CS230

Predicting Corporate Misconduct using Deep Neural Networks

Youngchul Joo
Stanford GSB
waldojoo@stanford.edu

Seyeon Kim
Stanford GSB
seyeonk@stanford.edu

Abstract

The present paper evaluates and compares the performance of existing standards – Statement on Auditing Standards no. 99 and naive forensic models – in predicting firms that have committed financial misconducts. We incorporate deep neural network approach by conducting one-layered and four-layered neural network models to assess how well we can accurately predict two outcomes: firms being investigated by the U.S. Securities and Exchange Commission and firms being found guilty of misconduct given that these firms have been investigated by the SEC. Using only publicly available information, we find that naive forensic models have higher accuracy in predicting the outcome. However, this result should be taken with caution as we had some limitation in collecting data for SAS 99.

1 Introduction

Because corporate misconduct has a substantial negative impact on all shareholders (Choi and Gipper, 2019; Blackburne et al., 2020), the U.S. Securities and Exchange Commission (SEC) has enforced federal securities laws that protect investors from misconducts by holding the wrongdoers accountable and preventing future misconducts (SEC, https://www.sec.gov/about/what-we-do). However, given limited time and financial resources, the SEC cannot investigate all potential cases. Doing so moreover poses critical problems in that investigating innocent firms will be a waste of resources while missing the wrongdoing leads to accusations of negligence. Such constraints make targeting the 'right' firms likely to engage in misconduct especially important.

The SEC investigation typically begins with a "lead", which includes a wide range of sources ranging from Divisions of SEC surveillance activities to press reports and investor complaints (Blackburne et al., 2021). Promising leads result in an investigation, either directly or indirectly via evaluation of non-public information. Non-public information are obtained from channels such as interviewing witnesses, reviewing trading data, and examining brokerage records. It is important that we seriously consider and evaluate current processes by which formal investigations are pursued, especially given the financial costs following investigations. For example, the median abnormal return after the opening of an investigation is -6% for the firm, regardless of whether the investigation process is disclosed or not (Blackburne et al., 2021). Targeting wrong firms may lead to type II errors, which is equivalent to punishing innocent firms.

Our study incorporates deep neural network approach to evaluate how well do existing standards – Statement on Auditing Standards no. 99, hereafter SAS 99 – perform in identifying firms that have committed frauds. Specifically, we compare the performance of SAS 99 to the performance of forensic accounting model. The input to our algorithm are public information in SAS 99 and forensic model. We then use a logistic regression to output a predicted probabilities of being investigated and

CS230: Deep Learning, Winter 2018, Stanford University, CA. (LateX template borrowed from NIPS 2017.)

being found guilty conditional on being investigated. The key contribution of this paper is that it serves not only the SECs but also the investors (the main mission of the SEC is to protect investors) and the firm well. Especially from the perspective of the SEC, the proposed project can help the SEC decide which firms to target with limited resources.

2 Related work

Existing work has shown how machine learning could improve enforcement in various dimensions. Recent paper by Mohammed and Mahmud (2020) has shown, using accident related variables from Occupational Safety and Health Administration (OSHA), that machine learning algorithms have been successful in predicting the required outputs. Feroz et al. (2000) illustrated the applications of artificial neural networks to predict the targets of the SEC investigation without actual data on firms investigated. Along similar lines, we use machine learning to predict the likelihood of SEC investigation and the likelihood of conducting financial misconduct.

Additionally, many practitioners have been curious to study what features make firms to be a subject of an investigation and ultimately to be guilty. Most of the SEC investigations are conducted without disclosure to protect the identity of those under investigation (SEC, 2019). This poses possible threats to investors as there may be insider trading (Blackcburne et al., 2021) or other possible managerial decisions during the investigation periods (Beneish, 1999).

