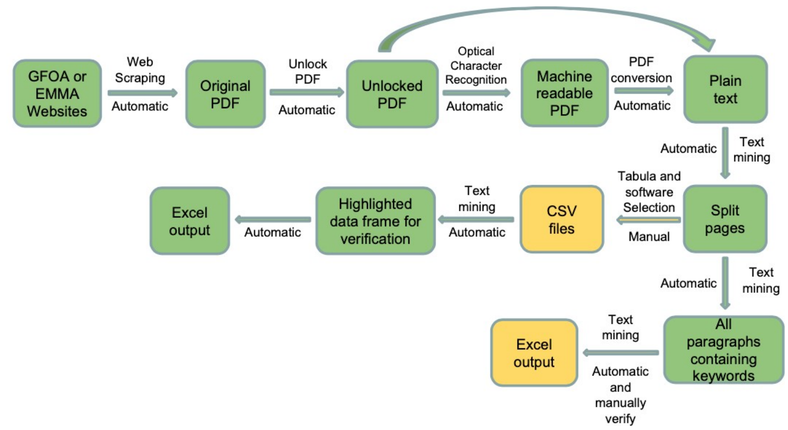
Results

- Item 5 is to examine the effects of changes in experience, assumptions and benefit terms for net pension liability.
- Our results show that we have 38 duplicate data points compared to 57 GASB results, 16 data points different from GASB results, 4 data points only found by code and also 2 data points only found by GASB.
- The overlapping rate of our result as GASB result is 66%. But when we scrutinize each of them, we find that 15 out of 16 different results are actually correct for our results. So we implement the following formular in our calculation. With this calculation, our code extraction accuracy for item 5 becomes 91%.

$$\frac{overlapping + different\ from\ GASB\ but\ correct + Code\ found\ only\ and\ correct}{overlapping + different\ from\ GASB + Code\ found\ only + GASB\ found\ only\ and\ correct}$$

- Item 6 relates to employer's proportion of collective net pension liability. Using the calculation method mentioned above, our code extraction accuracy is 89%.
- Item 8 related to fiduciary net position / total pension liability and total

GASB PIR - Process of unstructured data collection and verification



Summary of comparison results with GASB

	Overlapping with GASB	Different from GASB	GASB found only	Code found only	Code total	GASB total	Overlapping rate	Different from GASB but correct	Code found only and correct	Overall code correct rate
Item_5	38	16	2	4	58	57	66.67%	15	2	91.67%
Item_6	65	0	5	8	73	71	91.55%	0	5	89.74%
Item_8_1	114	7	5	11	126	128	89.06%	5	11	94.89%
Item_8_2	116	9	2	12	138	128	90.63%	9	12	98.56%

RUTGERS
Rutgers Business School
Newark and New Brunswick

Blockchain-Enabled Continuous Audit: Implementation of Blockchain-Enabled Smart Contract with the Integration of Business Process Management

Jumi Kim, Maurício Vasconcellos Leão Lyrio, Jun Dai, Miklos A. Vasarhelyi

Blockchain and Smart Contract

- **Blockchain Technology (BCT),** the underlying technology facilitating Bitcoin, was first proposed in 2008 (Nakamoto, 2008). Blockchain is defined as a cryptographically secured distributed ledger of mutually untrusting peers.
- Blockchain is characterized as follows: First, blockchain is **append-only and tamper-resistant** using a cryptographic hash function, making it immutable. Second, blockchain is a **distributed** ledger; thus, the ledger exists in multiple locations, consistent throughout the network. Third, blockchain is **transparent.** Here, transparency means that the information on the blockchain is visible publicly or amongst users. Lastly, blockchain runs on a **real-time** basis since it resides on the Internet.
- **Smart contracts** contain a set of rules, and if the predefined rules are met, the agreement is automatically enforced. Smart contracts promise to increase efficiency in operations, reduce operations costs, improve transparency, prevent data loss, and eliminate the risk of manipulation.

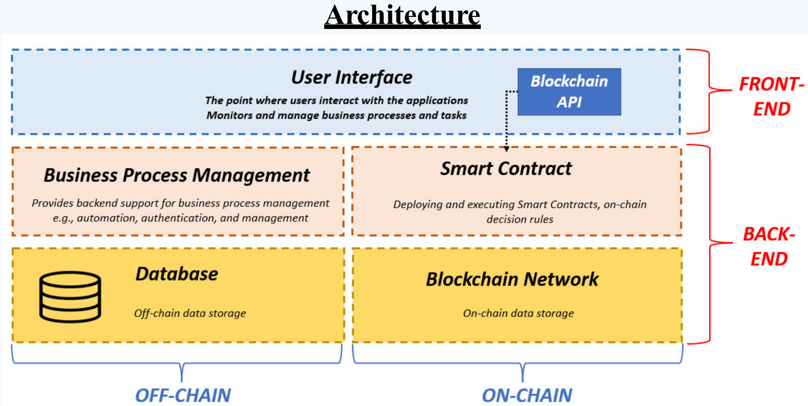
Business Process Management and Process Mining

- **Business Process Management (BPM)** is concerned with designing, executing, monitoring, and improving business processes (Mendling et al., 2018). BPM enables flexibility of business processes, reduces costs and risks, and is intuitive and easy to adopt. Besides, it helps to standardize and automate business processes, identifying and eliminating redundant activities and providing the processes' various aspects in real-time visualization (Barbu et al., 2020; Convex, 2019).
- **Business Process Mining (PM)** is one of the BPM technologies, aims to extract information from stored data (i.e., event logs) from the available information systems (Saylam & Sahingoz, 2013). Process Mining (PM) discovers a process model reflecting reality, verifying whether the model conforms to reality, and improving the existing model using reality (van der Aalst, 2014).
- PM provides the opportunity of **Continuous Audit** (CA), which is differentiated from traditional audit by its enhanced relevance with the timeliness of audit results. CA starting from the business processes has already been part of both theoretical frameworks (Chan & Vasarhelyi, 2011; Vasarhelyi et al., 2004; Vasarhelyi & Halper, 1991) and some practical implementations (Borthick, 2012; Alles et al., 2008). Especially, business processes are most related to the first levels of audit objectives for continuous assurance, transactional verification, and hold most promises to start implementing the principles of continuous auditing (Jans & Hosseinpour, 2019).

Architecture and Framework

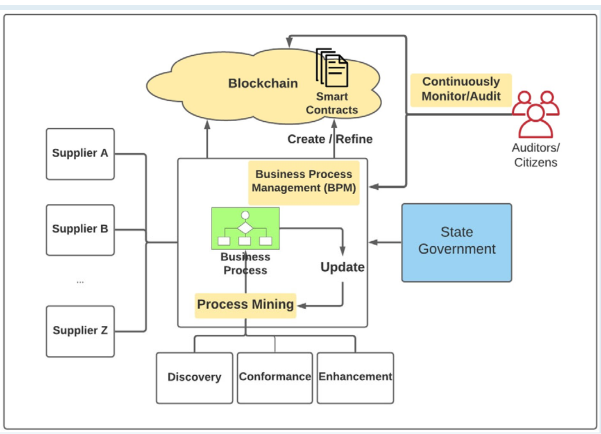
- The framework consists of five components: **User Interface, Business Process Management (including Process Mining), Smart Contract, Off-Chain Database, And Blockchain.**
- The first component is the **User interface (UI),** where users interact with the applications. UI includes users authentication, data entry, and monitoring the progress and metrics on business processes. The accessible and comprehensible UI supports the effectiveness of the application by enabling users to interact with the application better.
- The second component is **Business Process Management,** which enables automating, managing, and monitoring tasks and overall processes. The analysis from the BPM component updated on the UI enables users to monitor and control the processes. Meantime, the BPM component stores logs and data on the database system and Blockchain.
- The third component is **Smart contracts**, where on-chain rules are encoded.
- The fourth component is an off-chain database. We use the SQL database, considered as an industry-standard currently. An off-chain database should operate four basic storage operations: create, read, update, and delete (CRUD) and handle big data with comparatively lower costs.
- The last component is Blockchain, an on-chain database where smart contracts are executed and data is recorded permanently.

Architecture



On-chain vs. Off-chain

	On-Chain	Off-Chain
Definition	Data stored and rules executed on the blockchain	Data stored and rules executed outside the blockchain
Criteria	Data is relatively crucial and smaller	Data is relatively less crucial and enormous
	Data needs to be permanently stored	Data needs to be modifiable
	Data need to disclose publicly and related to the publics' interest	Data is privacy-sensitive and not related to the publics' interest
	Rules must be prevented from dominating by powerful individuals	Rules allowed being dominated by powerful individuals
	Rules are related to the publics' interest	Rules are for the business-specific purpose
Example	Rules for hiring one of the bidders	Contractor and contractee must sign to create a contract



The main flow of the **proposed framework** is as follows: 1. A business process model is extracted from event logs using PM. The process model is inserted into BPM software to monitor and control the operations. 2. The state government and related parties communicate on the web application for inter-organizational transactions while maintaining their own information system. 3. BPM component enables monitoring business processes. The business operation logs are recorded on the database system and the Blockchain. The process model is updated using PM if necessary. 4. Smart contracts are updated and refined through the PM. 5. Internal and external auditors continuously monitor business processes and operations while citizens monitor the Blockchain and smart contracts for fairness, accuracy, efficiency, and effectiveness of procurement processes.

RUTGERS
Rutgers Business School
Newark and New Brunswick

Blockchain for Improving Management and Transparency in Public Procurement: A Proposal for Smart Contracts Model for the Executive Branch of the State of Santa Catarina, Brazil

Maurício Vasconcellos Leão Lyrio, Jun Dai, Jumi Kim, Miklos A. Vasarhelyi

Blockchain applications in the public sector

- After its appearance, interest in the subject grew, reaching its peak around 2017, closely associated with one of its best-known applications, Bitcoin. The discussion on the subject has expanded to other forms of use in addition to the cryptocurrency market. The main innovation brought by the Blockchain is that it configures itself as a distributed ledger, in which information can be registered but later becomes irrevocable. Thus, it eliminates the need to maintain central intermediaries, generating profound economic, political, and social implications (Allessie et al., 2019; Dai & Vasarhelyi, 2017; Pierro, 2017).
- Blockchain strengthens information security due to its distributed manner on multiple servers, making it difficult for hackers to manipulate it and improving transparency and auditability in real-time.
- More than 100 Blockchain projects created to transform government systems are currently being carried out in over 30 countries (Jun, 2018), demonstrating that Blockchain, more than an abstract concept, is becoming a reality within the public sector.
- The prior experiences demonstrate its potential uses in the public sector to improve government efficiency, transparency, and accountability.
- Smart contracts are executable codes running on a blockchain platform to facilitate, execute, and enforce the terms of an agreement established between two untrusted parties.

