An academician's perspective on using machine learning to improve audit quality

Partha S Mohapatra

California State University

Introduction

- ML ML has gained substantial traction within audit firms

- ML research in auditing centered around three traditional themes
  - Frauds and Mistatements
  - Bankruptcies and Credit Risks
  - Estimations

- Behavioral Research on ML

- Research opportunities, and Research avenues beyond traditional themes

- What we need to teach to students in accounting programs

Leveraging Machine Learning for Enhanced Fraud Detection, Mistatements

- Identifying accounting and corporate frauds is an important aspect for auditors.

- Corporate Frauds and misstatements are difficult to detect.
  - The reasons for frauds and misstatements can be complex and related to a large set of factors.

- Relative to a linear statistical model, machine learning algorithms are best suited for problems in which the set of variables, their interactions, and the mapping into outcomes is not theoretically obvious.

Almost 80 papers….

- Researchers use a wide set of variables from accounting, capital markets, governance, and auditing datasets to detect material misstatements or fraud.

- Others have categorized misstatements as intentional or unintentional

- Examples: Cecchini et al. (2010) , Bertemou, Cheynel, Floyd, and Pan, 2020 , (2009), Bao et al. 2020, etc.

Datasets used

- Most US based researchers use publicly available data of fraudulent companies using accounting and auditing enforcement releases (AAERs) and match it with financials from EDGAR database.

Use of Topic Modeling in fraud discussion

- Recent trend to use text within annual reports to detect fraud, and classifying them as intentional or unintentional.

- Topic modeling on textual data on annual reports, and use it as a signal in detecting financial misreporting.

- Others Naïve Bayesian classification based on SEC comment letters to forecast misstatements.

References : Brown, Crowley, and Elliott (2020) , Ryans (2021), Hayes and Efrim (2023)

.

Challenge to fraud research

• Developing models to detect financial statement fraud involves challenges. There is imbalance of data because a number of frauds are very few. In many of these studies, the percentage of fraudulent to nonfraudulent cases <3.5%

  – Some address the imbalance between the low number of fraud observations relative to the number of non-fraud observations by creating multiple subsets of the original dataset that each contains all fraud observations and different *random subsamples of non-fraud observations*.

  – Some have matched one fraudulent company with 3 other non-fraudulent companies to have a better ratio of 25%

• References: (Nissim, 2022, Papik and Papikav, 2022, Kotsiantis et al. 2006)

SACRAMENTO STATE
Redefine the Possible

Credit Risk Defaults and Bankruptcies

- AS 2415 of PCAOB is about consideration of an entity's ability to continue as a going concern.

- AS 2415 provides "guidance to the auditor ….to evaluating whether there is substantial doubt about the entity's ability to continue as a going concern"

Credit Risk (default), Bankruptcy

- Academicians find that machine learning models show improved bankruptcy prediction accuracy over traditional models.

- Some models like random forest model provide better results.

- Reference: Barboza, Kimura, and Altman (2017)

- Researchers use text analytics and Topic models for this purpose too.

- They created a comprehensive measure of credit risk by using LDA on combined qualitative information disclosed in conference calls and in MDA's section of the 10-K.

- They could predict bankruptcies better using this method. The explanatory power is at least six times greater than that of other credit risk measures.

- Reference: Donovan, J., Jennings, J., Koharki, K. and Lee (2021)

SACRAMENTO STATE
Redefine the Possible

Going at more granular Level

- And I like this…

- You look at different components that makeup the risks.

- Xu and He (2020) used ML on online supply chain financial credit risk assessment using financial data from China.

Machine learning can improve accounting estimates

- Most balance sheet and income statement items are based on estimates (example, pension and employee stock options expenses)

- These estimates are affected by objective estimation errors as well as by managerial manipulation.

  – This diminishes the reliability and relevance of financial reports.

SACRAMENTO STATE
Redefine the Possible

PCAOB estimation

- Public Board Accounting Oversight Board has Auditing Standards related to estimates: *Auditing Accounting Estimates, Including Fair Value Measurements*.

- PCAOB states that estimates often have a significant impact on a company's reported financial position and results of operations because accounting estimates can vary widely.

- As PCAOB states, accounting estimates are often some of the areas of greatest risk in an audit, requiring additional audit attention and appropriate application of professional skepticism.

- ML can potentially improve accounting estimates.

- Ding et al., 2020 using insurance companies' data on loss reserves estimates and realizations (future customer claims), show that the loss estimates generated by machine learning were superior than estimation using traditional methods

Behavioral Research suggests..

- There is enough evidence that auditors are increasingly using ML in their joint production of financial statements

- However, *auditors* may not incorporate the output of ML into their decisions

- Managers from firms that are not using ML in their companies, may not accept the judgments made by auditors using ML.

- Even within an audit firm, experienced auditors may reluctantly take advice from ML specialists auditors' advice because experienced auditors may perceive the expertise difference between them and the specialist as a threat to their self-regard

- We need some behavioral interventions and solutions to these issues.

- References: Knechel et al. 2020). ; Commerford et al. 2022; Peecher, M. E., Pietsch, C. R., Stirnkorb, S., and Yamoah, I. L., 2023; Estep, Griffith, & MacKenzi, 2023

Directions for further Research

- Models developed on US financial data sometimes achieve worse performance measures for ML techniques(Papik and Papikav, 2022). So, we have to be careful when using it in Europe and specific to Netherlands. More research that is country specific are needed.

- What specific accounting estimates can we predict better using ML?

- Netherlands in the frontier for ESG reporting. How can ML be used to analyze the rich ESG data?

Directions for further Research

- Can we predict specific types of Frauds using ML?

  - Example, can we predict Bill & Hold Frauds, or fraud caused using Special Purpose Vehicle ?

- There is a need for cooperation between Netherlands and USA academicians for doing relevant research with rigor.

- Netherland auditing research more close to practice, we are good in theoretical development

- Because of the Global Audit firms, we are becoming one world

Academia-Industry collaboration

- We need more practice oriented research

- Audit firms are keen to collaborate with academia in the USA

- Many projects in ML are now funded from audit firms, and case studies are being developed to be taught in classrooms using the expertise of both academia and industry

- Example: Deloitte invested millions of dollars for data analytics in accounting in Univ of IL.

- KPMG had at one time 13 accounting university Masters programs sponsored 100%

What are we doing for our Accounting students

- We cannot make them experts in ML

- But, we can make them appreciate about various ML algorithms.

- We have started using more low-code ML programs so that accounting students focus on understanding the algorithms instead of expending the cognitive load on learning syntax of programming.

- We expect a collaboration between ML experts and auditors in the industry to find solutions.