3 Dataset and Features

With the aim of predicting firms that are targeted by the SEC for investigation and verification of conviction, we use Compustat dataset to obtain accounting information for all publicly listed firms in the US from year 2000 to 2018. Because it is mandatory for all public firms to report information on their performance, the dataset we use contains rich information about the company, including their sales, cost of goods sold, and short-term and long-term debt. See below for an example of part of a row in the data:

Observation	**Firm Identification**	**SAS 99**	**Naïve Model**	**Investigated**	**Guilty**
1	GVKEY	SALTA	TLTA	Binary	Binary
2	Date	INVSAL	TLTE	Binary	Binary
3	CIK	INVCA	LTDTA	Binary	Binary
4	…	…	…	…	…
5					
6					
7					

In addition to the Compustat dataset, we bring in a novel SEC data that provides information on all investigations conducted by the SEC from 2000 to 2018 along with the Accounting and Auditing Enforcement Releases (AAER) data that provides information on the firms involved in financial misconducts. The "investigated" variable in the SEC data takes on the value of 1 if the firm was investigated by the SEC and 0 otherwise. The "guilty" variable from the AAER data denotes whether the firm received AAERs upon being found guilty of misconduct – the value of 1 if the firm is guilty and 0 if not.

For each firm-year observation, we construct two input variables: SAS 99 variables and naive forensic accounting variables. It is important to note that both input variables suffer from limitations in that we only use publicly available information to construct them. For example, risk factors provided by SAS 99 can be categorized into three categories of the 'fraud triangle' or three conditions that are present when fraud occurs: financial pressures, opportunities, and rationalizations (Skousen 2004). We rely solely on financial pressure information in constructing the SAS 99 variable, because other factors – opportunities and rationalizations – are not publicly available.

There are 228,443 firm-year observations in the Compustat raw data, 12,900 investigations in the SEC raw data, and 2,100 AAERs sent to companies found guilty of misconduct in the AAER raw data. Because obtaining information on non-public firms is not feasible given our data, we exclude investigations with no public firm code (PERMNO), leaving us with a total of 3948 investigations.

We further filter out all ex-post variables so that we forecast the outputs of our interest - whether the firm was investigated and whether the firm was found guilty - with information that the SEC would have had at the time of the investigation and verification of conviction. We also do random sampling to reduce the number of firms that are not investigated by keeping only 10% of them so that the number of the firms that are and are not investigated is comparable. We choose this method instead of data augmenting for several reasons. First of all, data on investigation and AAER is a full population, so it is hard to find additional source. Moreover, method of augementation is not plausible as firm heterogeneity makes it difficult to decide what variables we need to alter. We present distributions of Compustat data by year before (top histogram) and after cleaning (bottom histogram) in Figure 1.

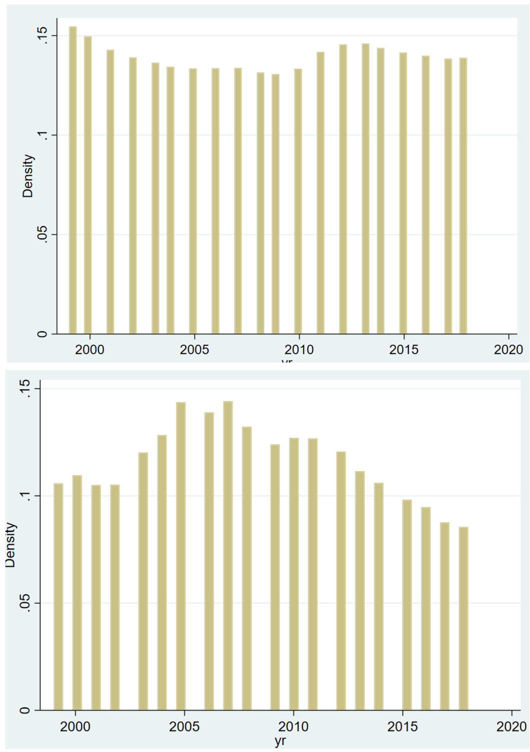

Figure 1: Distribution of Unprocessed Data (Top) & Cleaned Data (Bottom)

After merging the Compustat, SEC, and AAER data sets, we construct a total of 34 input variables where 20 variables come from SAS 99 and 14 come from naive forensic model. Our SAS 99 variable consists of financial ratios such as SALTA, net sales turnover divided by total assets, and INVSAL, total inventory divided by net sales turnover. Similarly, our naive forensic accounting variables are constructed using publicly available information in Compustat. The variables include TLTA, total liabilities divided by total assets, TLTE, total liabilities divided by stockholder equity parent, and LTDTA, total long-term debt divided by total assets, also shown in the table above.