Public Procurement

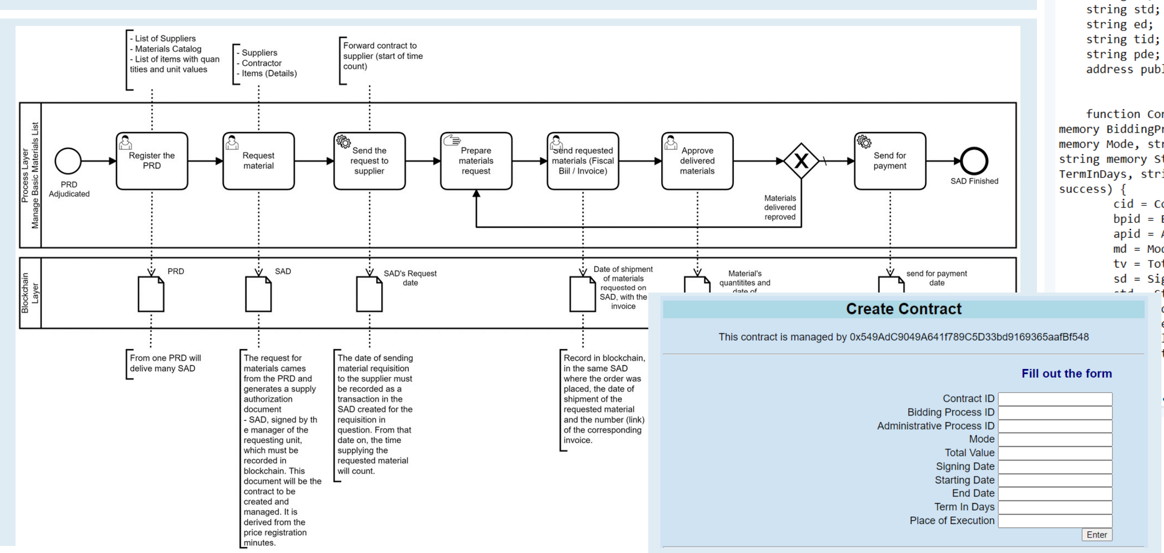
- Public procurement is "the acquisition of goods and services by a government or a public sector organization" (Bleda & Chicot, 2020; Rolfstam, 2013; Uyarra & Flanagan, 2010). Public purchasing performs as a strategy, making the public policies more efficient and effective, becoming an inducer of economic and social development, and benefiting society.
- The configuration of the purchasing process from heterogeneous information systems poses the complexity and risk associated with it. Currently, the purchasing cycle of the government of the State of Santa Catarina is fragmented and executed in multiple systems, which entails high management complexity, difficulties in integrating information, and duplication of databases in different principles. It becomes necessary to integrate the entire process "end-to-end," thus, the buyer should be able to access all stages of the procurement cycle through integration to make it more effective, reliable, and transparent.
- Besides, the current system does not offer a way to carry out a purchase plan based on the analysis of average consumptions and contract balances, preventing the parameterized alerts linked to the validity of contracts and related to available resources (minutes/contracts balance). Likewise, it is not currently possible to generate alerts considering the best time to purchase concerning seasonality and other issues inherent to the objectives to be acquired.

Blockchain to Improve Transparency

- Batubara et al. (2019) argue that there are three types of transparency: (i) data transparency - refers to the information on acts. It has been obtained through open data portals. (ii) process transparency – refers to the transparency of steps, behaviors, and interactions during the various public administration processes. This type of transparency, in general, is still not common in Brazil. Initiatives that seek to present this type of information are rare. Finally, (iii) Transparency of decisions – advancing on a deeper level of discussion, this type of transparency concerns the intentions and rationality behind the Government's decisions, actions, and policies. It seeks to explain and justify to society the reasons for the decisions taken by public managers
- Blockchain enables the improvement of transparency in its three types, deepening and increasing the level of information available to citizens in a way that has not yet been achieved on a large scale in Brazil, justifying its relevance. Regarding data transparency, Blockchain allows information to be made accessible, understandable, and in real -time. Regarding process transparency, through the Blockchain, it is possible to track and provide information about the progress of processes, allowing all parties involved to know exactly which stage a particular process is at and its history. Finally, regarding transparency of decisions, the business rules that guide decisions may be codified for automation, and the created algorithms can be made available to society for social control.

Proposed Prototype

- This project aims to develop a Blockchain-enabled Smart Contract prototype for the Executive Branch of the State of Santa Catarina to manage the list of the essential materials.
- The prototype uses the Blockchain Ethereum, a public blockchain without permission, to achieve complete transparency for the object of the prototype. Due to avoid unnecessary costs incurring to process a transaction on the Mainnet, the prototype has been mock tested on one of the Testnet. The authentication and registering functions are borrowed from MetaMask, a digital wallet, to interact with Ethereum's network.
- This project built a Dapp to integrate Ethereum blockchain, smart contracts, and user interface.



Smart contract example— initiate a contract

```
pragma solidity >=0.7.0 <0.9.0;

/**
 * @title SC-contracting
 * @dev Contract registration form for users to create a contract.
 */

 contract SCContract {

    string cid;
    string bpid;
    string apid;
    string md;
    string tv;
    string sd;
    string std;
    string ed;
    string tid;
    string pde;
    address public manager;


    function ContractInformation (string memory ContractID, string
memory BiddingProcessID, string memory AdministrativeProcessID, string
memory Mode, string memory TotalValue, string memory SigningDate,
string memory StartingDate, string memory EndDate, string memory
TermInDays, string memory PlaceofExecution) public returns (bool
success) {
        cid = ContractID;
        bpid = BiddingProcessID;
        apid = AdministrativeProcessID;
        md = Mode;
        tv = TotalValue;
        sd = SigningDate;
        std   tartingDate;
              dDate;
              ermInDays;
              laceofExecution;
              true;
```

Rutgers Business School
Newark and New Brunswick

Discuss the Financial Distress of Local Governments: Using Machine Learning to Predict the Possibility of Bankruptcy

Ruanjia Liu, Huaxia Li, and Kyunghee Yoon

Introduction

- The Census Bureau classifies local governments as

  1) General purpose: cities and counties

2) Special-purpose: school districts, water authorities and other narrowly-defined municipalities

- Consequences of bankruptcy protection

  Credit ratings drop, borrowing costs increase; Temporary cash flow relief (dissipate within a few years)

- Various literature on assessment of municipality financial condition

  1)Brown's ten – point test10-point test of financial condition: Toward an easy-to-use assessment tool for smaller cities, Brown, K. W. (1993)

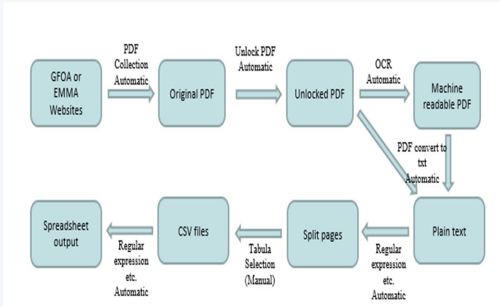
  2)The Revisiting Kenneth Brown's "10-Point Test", Maher and Nollenberger, 2009

  3)Thomas G (2002) review preliminary information risk assessment (1 study), control risk assessment (2 studies), errors and fraud (6 studies), going-concern audit opinion (3 studies), financial distress (3 studies), and bankruptcy (12 studies)

  4)Fanning and Cogger (1998) compared logistic regression, linear discriminant analysis, quadratic discriminant analysis, and an NN model

Data and Selected Financial Ratios

- Entities: only cities and counties (general purpose)
- Data Source: CAFRs from 2015-2017
- Data selection



- Financial variables (4 types)

  Financial position Financial

  performance Financial

  capability

  Other ratios

Methodology

- Resampling Techniques

  1) Oversampling

  Random Over-Sampling Examples (ROSE)

  Synthetic Minority Oversampling Technique (SMOTE)

  2) Under-sampling

  Each subset contains all the target observations (defaults) and different random subsamples of non-defaults observations

  3) Combination of over- and under- sampling

  Smotetomek

- Algorithms

  Zero Rule Algorithm as baseline model for comparison

  Gradient Boosting Machine (GBM)

  Support Vector Machine (SVM)

  Artificial Neural Networks (ANNs)

  Logistic Regression

  Naive Bayes Classification

Descriptive Statistics

Table 1. Descriptive Analysis of Illinois State

Interperiod equity** (GA)		BTA self-sufficiency***		Revenue dispersion**		Change in net position to total revenues *		(net OPEB + net pension liability)/Total Liabilities *** (GA)		(net OPEB + net pension liability)/Total Liabilities *** (BTA)		Current Ratio		Unrestricted net position / annual revenue	
Mean	0.5382286	Mean	0.9535598	Mean	0.7801713	Mean	0.3365964	Mean	0.1356493	Mean	0.0723972	Mean	6.6857693	Mean	-0.69919
Standard Erro	0.1419532	Standard Erro	0.0514688	Standard Erro	0.0541479	Standard Erro	0.224258	Standard Erro	0.0472105	Standard Erro	0.024948	Standard Erro	0.4097373	Standard Erro	0.5570352
Median	0.2448858	Median	1.0001956	Median	1.0090598	Median	0.0139247	Median	0.0047792	Median	0.0390765	Median	5.863672	Median	0.09603
Mode	#N/A	Mode	#N/A	Mode	1	Mode	#N/A	Mode	#N/A	Mode	#N/A	Mode	#N/A	Mode	#N/A
Standard Dev	1.5022911	Standard Dev	0.5069085	Standard Dev	0.5781418	Standard Dev	2.2425805	Standard Dev	0.3532911	Standard Dev	0.1115706	Standard Dev	4.4508865	Standard Dev	5.8687301
Sample Varia	2.2568786	Sample Varia	0.2569562	Sample Varia	0.334248	Sample Varia	5.0291671	Sample Varia	0.1248146	Sample Varia	0.012448	Sample Varia	19.81039	Sample Varia	34.441993
Kurtosis	33.212087	Kurtosis	48.769522	Kurtosis	11.385365	Kurtosis	47.651107	Kurtosis	3.6687491	Kurtosis	6.9964928	Kurtosis	7.0682498	Kurtosis	44.516432
Skewness	-0.905407	Skewness	-6.123254	Skewness	0.0326734	Skewness	6.6267751	Skewness	-0.199398	Skewness	2.4249908	Skewness	2.1364255	Skewness	-5.110444
Range	19.547718	Range	5.0544302	Range	5.8233367	Range	20.672727	Range	2.1324922	Range	0.5272642	Range	29.773798	Range	67.353074
Minimum	-10.02941	Minimum	-3.219831	Minimum	-1.878454	Minimum	-2.33673	Minimum	-1.246897	Minimum	-0.077169	Minimum	0.2578419	Minimum	-49.4791
Maximum	9.5183097	Maximum	1.8345989	Maximum	3.9448822	Maximum	18.335997	Maximum	0.8855947	Maximum	0.4500947	Maximum	30.03164	Maximum	17.873972
Sum	60.281599	Sum	92.495302	Sum	88.939528	Sum	33.659639	Sum	7.59636	Sum	1.4479443	Sum	788.92078	Sum	-77.61005
Count	112	Count	97	Count	114	Count	100	Count	56	Count	20	Count	118	Count	111
Confidence L	0.2812897	Confidence L	0.1021647	Confidence L	0.1072768	Confidence L	0.4449766	Confidence L	0.094612	Confidence L	0.0522167	Confidence L	0.8114633	Confidence L	1.103913

Model Evaluation

- The evaluation metrics for learning algorithms in this research are AUC (Area under the ROC Curve) (Fawcett, 2006; Narkhede, 2018) and Precision-Recall curve (PR) (Davis and Goadrich, 2006)
- Two hypotheses:
  - Null hypothesis: the municipality is predicted as default. We define it as positive class with value of 1
  - Alternative hypothesis: the municipality is predicted as non-default. We define it as negative class with value of 0
- Further research
  - Machine learning algorithms VS advanced statistical models (i.e., logistic regression)

RUTGERS
Rutgers Business School
Newark and New Brunswick

A Predictive Analytical Approach to the Evaluation of Internal Controls

Huijue (Kelly) Duan, Miklos Vasarhelyi, Mauricio Codesso

Motivations

- The Sarbanes Oxley Act (SOX) has fundamentally changed financial reporting, auditing, internal control, standard-setting, and corporate governance. It has both intended and unintended benefits and consequences.
- SOX has imposed significant requirements on publicly traded companies and auditors, and it has significantly impacted companies' operational functions and audit procedures.
- Companies implemented various controls at different levels within the organizations to meet the SOX requirements. However, some of these controls might be redundant or might not be appropriate. Companies might run into the risk of not operating efficiently and effectively, and they might miss key controls that could lead to catastrophic results. The number of control testing that the companies and the auditors need to perform has become an intensive task each year, and it requires extensive resources and time to complete the examination.
- Internal control evaluation research has been neglected, with no new significant research paradigms being developed since the '70s and the '80s.