After some basic cleaning, we end up with a total of four data sets. The first data set consist of SAS 99 variables to predict SEC investigation (hereafter referred to as “SAS Investigate”); the second consist of naive forensic variables to predict SEC investigation (hereafter “Forensic Investigate”); the third are SAS 99 variables to predict guilty or not (hereafter “SAS Guilty”); and finally, the fourth are naive forensic variables to predict guilty or not (hereafter “Forensic Guilty”). There were 28,000 training examples used for each SAS Investigate and Forensic Investigate data, and respectively, and 3,396 and 3,393 validation examples were used for the SAS Investigate and Forensic Investigate data respectively. For the SAS Guilty and Forensic Guilty data, 9,000 training examples were used for both data, and 499 and 502 validation examples were used in SAS Guilty and Forensic Guilty data respectively.

Finally, our neural network models as below. We use logistic regressions and first is our one-layer neural network model, where we have input layers, one hidden layer, and a sigmoid activation function that outputs one output layer. Another is our model with four hidden layers, where we chose relu, sigmoid, relu, and sigmoid activation functions to output a binary classification of whether a company is investigated or found guilty. We also tested two- and three- layer neural network models instead of a four-layer model, and we chose the four-layer model for its accuracy was the highest.

4 Methods

The main purpose of our model is to use financial information to decide whether a firm is likely to be investigated by the SEC and if so, how likely are they are to be found guilty. Recall that the fraud triangle stipulates three conditions associated with fraud – incentive/pressure, opportunity, and rationalization – and that for the purpose of this project, we limit our variables to only include incentive / pressure factors, which are the only public data to which we have access. It is therefore difficult to set a lower bound for human error and bayes error because our data is limited only to the publicly available dataset.

We used PyTorch as our primary implementation framework. First, we shuffle the raw merged data that has a total of 228,443 observations and split it into a train set of 200,000 examples and a test set of 28,443 examples. There are 14 unique inputs derived from descriptive information provided in SAS 99 and 20 unique inputs based on the naïve forensic model. We began our analysis by standardizing the raw merged data and running a one-layered (sigmoid) logistic regression on the normalized data sets consisting of SAS Investigate, Forensic Investigate, SAS Guilty, and Forensic Guilty data. To control for possible overfitting, we applied regularization to the neural network models. We set $\alpha = 0.01$ for the learning rate and binary cross entropy as a criterion between target and input probabilities. After trying several different types of optimization method such as momentum and ADAM, we ultimately decided to use stochastic batch gradient as our main optimizer.

We then trained our model using epoch of 100 and found an accuracy of 94% or over. This was an extremely skeptical result, because the accuracy was too high for such a simple model. To investigate the source of the result further, we ran a couple of experiments and realized that the issue was due to the number of binary classifications. That is, there were more than 200,000 firm-year observations that were not investigated by the SEC, while there were merely 9,396 firm-years that were investigated. Since the data we obtained is already a full population data, meaning that all firms that were investigated were included in our data, it was especially difficult to augment this type of data. We therefore conducted random draws to keep only 10% of the firms that were not investigated. With this adjustment to make the number of firms that were and were not investigated similar, the modified data set contains 22,000 firms that were not investigated and 9,396 that were investigated.

In running our four-layered model, where we use relu then sigmoid then relu then sigmoid functions, we run the same model for the one-layered model explained above using SGD and BCE. See the next section for results.