Objectives

- Develop a predictive analytical approach to establish an internal control evaluation model by incorporating process mining and machine learning algorithms into traditional audit procedures.
- Apply the analytical approach to provide a more informative risk assessment, systematically analyze companies' internal controls, and improve the quality of internal control testing as well as substantive testing.

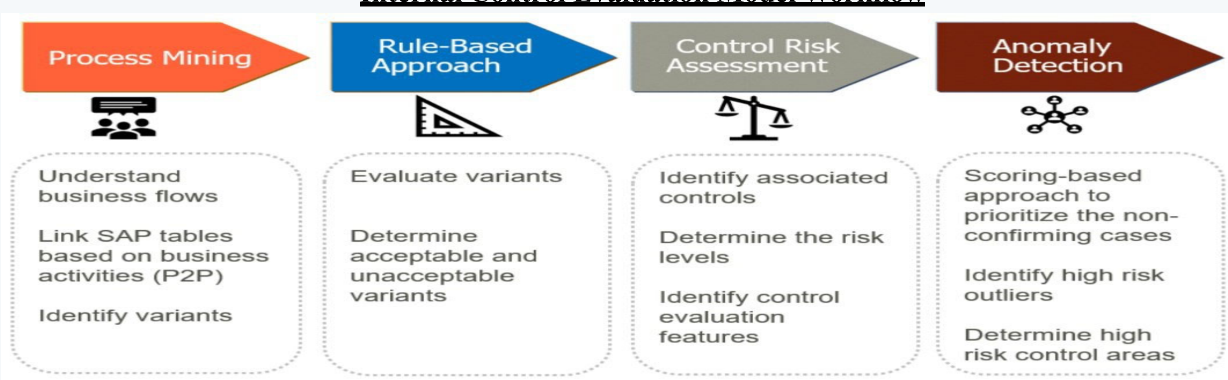
Contributions

- This study presents a new paradigm for internal control evaluation. It touches a long-neglected area in academia. The paradigm enables auditors to systematically evaluate the controls based on full population testing without extensive manual testing. The model solves the numerous exceptions noted, allows the auditors to focus on high anomalous transactions and high-risk control areas.
- New audit evidence, process mining, is presented to evaluate internal controls that allow auditors and management to facilitate a proactive investigation of the transactions that are deviated from the normal path.
- The research presents a way of performing the full population testing. A manufacturing company's procurement business process for 2019-2020 and controls over the procurement processes are analyzed.
- The internal control evaluation model contributes to different audit procedures, including risk assessment, internal control testing, and substantive testing.

Findings

- Concerns over the purchase approval and disbursement processes (e.g., accurate payment, authorized purchases, segregation of duties).
- Missing controls are identified.

Internal Control Evaluation Model Workflow



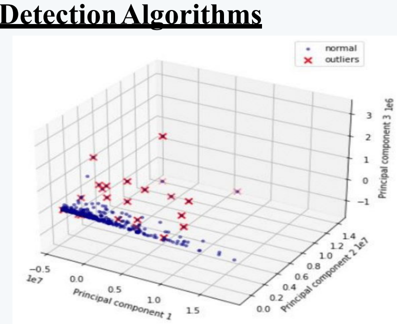
Process Deviations Evaluation

Index	Non-Confirming Type	# of Non-Confirming Cases	Business Evaluation Rules To Determine Acceptable or Not
1	Invoice Receipt is followed by Clear Invoice	26,623	Acceptable
2	Clear Invoice is followed by Tax Release	24,375	Acceptable
3	Create PO is followed by Purchase Requisition	43	Not Acceptable
4	Good Receipt is followed by Clear Invoice	115	Acceptable
5	Tax Release is followed by Clear Invoice	585	Acceptable
6	Start with Invoice Receipt	100	Not Acceptable

Anomaly Detection Algorithms

Scoring-based techniques

- Isolation Forest (iForest)
- K Nearest Neighbor (KNN)
- Local Outlier Factor (LOF)



Conclusion

- This study applies a Predictive Analytical Approach to establish an Internal Control Evaluation Model. It integrates process mining and machine learning into the traditional audit procedures, performs full population testing, systematically evaluates business transactions, assesses control risks, and identifies potential control issues.
- The internal control evaluation model identifies potential control issues and missing controls in the purchase approval and disbursement processes, including segregation of duties. It directs the auditors' investigation to high-risk control areas.

RUTGERS
Rutgers Business School
Newark and New Brunswick

Using Machine Learning to Detect and Predict Restatements - The Enigma of Unlabeled Positives

Lanxin Jiang, Miklos Vasarhelyi , Chanyuan (Abigail) Zhang

Introduction

- This study explores the issue of "unlabeled positives" in accounting research that uses machine learning to detect or to predict financial statement restatements.
- In restatement prediction and detection, extant accounting research adopts supervised learning which assumes that the dataset has complete labels and that the absence of a restatement label indicates a clean financial report.
- To address the existence of unlabeled positives in restatements, we introduce a semi-supervised learning technique called Positive-Unlabeled (PU) learning to accounting researchers. Our initial evidence shows that PU Learning is superior to supervised learning in detecting and predicting restatements, especially the unlabeled positive instances.

Data and Methods

- We obtained COMPUSTAT population from 2000 to May 2021 and matched with restatement data from the Audit Analytics database. After handling missing values, we retained 65,817 firm-year observations with 25 variables.
- PU learning is the setting where a learner only has access to positive and unlabeled data.
- The most commonly made assumptions is the Selected Completely at Random (SCAR) assumption: the labeled examples are selected independent of their attributes from the positive distribution. We use SCAR-Elkanoto (Elkan and Noto, 2008), SCAR-KM2 (Ramaswamy et al., 2016), SCAR-TIcE (Bekker and Davis, 2018) in our experiment.
- A less restrictive assumption is Selected at Random (SAR) assumption, which assumes that the probability is a function of a subset of an example's attributes (the propensity score). We also use SAR-EM (Bekker et al., 2019) under this assumption in the experiment.

Experiment Design

- Assumption: we have the "complete" labels for 2003-2012. Thus, we can use information available as of today to evaluate the performance of algorithms.
- Setting: learning from PU datasets in order to detect and predict unknown restatements. P here is the labeled positives, i.e., the restatements announced as of the prediction year; U here is the unlabeled data points, which includes normal data points and restatements that had not been announced as of the prediction year.
- Rolling-window design: we trained the models with five years of historical data and then detected unlabeled positives in the same five years and predict restatements in the following year.
- Preprocessing: we built 10 datasets at the beginning. For each subset, we adopted the same preprocessing procedures.
- Evaluation metrics: ROC AUC, partial AUC, PR AUC, f1 score, precision, and unlabeled true positive rate (# unlabeled positives identified / # unlabeled positives).

Experiment Results

Table 2 Average Detection Performance for Unlabeled Positives in the Training Sets

	ROCAUC	*pROCAUC*	*PRAUC*	*TPR*	*FPR*	*f1*	*Accuracy*	*Precision*	*UTP/UP/FPR*
LR	0.5757	0.0317	0.0532	0.8799	0.8642	0.0743	0.1649	0.0391	1.1890
SAR-EM	0.5686	0.0330	0.0475	0.7893	0.7080	0.0792	0.3111	0.0418	1.2614
SCAR-KM2	0.5895	0.0390	0.0545	0.8350	0.7349	0.0812	0.2864	0.0428	1.2937
SCAR-TIcE	0.5897	0.0391	0.0544	0.8162	0.6999	0.0827	0.3195	0.0436	1.3220
SCAR-Elkanoto	0.5896	0.0390	0.0544	0.8293	0.7213	0.0818	0.2995	0.0431	1.3043

Table 3 Average Prediction Performance for Unlabeled Positives in the Testing Sets

	ROCAUC	*pROCAUC*	*PRAUC*	*TPR*	*FPR*	*f1*	*Accuracy*	*Precision*	*UTP/UP/FPR*
LR	0.5902	0.0354	0.1781	0.8729	0.8481	0.2133	0.2420	0.1248	1.2533
SAR-EM	0.5755	0.0305	0.1582	0.7630	0.6741	0.2216	0.3786	0.1299	1.2872
SCAR-KM2	0.6052	0.0411	0.1824	0.8192	0.7118	0.2270	0.3498	0.1326	1.3215
SCAR-TIcE	0.6052	0.0411	0.1822	0.7880	0.6729	0.2282	0.3807	0.1340	1.3405
SCAR-Elkanoto	0.6050	0.0409	0.1822	0.8069	0.6970	0.2272	0.3619	0.1327	1.3259

Table 4 Paired t Test Results of the Training Sets Average Detection Performance

	ROCAUC	*pROCAUC*	*PRAUC*	*TPR*	*FPR*	*f1*	*Accuracy*	*Precision*	*UTP/UP/FPR*
LR, SAR-EM	2.8956**	2.0042	2.6056*	2.9516**	2.6232*	-0.1051	-2.5843*	-0.2871	-0.2900
LR, SCAR-KM2	-5.2891***	-2.6170*	-3.2704**	1.3652	2.1240	-2.7100*	-2.1328*	-2.3220*	-2.1262
LR, SCAR-TIcE	-5.4545***	-2.6906*	-3.1726**	1.3118	2.0951	-2.7612*	-2.1031	-2.3953*	-2.1653*
LR,SCAR-Elkanoto	-5.4209***	-2.6769*	-3.0932**	1.5752	2.4368*	-2.9505**	-2.4463*	-2.5835*	-2.4469*
SAR-EM, SCAR-KM2	-3.9930**	-2.9850**	-3.1252**	-2.3588*	-0.8178	-2.2814*	0.7411	-2.0310	-2.2005*
SAR-EM, SCAR-TIcE	-4.0341**	-3.0357**	-3.1231**	-3.4175**	-0.5960	-2.2291*	0.5006	-1.9748	-2.2356*
SAR-EM, SCAR-Elkanoto	-4.0424**	-3.0412**	-3.1433**	-2.8087**	-0.5871	-2.4493*	0.4981	-2.1883*	-2.5665*

Figure 4 PR AUC (for Detection) through Time

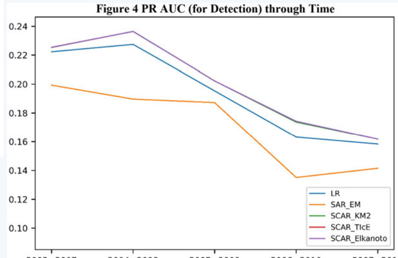
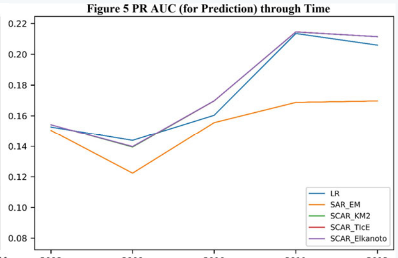


Figure 5 PR AUC (for Prediction) through Time



Conclusions

- We demonstrate that when the restatement dataset contains unlabeled positives, PU Learning is better than supervised learning in detecting and predicting restatements, especially those unlabeled positives.
- This paper adds to the line of accounting research that adopts machine learning by introducing a methodology – PU Learning – to deal with datasets that have unlabeled positives. The finding from this research can inspire future accounting research to adopt semi-supervised learning in studying accounting events that similar features, such as fraud and accounting misconduct.
- This paper sets an example on how to properly set up ML experiments when the dataset has unlabeled positives.

RUTGERS
Rutgers Business School
Newark and New Brunswick

Do Auditors Respond to Cues that Provide Indications of A Heightened Risk of Material Misstatement for Fair Value Measurement?

Helen Brown-Liburd, Dereck Barr-Pulliam, Stephani Mason, Meehyun Kim

Background and Introduction

- Fair value measurement involves subjective judgments, with greater uncertainty regarding the reliability of the measurement process. Thus, auditing fair value measurement is challenging for auditors. (Public Company Accounting Oversight Board 2007a, 2007b; Bratten, Gaynor, McDaniel, Montague, and Sierra 2013; Christensen, Glover, and Wood 2012; Glover, Taylor, and Wu 2016)
- Because of the complexity, relevant cues may go unnoticed if auditors see them as irrelevant pieces of information and do not effectively process the information into their risk judgment. Therefore, auditors experienced in fair value measurement and motived to include relevant cues into their problem statement (Hammersley 2006; Griffith 2018) are more likely to appropriately assess the risk of material misstatement.
- Therefore, we examine whether auditors respond to cues that provide indications of a heightened risk of material misstatement for fair value using text mining techniques. We compare cues against auditors' assessment of the risk of material misstatement and nature and extent of planned audit procedures.

Experimental Design

- We use a 3x2 factorial design where we manipulate the audit client's choice of specialist at three levels (e.g., in-house, another audit firm, or a consulting firm) to prepare a material fair value measurement between-subjects and the financial instrument context at two levels (either asset/liability, gain/loss, held to maturity/settled soon, or trading/held-to-maturity) within-subjects.
- We leverage the strength of experimental methods by holding constant the characteristics of the financial instrument (i.e., a publicly traded fixed-rate security) for which there is a fair value change and focusing on its position in the balance sheet (asset or liability) or realization time frame(hold to maturity or sell/settle soon).
- Participants review a case that includes a client's investments and/or debt issuances with the related fair value measurements.
- Participants were asked the following questions:
  1) The most important factors that influenced your assessment of inherent risk
  2) The most important factors that influenced your assessment of control risk
  3) What factors you considered in arriving at your judgment regarding the risk of material misstatement for the investment in convertible notes
  4) What factors you considered in arriving at your judgement regarding the risk of material misstatement for the bonds payable
  5) Planned audit procedures
- Participants are 499 auditors from 20 firms.
- Questionnaires were mostly completed via hard copy; a few were completed online.

Textual Analysis

- Keywords for textual analysis
  - Direct cues are based on appendix A of AS 2501(revised), standard for financial instruments
    - Direct cue 1-terms and characteristics of the financial instruments
    - Direct cue 2-extent to which the fair value of the type of financial instruments in based on inputs that are observable directly or indirectly
    - Direct cue 3-other factors such as credit/counterparty risk, market risk, liquidity risk
  - Indirect cues are based on AS 2501(revised), which is for all accounting estimates
    - Indirect cue 1-understanding processes used to develop estimates
    - Indirect cue 2-brainstroming(e.g., change in method and so on)
    - Indirect cue 3-risk factors for identifying significant accounts and disclosures

Textual Analysis Process

Rutgers Business School
Newark and New Brunswick

Expected Loan Loss Provisioning Using a Machine Learning Approach

Nichole Li, Alexander Kogan

Background

•Accounting estimates are a critical part of financial statements. Most companies' financial statements reflect accounts or amounts in disclosures that require estimation. Accounting estimates are pervasive in financial statements, often substantially affecting a company's financial position and results of operations

• Banks are crucial for financial stability. Due to the na- ture of their assets (i.e. mainly loans) and their finan- cial structure (i.e. highly leveraged and financed large- ly via deposits) they have specific information asym- metry problems with stakeholders different to those they may have with share- holders.

•Estimating expected Loan Loss Allowance (LLA) in banks is a critical but also difficult problem in accounting estimates. The issue has become of increasing interest to academics and regulators with the FASB and IASB issuing new regulations for loan impairment.

•However, till now, few studies have looked at the application of machine learning in managerial subjective estimates, especially in the LLA. And no published research has been done to model and predict loan losses using other types of machine learning algorithms than regression so far. Therefore, in this paper, I want to fill in the gap by using multiple machine learning algorithms to model and predict loan losses in banks.

Measuring Loan Losses

•Interest income is recognized over time and is derived from a yield that includes at least four components: the time-value of money, expected loan losses, risk premia, and economic profit (Harris et al., 2018). Measuring expected losses is particularly complex, but the loan loss allowance and provisions estimated by managers are based in part on a series of primary indicators, many of which are available in public disclosures. So, in this research, I focused on these primary indicators below in constructing an alternative summary measure of the expected loan losses.

•Loan Balances and Loan Composition. Characteristics of the borrower and of the collateral, affect both the probability of default and the loss-given-default. In my research, I include the proportions of the three largest loan categories: real estate, commercial and industrial (C&I), and consumer.

•Loan Duration

•Nonperforming Loans. Loans that are not paying interest or principal due to a borrower's credit problems are classified as nonperforming loans ,which include nonaccrual loans, restructured (troubled) loans, and some past-due loans.

•Net Charge-offs. Net charge-offs (NCOs) are measures of realized loan loss in a given period and indirectly impact the balance sheet and income statement through the ALLL and the PLLL

Variable Definition

Independent Variables	Definition
Bank Variables	
Log Assets	Log of Total Assets
Total Loans	Total loans in banks portfolio
Loans to Assets	Ratio of loans to assets
Securities to Assets	Ratio of the securities to assets
NCO	Net charge-offs
Four-Qtr NCO + NPL	Sum of rolling four-quarter Net charge-offs plus ninety-days past due and non-accrual loans at the end of the rolling window's fourth-quarter
Pct Four-Qtr NCO + NPL	Four-Qtr NCO + NPL scaled by total loans at the beginning of the quarter
Charge-offs	Amount that is charged-offs
Recoveries	Amount recovered in previously charged-off loans
Allowance	Loan Loss Allowance
Pct Allowance	Loan Loss Allowance scaled by total loans at the beginning of the quarter.
Pct RE Loans	Real estate loans as a percentage of total loans.
Pct CI Loans	Commercial and Industrial loans as a percentage of total loans.
Interest Receivables	Income accrued but not yet collected on loans
Loan Yields	The ratio of tax-equivalent interest income divided by total loans
External Environment Variables	
Unemp Rate	Unemployment rate
HPI	Home price index
HPI Growth	Home price growth
Inflation	Personal consumption expenditure growth in the previous year
GDP	GDP level in the previous year
GDP Growth	GDP growth in the previous year

•

Methodology

Based on the discussion above, the manager would estimate loan loss allowance by predicting future loan losses to be realized in subsequent periods. The accuracy of the prediction will be assessed by how well the estimated allowance captures actual net-charge-offs. This idea is consistent with the accounting identity:

$$\begin{aligned} LLA_t &= LLA_{t-1} + LLP_t + RECOV_t - CO_t \\ &= LLA_{t-1} + LLP_t - NCO_t \end{aligned}$$

Where LLAt is allowance for loan losses at end of t, RECOVt is the amounts that have previously been charged off but are recovered during this time period, COt is the gross amount of all loans charged off against the LLA losses. The income statement effect is captured by LLP, the loan loss provisions. Thus, the precision of the LLAt is assessed by how well it predicts future net charge-offs at t.

Sample and data

Following Harris et al. (2018), I focus on bank holding companies (BHCs) and extract accounting data from regulatory consolidated financial statements (FR Y-9C reports) for the period 1996–2017. The sample period starts from 1996 because information required for measuring certain FR Y-9C variables is unavailable before then.

Dependent variables

To predict the losses, I want to follow Fillat and Montoriol-Garriga (2010) and consider the sum of rolling four quarter net charge-offs, and add non-performing loans at the end of the rolling-window's fourth quarter, which is the dependent variable in my study. At time t the loss is measured as,

$$CL = \sum_{\tau=T+1}^{t+4} NCO_\tau + NPL_{t+4}$$

where NPL is Non-performing loans, which are defined as loans past due more than 90 days and nonaccrual loans (i.e., loans on which a bank has ceased to accrue interest).

Independent variables

My independent variables (predictors) consist of information already known at the time of estimation. Inspired by Vijayaraghavan(2019), I include two sets of independent variables. One set is the bank variables that contains the characteristics of the banks themselves. Another set contains the exogenous environmental variables that reflect the macro-economic factors that may influence the loan loss estimation. The independent variables I want to include are shown in the table above.

Machine learning algorithms

Based on prior literature, I try five different machine learning algorithms in my research: Lasso regression, support vector machine, random forest, artificial neural networks and gradient boosting machine to make the prediction.

RUTGERS
Rutgers Business School
Newark and New Brunswick

Continuous Monitoring with Interactive Data Visualization: An Application to Healthcare Payroll Process

Guangyue Zhang, Hilal Atasoy, Miklos Vasarhelyi

Design Science Research

- Build and evaluate artifacts to meet business needs, and the goal is utility (Hevner et al., 2014)

Research Question

1. Identify unique business problem

- Improve assurance over business processes with proactive and intelligent continuous monitoring
- Alleviate information overload for management

2. Design artifacts

- A systematic dual-purpose CM artifact design within a framework
- Guide management to analyze risk factors, formalize control checks, and utilize emerging data analytics technologies

3. Integrate the artifact within the technical infrastructure of the business environment and evaluate the design

- An implementation to a hospital payroll process to obtain facts on management reaction and the incremental economic value

CM design framework

1. Stage 1: Risk assessment and control check definition

- Determine target business process, reconsider significant risk factors, and reorganize existing control checks

2. Stag2: Data preparation

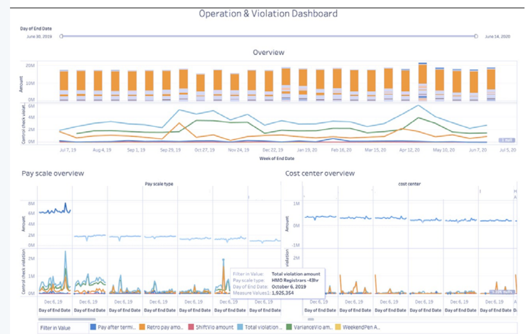
- Understand relational database's structure of the information system in use
- Preprocess system data extraction: removing missing values, correcting data format, and merging data from different sources

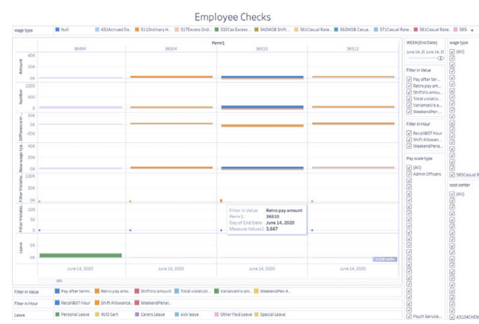
3. Stage 3: CM model development

- Risk reassessment with operation & violation dashboard
- Anomaly detection dashboard for notable item identification
- Materiality and inspection scope confirmation & Notable item examination

4. Stage 4: CM model implementation and evaluation

- Verify the functions and reengineer the control monitoring process





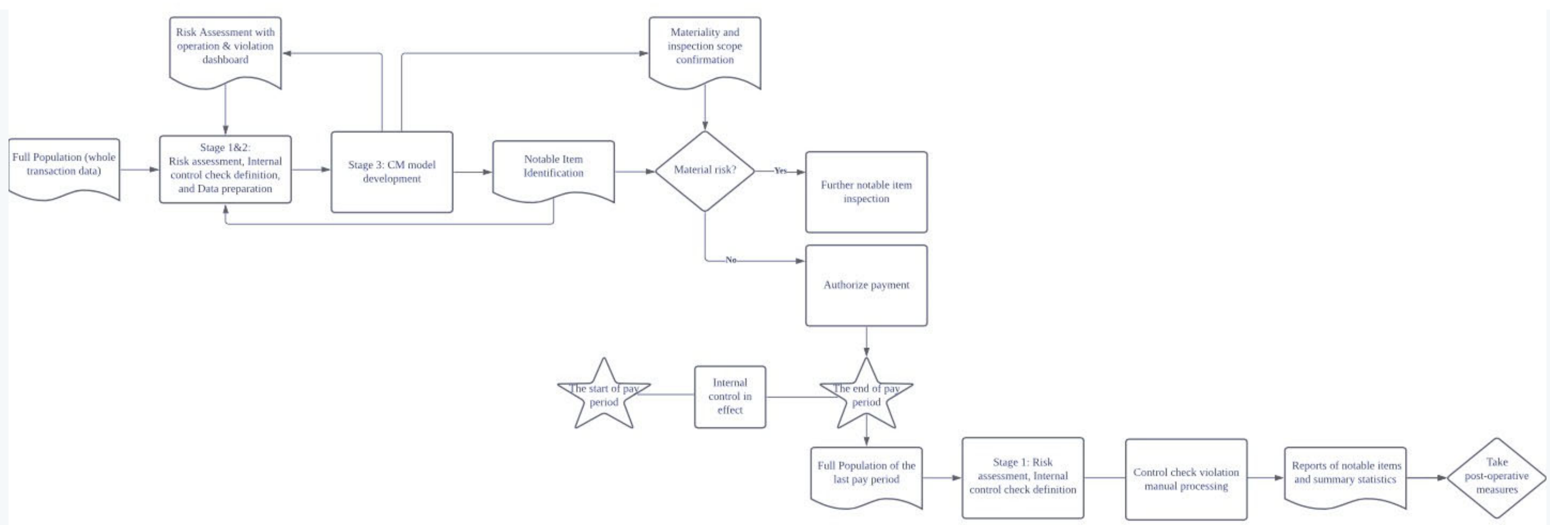


Illustration of the CM design

- We work with an audit firm and one of its hospital clients in Australia to improve the level of assurance over the payroll system for FY 2019 with the proposed CM framework
- Our hospital partner is highly motivated to implement the CM framework
- Their payroll system is very representative of the unique challenges hospital payroll systems face
- The current payroll service still heavily relies on time-consuming manual data processing. Risk assessment and anomaly detection only complete after the end of the pay period when payments are all finalized, and management can only take post-operative measures
- Preliminary evaluation: emerging technologies substantially enhance the efficiency and timeliness of risk assessment, anomaly inspection, loss prevention, and business operation

RUTGERS
Rutgers Business School
Newark and New Brunswick

Audit with Machine Learning: Applying an Unsupervised Algorithm on Journal Entries of an Australian Bank

Danyang (Kathy) Wei, Soohyun Cho, Miklos A. Vasarhelyi

Introduction

- The nature of audit determines that it is an outlier detection process under a specific context. On the other hand, the effectiveness of machine learning on outlier detection has been proved in fields such as chemistry and medicine. However, extant literature on using machine learning for audit purpose has not solved several challenges. First, how to fit data analytics with machine learning in audit context is still uncertain and ambiguous. Second, consequence of applying machine learning is still unclear. Third, auditors have difficulty to document audit evidence from machine learning due to its unobservable working process.
- Taking the concerns into account, we proposed a methodology called audit with machine learning (AML) that can serve as a complementary tool in journals testing. It identifies entries with potential risk of misstatement among full population and convert machine learning results into understandable fact to auditors who will make the final decision. In addition, to address large number of outliers, AML is not identifying every risky entry but potential risks with a sample of entries so that auditors can determine if the risks are material or not. If auditors do find the risk worth their attention, they can develop a series of rules to extract all the related entries by themselves.
- AML includes three stages: outlier pool formation, machine learning application, and interpretation. Auditors will apply judgment in the first and the third stage to make sure the final result is audit-related. Meanwhile, a framework is provided in the third stage to help auditors interpret the results of machine learning. After introducing AML, we applied it on entries of an Australian bank and reported feedback from its external auditors.

Methodology

- We argue that the variables of entries play different roles in audit analysis. Therefore, we categorized them into two groups: standard and add-on variables. Standard variables are variables that are directly related to economic substance and less likely to be modified for management purpose (i.e., net amount), while add-on variables are developed for management purpose and may be modified to better fit management objectives (i.e., user id). The two groups are separately utilized in AML.
- There are three steps in AML (Figure 1). The first one is outlier formation where auditors identify transactions that worth their attention based on experience with a certain client. This process could be achieved through rule-based approaches (i.e., transactions over $1 million). Standard variables are used here. The second step is machine learning application. We adopted an unsupervised algorithm called Local Outlier Factor (LOF) to assign each entry in the outlier pool a score. LOF is an effective algorithm for outlier detection. The larger the score is, the higher likelihood the entry has to be an outlier. The last step is interpretation. Add-on variables are used here and we further divided them into three subsets: local, global, and complementary. The separation is based on how you group the entries. We argue that LOF should be applied on entries that are similar to each other based on given variables so that any entries that are different can be treated as outliers. One example could be grouping entries with primary categories. Then, if the values of an add-on variable within one group are overlapping with the values of that variable outside that group, then it is a global variable. In contrast, if the values of an add-on variable within one group are not overlapping with values outside the group, it is a local variable. Complementary add-on variables are the textual variables such as journal description since they contain too much noise to be used for data analytics. With the framework illustrated as Figure 2, a subset of transactions in the outlier pool will become the exceptional transactions which are prioritized for investigation. It will solve the "alarm flood" for auditors because they do not need to investigate each entry in the outlier pool.

A Case Study—An Australian Bank

- For experiment, we applied AML on journal entries of an Australian bank (the Bank). The data is all entries in Fiscal Year 2019. The original variables are in Table 1. After discussing with the Bank's external audit team, we selected five primary categories for experimental purpose: cash and cash equivalents; creditors, accruals and settlement accounts; depreciation and amortization; derivative assets held for hedging purposes; employees' compensation and benefits. The entries pulled out into the outlier pool were selected by the audit team. The variable categorization result based on our definitions is in Table 2.
- After AML, the final result a list of exceptional transactions and the reasons why they were identified. In Table 3, we reported a summary that includes the number of exceptional transactions identified under each pre-defined reason based on each primary category. The external audit team had positive reactions to the result. They believe that AML will become an effective complementary tool in journals testing and help them obtain extra insights about the client. The framework at the interpretation stage also makes the AML process much more understandable to auditors.
- AML contributes to extant literature in three aspects. First, it provides an approach to address a large number of outliers when using data analytics on the full population. Second, the interpretation process allows auditors to document audit evidence generated from machine learning through a series of predefined explanations. Third, AML leaves enough flexibility to auditor's judgment through the decisions about the threshold for "abnormal." However, this research also has some limitations. First, AML has only been applied on the Bank's entries. Therefore, future experiment is necessary to draw a solid conclusion on its effectiveness. Second, we are not able to validate if the exceptional transactions identified are riskier than the remaining transactions because it is almost impossible for auditors to assess the risk of misstatement for every single entry. Therefore, future research can focus on the validation process of audit data analytics.

Figure 1. Three Stages of AML

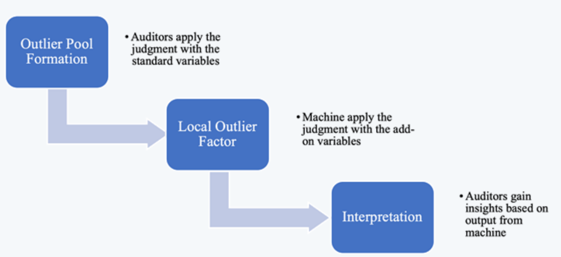


Figure 2. Framework for AML Interpretation Stage

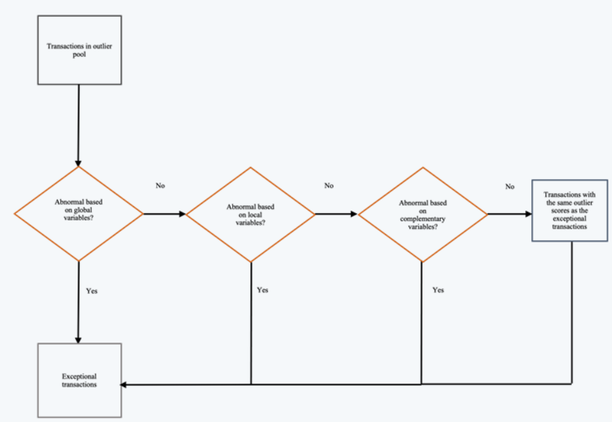


Table 1. Original Variables of the Bank's Ledgers

Variables	Description	Type
Account Code	General ledger account ID (corresponding to the "Account Number" in TB)	Categorical
Transaction Id	Journal entry transaction ID (not unique)	Categorical
Net	Net debit/ credit amount posted by the journal line	Numerical
Effective Date	Date the transaction occurred	Date
Created Date/ Time	Date the transaction was processed into the system	Date
Document Type	Type of transaction and source that was posted	Categorical
User Id	Individual or System ID that entered the transaction	Categorical
Journal Description	Narration for the journal	Textual
Line Description	Specific narration of the line of the journal	Textual

Table 2. Result of Original Variable Categorization

Variable Group	Variables in Original Data
Standard variables	Net amount; effective date
Global add-on variables	Transaction id; created date; document type; user id
Local add-on variables	Account code
Complementary add-on variables	Journal description

Table 3. Summary of Exceptional Transactions in the Selected Five Primary Categories

Primary Category	Abnormal Document Type and User Id	Unlikely Account Code	High-Risk Account Code	Suspicious Journal Description	Similar Outlier Score	Number of Exceptional Transactions
Cash and cash equivalents	4	7	5	0	0	**16**
Creditors, accruals and settlement accounts	0	2	0	0	4	**6**
Depreciation and amortization	0	1	0	0	4	**5**
Derivative assets held for hedging purposes	0	1	8	0	0	**9**
Employees compensation and benefits	0	3	0	0	3	**6**

RUTGERS
Rutgers Business School
Newark and New Brunswick

Channel Stuffing Detection Model: Using Processing Mining

Huijue Duan, Jumi Kim, Qing Huang, Miklos A. Vasarhelyi

Instruction

Sales activity management is a big challenge for multinational corporations, especially corporations in manufacturing industry. Since these corporations usually have so many branches around the world, and different branches may have different sales strategies, policies, SAP systems, etc. So it difficult for management and internal control department to monitor and compare different branches' performance and compliance.

This study aims to develop a channel stuffing detection model, as well as a continuous monitoring system to ensure that all sales are being recorded to the correct period, sales activities are conducted in accordance with the company's policy.

Methodology

This methodology is built on a case of one multinational manufacturing corporation's order to cash process (OTC):

- First, setup the data model and data connections in an process mining and execution management software. This is the foundation of all the further analysis, because this corporation has more than one SAP systems. Different data is stored at different systems, and there's no or rare connections between systems. Data model can help to build the relationships between data.
- Then, some key performance indicators (KPIs), key risk indicators (KRIs), and process models will be developed based on this corporation's OTC process.
- Later, outlier detection technology will be used to identify potential abnormal data and/or process model.
- Finally, a continuous monitoring system which include all the channel stuffing methods mentioned above will be built.

Contributions

This study combines the data analytics, process mining, machine learning, and continuous monitoring together.

- It will provide the management a holistic and detailed view of the OTC process and sales activities, visually to see the problematic area and unusual trend/patterns of transactions, and identify the areas with high control risk as well as high operational risk.
- This continuous monitoring dashboard can also improve auditors' performance in risk assessment and deciding the nature, timing, and extent of the audit procedures.

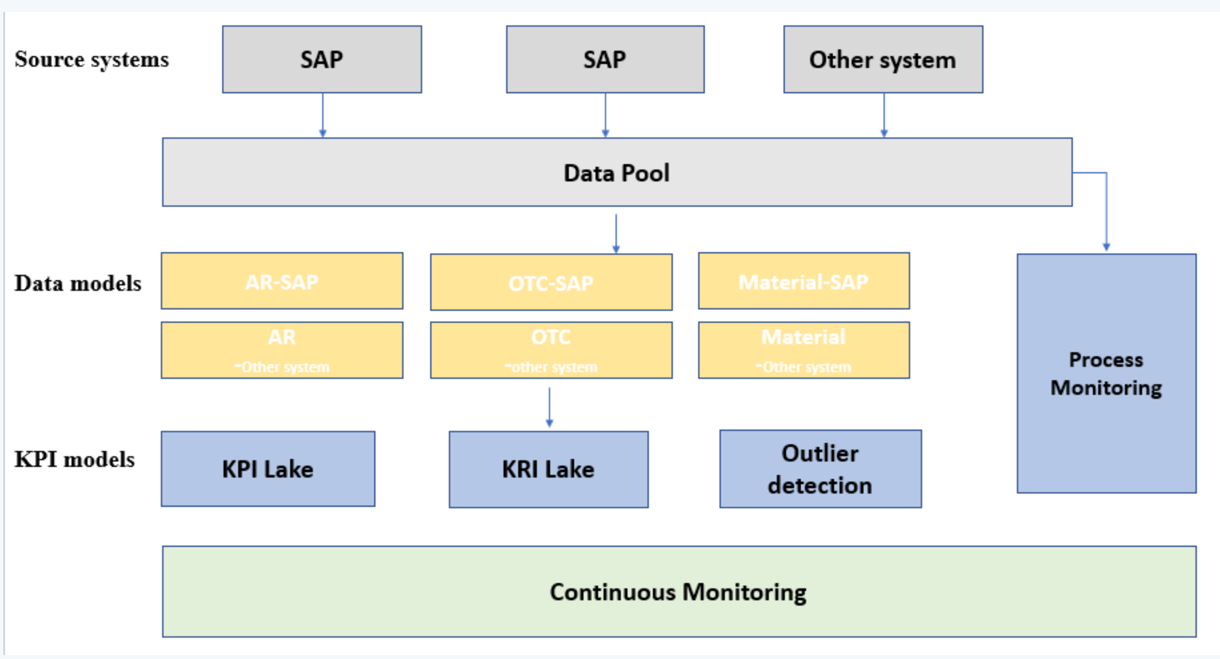


Figure 1. Channel stuffing diagram

RUTGERS
Rutgers Business School
Newark and New Brunswick

Cybersecurity Risk Disclosure During the COVID-19 Pandemic

Qingman Wu, Won Gyun No

Introduction

- With the rapid development of Information technology (IT), firms are relying on technology almost in every aspect of their business operation. Because of this, there are more and more cybersecurity breaches happened in firms' daily activities. Breached firms experienced substantial loss of revenues, customer base, reputation, and business opportunities, and most of the breached firms spent millions of dollars improving security solutions and expanding security procedures following the attacks.
- Since the outbreak of Coivd-19 in January 2020, people are required to work from home, which significantly reduce the cost on commuting but also brought other risks. Cybersecurity Ventures' estimation that cybercrime damage costs could potentially double during the Coronavirus outbreak period is concerned not only with phishing scams, but also with ransomware attacks, insecure remote access to corporate networks, remote workers exposing login credentials and confidential data to family members and visitors to the home, and other threats. According to many surveys and prior works, firm costed more to handle their cybersecurity risks during the Covid-19 pandemic than before.
- This study aims to find out whether firms provide more cybersecurity risk disclosures related to the remote work environment and the pandemic after the outbreak of the Covid-19, following the guidance from the SEC.

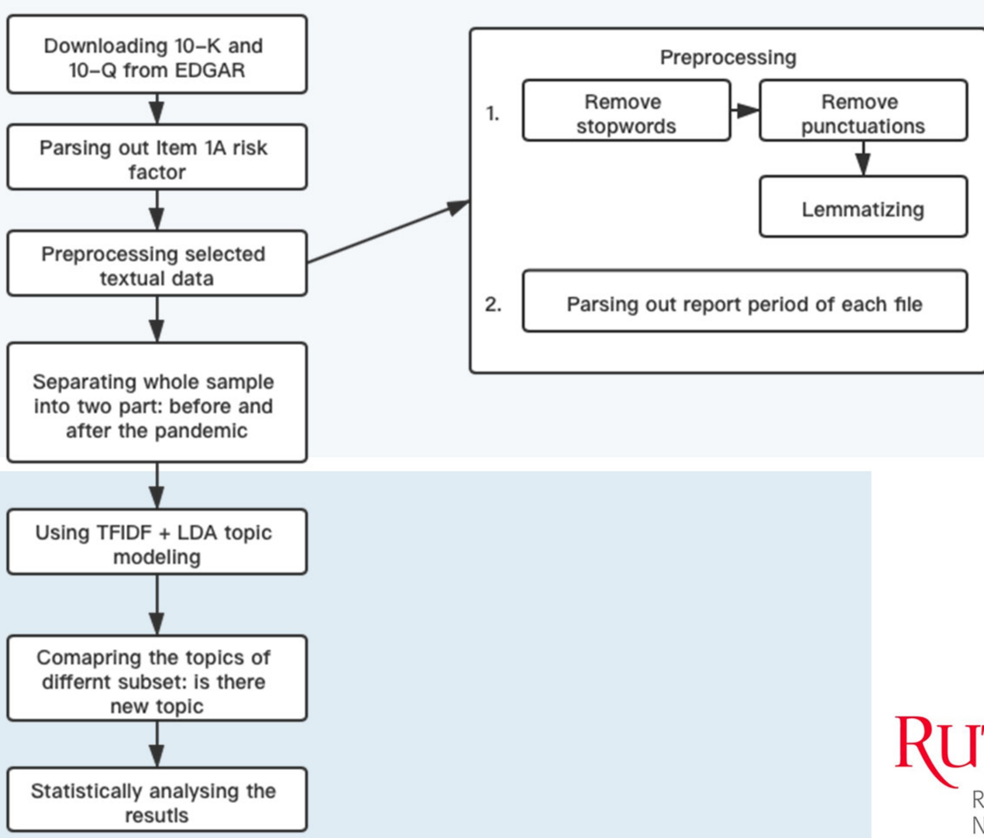
Literatures

- There are two different views on risk-factor disclosures in the 10-K files. One stream of studies claimed that the risk disclosure is informative, which gives investors better insight of companies' risks (Gordon et al., 2010; Kravet and Muslu, 2013; Campbell et al., 2014). Other studies found results that most part of risk disclosures have no impact on investors' risk perception, which indicate the limit informativeness for users (Bao and Datta, 2014).
- For firms' cybersecurity risk disclosures, on the one hand, providing cybersecurity related disclosure could reduce information asymmetry and provide stakeholders with better insight of cybersecurity risk detection and prevention strategy, which could result in positive market reaction (Gordon et al., 2010; Berkman et al., 2018). On the other hand, prior literatures found results supported that firms would strategically provide cybersecurity risk disclosures to the public to avoid the negative impact of those information (Amir et al., 2018; Li et al., 2018; Gao et al, 2020).
- According to the surveys and studies, the volume, severity and costs of cybersecurity incidents are increasing sharply during the pandemic (IBM, 2021; Google, 2021; Gartner, 2021; Rock et al., 2020).

Hypothesis

- Hypothesis 1: firms disclose more information related to the cyber-attacks they suffered, like phishing, ransomware, the third party risk.
- Hypothesis 2: firms provide more risk-mitigation information to mitigate the negative impact of the cyber incidents they suffered.

Research Method



Rutgers Business School
Newark and New Brunswick

Measuring Text Information of Green House Gas (GHG) Emission

Dong Gil Kim and Won Gyun No

Background

- The California Global Warming Solutions Act of 2006 or Assembly Bill 32 (AB 32) requires reporting of greenhouse gas (GHG) emissions by facilities located in California
- Mandatory reporting first started on January 1, 2008 and it is about financial disclosure
- The goal of AB 32 is to reduce GHG emissions in California to the 1990 level by 2020
- Starting on January 1, 2010, the Final Mandatory Reporting of Greenhouse Gases (GHG) Rule issued by the EPA will require mandatory reporting of GHG emissions from large sources in the U.S.
- Therefore, California firms needed to report GHG in 2008 and 2009 while non-California firms did not, which is our variable of interest

Hypothesis

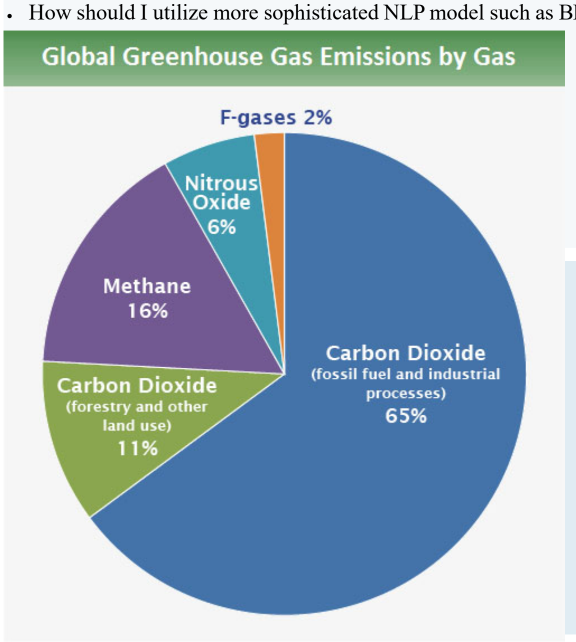
- Main hypothesis: mandatory greenhouse gas reporting law, which is about financial information, also lead to non-financial disclosure about GHG/Carbon
- Imagine firms have high GHG emission. They have to reveal the number of emission by mandatory law
- They may not want to reveal the information since it could cause bad image to stakeholders. If firm concerns this, they may want to explain/excuse the high emission with non-financial information
- Of course, they may not care at all and that is why they have high emission. In that case, they may not disclose additional non-information to stakeholders
- If some firms have low GHG emission, they may want to appeal more their low emission with non-financial information
- Or they may not since they think their financial information(number of emission of GHG) enough.
- Verifying the hypothesis shows whether non-financial information complements or substitutes financial information

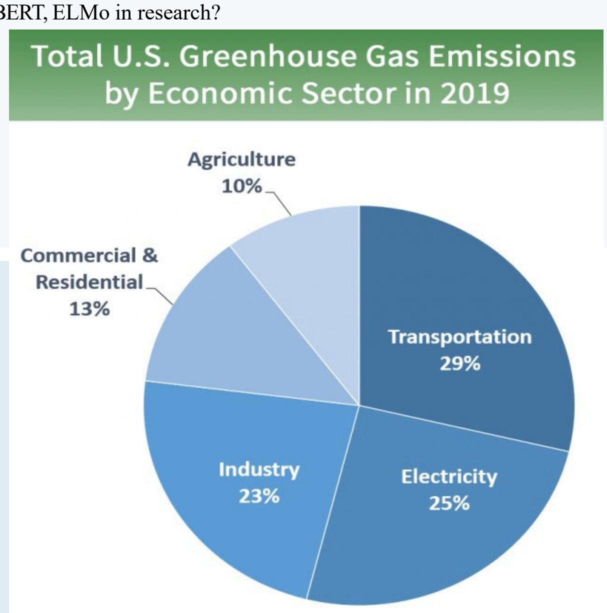
Measuring non-financial information of GHG/Carbon

- Data: MD&A section in 10-K, parsing sentences which contain keywords related with GHG/ Carbon
- Easiest way is just counting total number of the sentences parsed or use growth of it ex) ΔNon_Financial_GHG = change in number of sentences related with GHG in 10-K
- There could be more creative ways to quantify non-financial information
- Using parsed sentence, you can calculate cosine similarity between at year t and t-1. The higher cosine similarity, the less non-financial information disclosed by the firms
- Applying LDA topic model to the parsed sentences, calculate number of topics. The more topic diversity, the more non-financial information disclosed by the firms.
- Since each measure has limitation, developing effective measure, that is how to quantify meaning ful non-financial information is the key

Future research:

- Divide firms as three groups(High, Medium, Low) based on GHG emission and investigate how firm discloses non-financial information strategically
- Does firm with high GHG emission disclose more non-financial information about GHG to greenwash their image?
- Is there different degree of disclosure of non-financial information across different industries ex) Does auto industry disclose more non financial information about GHG more than other industry?
- How does government subsidy ex) growth of electric car market affect to disclosure of non-financial information?
- Measuring more sophisticated non-financial information of GHG/Carbon with criteria such as Risk Disclosure, Investment, Sentiment, Governance and Stakeholders
- Using Word2Vec, training parsed sentences or paragraphs from 10-K to embeddings and utilize it for research
- How should I utilize more sophisticated NLP model such as BERT, ELMo in research?





Rutgers Business School
Newark and New Brunswick

Impact of Business Analytics on Managerial Accounting

Wenru Wang, Miklos A. Vasarhelyi

Introduction

Over the years, the role of management accountants has expanded from historical reporting to more real-time reporting. Currently, managerial accounting has evolved into the "Predictive Analytics Era," and managers are interested in knowing what has happened and what will happen in the future (Cokins, 2013).

Previous literature discussed the impact of business analytics on managerial accounting and how business analytics measure corporate performance (Appelbaum, Kogan, Vasarhelyi, & Yan, 2017; Nielsen, Nielsen, Jacobsen, & Pedersen, 2014), but little research explored the application of business analytics in management accounting in a government procurement point of view.

Methodology

This paper provides a use case of applying the continuous monitoring methodology and different business analytics techniques to illustrate how business analytics could serve as efficient managerial accounting tools.

An analysis diagram is created to illustrate how the study achieves to answer the three questions that management accountants are interested in, as presented in Figure 1.

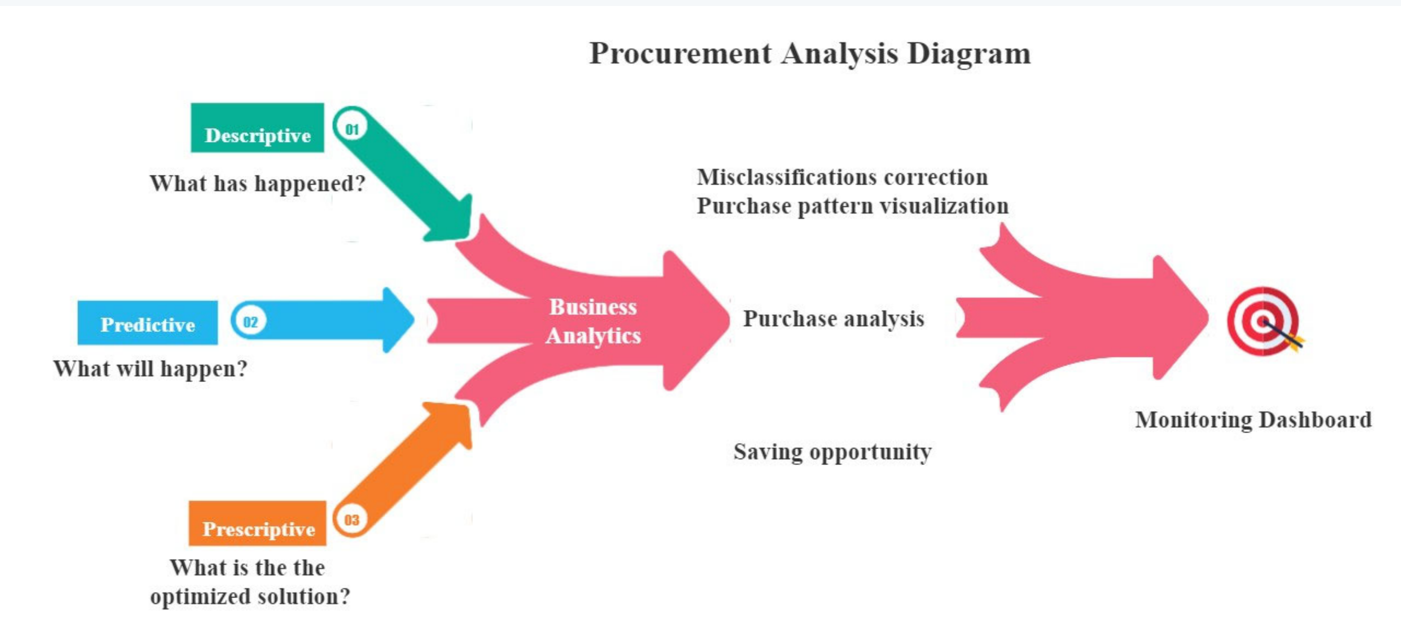
Specifically, we also utilize fuzzy matching, a text mining technique, to enable accountants to search the medication of their interest in the national medication catalog and locate a supplier that provides a better price but does not serve the city before.

Main Contributions

First, it introduces different business analytic techniques such as continuous monitoring and visualization into the field of managerial accounting under a government procurement setting.

Second, this paper provides evidence of applying continuous monitoring and other business analytic techniques to analyze aggregate government procurement data efficiently. The proposed methodology could identify and data misclassification, generate purchase patterns, and detect possible procurement wastes from thousands of procurement records with little human intervention.

Third, this paper highlights the usability of different business analytics tools. Business school curriculums could be inspired by the similar or different analyses that are achieved by the analytics tools used in the paper.



Rutgers Business School
Newark and New Brunswick

Impact of the Remote Working Environment on Cybersecurity Risk in Organizations

Hongmin (William) Du, Won Gyun No

Background

- Cybersecurity is the way that organizations can protect their information, intangible assets, and communications while operating within the cyberspace. Since cybersecurity issues have seen multiple SEC regulations over the years, they have begun to become more important to companies as regulation on cybersecurity disclosure becomes stricter.
- In 2019, with the start of COVID-19, almost all company operations were moved to an virtual environment including online meetings , online transactions, and conducting overall online business.
- Software's in the remote working environment include telecommunication services such as WebEx, Skype, and Zoom.
- We want to test whether the increase in the use of the remote working environment has increased cybersecurity risk of firms.

Hypothesis

- Our first hypothesis involves whether the remote environment has introduced new security concerns to organizations.
- Hypothesis 1: Cybersecurity Firms issue more cybersecurity risk disclosure in their 10-K relating to the use of the remote working environment.
- Our second hypothesis involves the risk space controlled by the internal controls of an organization. Typically organizations have internal controls put in place to monitor alleviate risk within the organization; however, the remote working environment is considered in a external environment outside of the controls of a firms typical internal controls. If we find an increase in cybersecurity breach after the increase in use of the remote working environment, we can infer that the remote working environment introduces more risk to organizations.
- Hypothesis 2: The remote working environment has introduced more security risks to organizations.

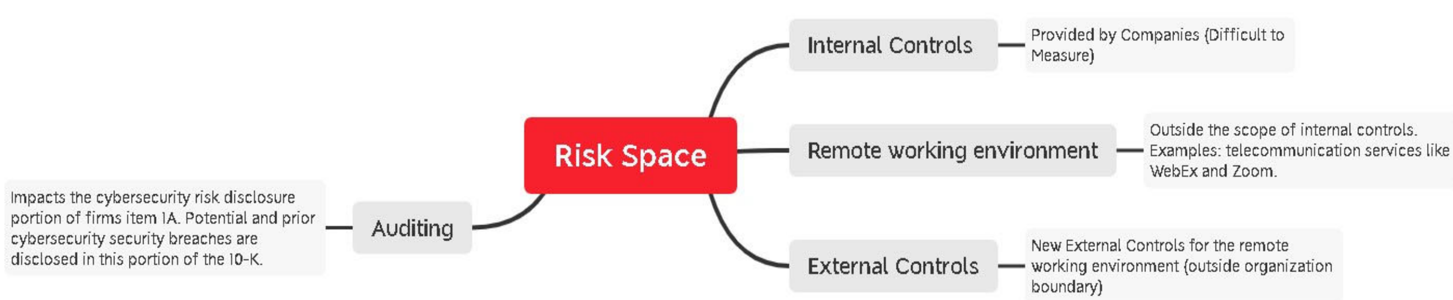
Dataset

- The first dataset that will be used is from NIST – one of the big institutions of science technologies. NIST provides a database called the NVD (National Vulnerability Database). When you go to NVD, you can see a list of vulnerabilities. It is a public database that provides both vendors, and software that have potential cyber security vulnerabilities. For each vulnerability, there is also a risk score assigned by NIST's risk measure for each vendor and software. Using this database, we can find out that the vulnerabilities in the database are more related to the internal controls of an organization, but what we want to find are the vulnerabilities that are introduced by the remote working environment. The vulnerabilities related to the internal controls include the risk that pertain to the servers, systems, and security services provided within the organization.
- The second dataset will be using the 10-K financial dataset of Fortune 500 companies. We limit our dataset to only a small subset of companies and believe that hackers would more likely target a company with more revenue resulting in bigger impact than a company of smaller revenue. From the 10-K we will parse out the Item 1A, and Item 7. From these two sections we will use a selected keyword list to identify sentences related to cybersecurity.

Methodology:

Our dataset will consist of two times. Our first dataset will consist of the period before from 2017 to 2019 and our second dataset will consist from the period from 2019 to the present. We plan to perform a Difference in Difference experiment design to see the change in effect caused by the remote working environment.

There is a keyword list for cybersecurity related terms; however, there are no keyword list actively available for the remote working environment. For this reason, we need to develop a keyword list including terms that are related to the remote working environment. Examples of potential keywords to be included in this list could be online, virtual, remote, and zoom. To create this keyword list, we will use the NIST database to search for the vulnerabilities that are associated with the remote working environment and include them in our list. Using the keyword list related to the remote working environment we can scan the our 10-K dataset for each company for matching keywords to see if companies mention these risk concerns.



Rutgers Business School
Newark and New Brunswick

Predicting the Discontinuity of Non-Profit Organizations Using the Machine Learning Approach

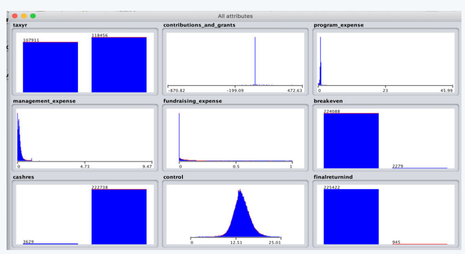
Xinxin Wang , Heejae Lee, and Richard Dull

Introduction

- In the United States, NPO usually defined as the organization granted the tax-exempt status by the Internal Revenue Service (IRS). Organizations are tax-exempt if they meet the requirements of Internal Revenue Code Section 501(c)(3). For accountability and transparency purposes, IRS required NPOs that have gross receipts that are greater than or equal to \$200,000 or \$500,000 in total assets to fill form 990 annually (Internal Revenue Service, 2019).
- According to IRS, if the organization is facing the situation as Liquidation, Termination, Dissolution, or Significant Disposition of Assets, it means that NPO is having discontinuity issues and should file Schedule N of form 990 (Internal Revenue Service, 2020).
- Non-profit organizations worked as public service providers deserve a healthier financial environment. It is important to the economic and social benefits that a large number of healthy NPOs keep serving the people and the country. To do that, identifying the financial challenges and the cause of the discontinuity became the focus of this paper. The research question became what are the signs of NPOs' discontinuity by examining the Form 990 database.

Data and Methodologies

- IRS started electronic filing Form 990 since 2011 and made the data from over 1,000,000 E-filing Form 990 available for everyone through Amazon Web Services. Wu and Dull (n.d.) developed a database that transfers the unstructured data to a useful format for analysis.
- The whole dataset contained Form 990 E-filing from the tax year 2012 to the tax year 2015. The original dataset contained more than 1,000,000 records and more than 660 attributes. In this research, I will select the attributes related to financial information and using the feature selection to find out the most important financial indicator for the NPO discontinuity (financial distress/vulnerability, liquidity, and solvency).



Preliminary Results

- **Descriptive Analysis**

After combining the data from 2012 to 2015, we get the final record of 621,899, which includes 447,596 data entry as the training set and 174,303 as of the test set. These are the organizations filed Form 990 and have continuous records in the database. In addition, we remove the duplication and missing values from the dataset. The final training dataset contains data entry, and the ultimate test set contains163,829 records. It is worth noticing that this dataset is imbalanced data; the cost associated classifier might work better in this situation.

- **Machine Learning Classification**

In this section, we applied five different classifiers in our training set and compared the result of different algorithms and then apply the best in the test set. Due to the imbalanced dataset, we will apply the cost-sensitive classifier and the cost matrix set up to 10. This study applied five different machine learning algorithms, Naïve Bayes (NB), Bayesian Network (BBN), Logistic Linear Regression (LLR), Decision Tree (DT), and Random Forest (RF). The results of each classifier presented in Table 4.

Table 1 - Variable Selection

Column Name	Form 990 Match	Description
taxyr	Head A	Tax Year
ein	Head D	Empolyer Indentification Number
finalreturnind	Head B4	Final Return/ Terminated
cycontributionsgrantsamt	Part 1 Item 8	Contribution and Grant(Current Year)
cytotalrevenueamt	Part 1 Item 12	Total Revenue (Current Year)
cytotalexpensesamt	Part 1 Item 18	Total Expense (Current Year)
totalassetseoyamt	Part 1 Item 20	Total Asset (End of year)
totalliabilitieseoyamt	Part 1 Item 21	Total Liabilities (End of year)
totalfunctionalexpensesprogramservicesamt	Part 9 Item 25(B)	Total Functional Expenses Program Service Expense
totalfunctionalexpensesmanagementandgeneralamt	Part 9 Item 25(C)	Total Functional Expenses Management Expense
totalfunctionalexpensesfundraisingamt	Part 9 Item 25(D)	Total Functional Expenses Fundraising Expense
New Variables		
breakeven		If total revenue >= total expense
cashres		Log of Cash and Cash Equivalents
control		Log of Gross Receipt

Figure 1– Regression Model

$$Dicontinuity_{n+1} = F(Breakeven_n, CashReserve_n, Con\&grant_n$$
$$ProgramExpR_n, MgnExpR_n, FundraisingExp_n$$
$$+control) + \varepsilon_n$$

Figure 2– Correlation Table

	contributions and grants	program expense	management expense	fundraising expense	breakeven	cashres	control
program expense	-0.041952						
	<0.0001						
management expense	-0.000294	-0.885172					
	0.9766	<0.0001					
fundraising expense	0.103891	-0.401298	0.010152				
	<0.0001	<0.0001	0.3100				
breakeven	-0.037110	-0.078704	0.067378	0.050309			
	0.0002	<0.0001	<0.0001	<0.0001			
cashres	0.016681	-0.009021	0.003342	0.023323	0.005199		
	0.0953	0.3671	0.7383	0.0197	0.6032		
control	-0.059138	0.019834	-0.011563	0.008696	-0.008833	0.043944	
	<0.0001	0.0473	0.2476	0.3846	0.3771	<0.0001	
finalreturnind	-0.000457	-0.022252	0.025578	0.000575	-0.006895	-0.032566	0.001115
	0.9635	0.0261	0.0105	0.9541	0.4905	0.0011	0.9112

Cell Contents: Pearson correlation
P-Value

Conclusion

Table 2 - Classification Results:

	Accuracy	TP Rate	Precision	Recall	ROC Area
NB	98.15%	0.981	0.992	0.981	0.514
BBN	99.58%	0.996	0.996	0.996	0.5
LLR	99.58%	0.996	0.996	0.996	0.5
DT	99.58%	0.996	0.996	0.996	0.5
RF	99.99%	1	1	1	0.998

This study found the information from Form 990 tax filing of the NPOs can differentiate a pattern that implies the discontinuity of NPOs. Innovatively, we created new variables based on the necessary information of form 990 that would provide good prediction of the NPO bankruptcy. This paper applied five different supervised machine learning algorithms to find the best classification method, the results that Random Forest was the most effective one for the classification. For future research, the author would like to expand the database to include at least three years of data previously. The financial ratios would also be some important financial indicators to look into.

Rutgers Business School
Newark and New Brunswick

Audit 4.0-based ESG Assurance: An Example of Using Satellite Images on GHG Emissions

Yu Gu, Jun Dai, Miklos A. Vasarhelyi

Introduction

- As Environment, Social, and Governance (ESG) information has become an essential resource for investors globally (Amel-zadeh and Serafeim, 2018), regulators are making significant efforts on standardizing its reporting process and assuring its quality.
- While an increasing number of companies have started ESG disclosure, it is usually not fully substantiated with supporting information. Thus, assurance that attests the metrics on ESG reports free of material errors is in urgent demand. However, the existing ESG assurance procedures face some key challenges, such as how to exercise professional skepticism and judgment.
- This study proposes an innovative method to improve ESG assurance base on Big Data and emerging technologies in Audit 4.0 to enable timely and accurate auditing. This study first discusses a novel process of assuring ESG reports via extending the Audit 4.0 paradigm to the ESG domain. Following the process, a case study is conducted to explore new audit evidence of satellite images on the estimation of methane, one major component of greenhouse gas (GHG), and creates a continuous assurance dashboard to reduce the risk of greenwashing.
- This study pioneers the use of satellite images to provide insights for regulators and companies to monitor and assure ESG reports continuously.

Design of Audit 4.0 on ESG

- The concepts of the Audit 4.0 paradigm, originally designed for conducting financial audits (Dai and Vasarhelyi 2016), could be extended to ESG, by piggybacking on the Industry 4.0 technologies to provide efficient, effective, and timely assurance.
- As Figure 1 shows, the paradigm would mainly utilize technologies to capture the relevant information from the physical world and transmit it to the mirror world in which comprehensive analyses are performed. The paradigm would contain four layers, i.e., technology, data, function, and reporting. The technology layer would facilitate data collection, data storage, and data analysis. Metrics in the data layer may differ from those in ESG reports but can be compared by trend analysis in the proposed verifying process in Figure 2. In the function layer, prediction, monitoring, investigation, and annual audit serve as the main functions for continuous assurance. Based on the results of the function layer, timely alerts could be sent to stakeholders. The results of the function layers could also be included in integrated reports.
- In Figure 3, this study further explores the Audit 4.0-based ESG assurance by identifying two specific objectives: 1. verify the ESG reports from two perspectives, numeric data and textual claim; 2. design an ESG Continuous Assurance (CA) model.

Case Study — Satellite Images on GHG Emissions

- We conduct a case study that uses satellites and aircraft to gather exogenous data regarding GHG emissions as a proof-of-concept for the Audit 4.0-based ESG assurance. Greenhouse gas (GHG) emissions are one major subject matter of ESG report topics (ISAE 3410, 2011). Data regarding the emissions (methane concentrations) are collected and estimated by a private satellite firm GHGSAT.
- Figure 4 demonstrates how to visualize the data in satellite map and dashboard in Tableau. The satellite map shows the geographic information of methane emissions over wells. Using the satellite map, auditors can estimate the methane emissions of a particular well. Managers could monitor the wells owned by the company on a real-time basis to improve its performance. The dashboard combines the normal map, satellite map, operator sample, site type, and trend diagrams of average estimated source rate, allowing auditors to interact with the dashboard and investigate the relationship between those graphs.
- As a proof-of-concept case, this paper uses satellite images to assure companies' claims regarding methane emissions in ESG reports. It also shows how to build a dashboard to provide continuous emission analysis and monitoring.

Figure 1. Design of Audit 4.0 on ESG

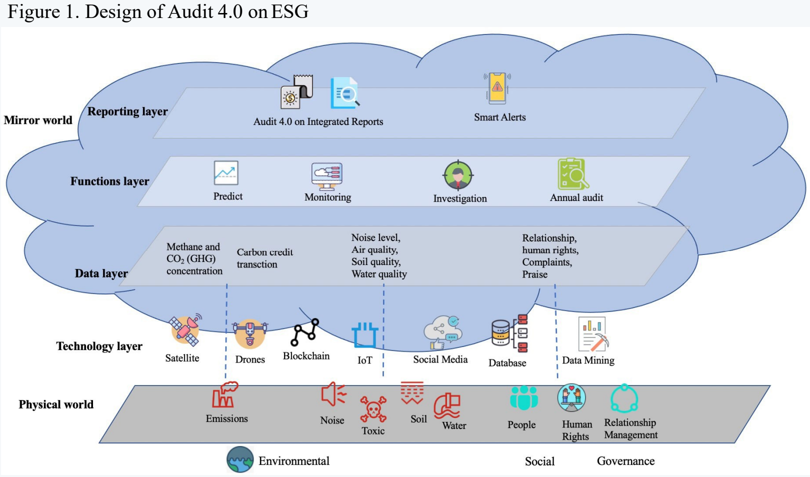


Figure 2. Proposed verifying process

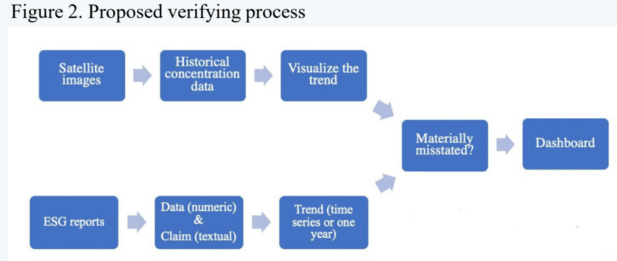


Figure 3. Objectives for ESG assurance

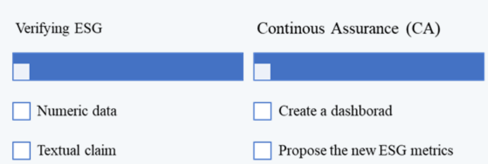
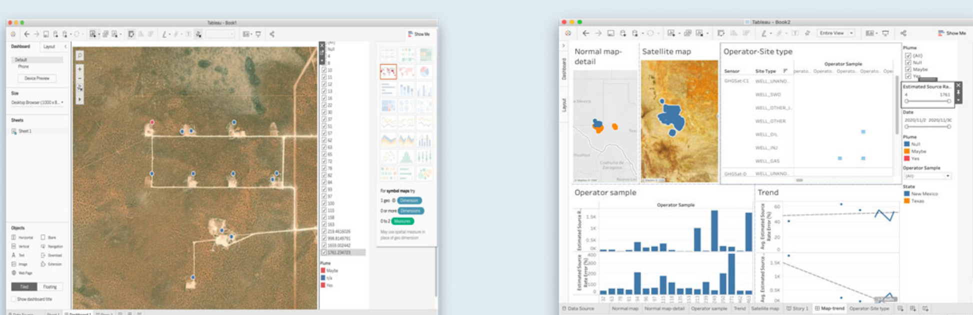


Figure 4. Demonstration of Sample Data in Satellite Map & Dashboard



RUTGERS
Rutgers Business School
Newark and New Brunswick