5 Experiments/Results/Discussion

The present study has 9 main models as illustrated in the next page. The first model is a logistic regression on an unenhanced data, which has an unusually high validation accuracy. As discussed earlier, we realized upon further investigation that a key error with this model was that there were too many values of 0 and too little values of 1 when we try to find firms that have been investigated by the SEC. We therefore address such issue by making an adjustment to our model by reducing the number of firms that were not investigated by 10% so that the number of investigated firms and the number of non-investigated firms are somewhat similar.

The second row in the results table below describes validation accuracy and training loss of the one-layer neural network predicting investigation using forensic model and shows that this model has 70% accuracy. A four-layer neural network predicting the same outcome, investigation, using the same forensic model illustrates that the accuracy improved by 3%. We see a similar pattern of four-layer neural network performing better than one-layer network in the models predicting investigation using SAS. The validation accuracy of one-layer neural network predicting investigation using SAS 99 is 50% while that of four-layer neural network predicting investigation using SAS is 69%.

The last four rows in the table illustrate similar patterns for models predicting guilty outcomes in that using four-layer neural network produces better accuracy than using one-layer neural network. We also see higher accuracy for these four models. It is important to note that these models predicting the guilty outcome were conducted on observations that have been investigated, rather than on all observations that include non-investigated firms. We made this decision, because only the firms that were investigated are 'eligible' for verification of conviction and therefore it is more theoretically important and relevant to evaluate models that predict firms that are found guilty among the investigated firms. However, we do understand limitations stemming from a smaller sample size for the models predicting the guilty outcome, as we limit our observations to 8,900 investigated firms as opposed to 210,000 firms that include the non-investigated firms.

It is also worth examining why SAS 99 performs so poorly compared to the naïve forensic model. While the fact that the naïve model performs better than SAS 99 model seems counterintuitive or even wrong at first glance, this is because SAS 99 requires access to private information that we do not have access to. For this reason, it would be a fruitful investigation for future research with better access to non-public data to pursue this line of research and compare models predicting firms that are to be investigated and found guilty of misconduct.

Model	**Validation Accuracy**	**Training Loss**
Logistic Regression on Unenhanced Data	0.9436	0.1901
1-Layer Neural Network Predicting Investigation Using Forensic Model	0.7033	0.6017
4-Layer Neural Network Predicting Investigation Using Forensic Model	0.7333	0.6174
1-Layer Neural Network Predicting Investigation Using SAS 99	0.5015	46.720
4-Layer Neural Network Predicting Investigation Using SAS 99	0.6966	0.6025
1-Layer Neural Network Predicting Guilty Using Forensic Model	0.9228	0.2859
4-Layer Neural Network Predicting Guilty Using Forensic Model	0.9472	0.3299
1-Layer Neural Network Predicting Guilty Using SAS 99	0.9218	5.0450
4-Layer Neural Network Predicting Guilty Using SAS 99	0.9300	0.1884

6 Conclusion/Future Work

All in all, we evaluated the SAS 99 and naive model in their predicting the firms that are likely to be investigated and found guilty. We tried different experimentation such as adjusting hyper parameters, deciding best performing optimization, and choosing optimal level of layers. We found that a simple naive model can predict 73 percent of the firms that are likely to be investigated and SAS does 69 percent. If we have access to other firm data relating to opportunities and rationalization as suggested in SAS 99, we believe that the SAS prediction will be higher. This accuracy is relatively high as we are looking at simple firm related variables and are accurately pointing out what firms are likely to be investigated. Moreover, our model has a 94 (93) percent accuracy to predict guilty for forensic (SAS 99) model, implying that if we know which firms are currently being investigated, we have a high probability of accurately identifying whether they are guilty or not.

7 Contributions

All two members of the team were heavily devoted to the project but on different areas of the project based on their skills and background. Youngchul Joo, as an Accounting PhD student with background in Economics and Mathematics, took the lead not only on acquiring and cleaning the data but also testing and running various advanced models used across multiple areas of research. Seyeon Kim, as an Organizational Behavior PhD student with background in Sociology, took the lead on the academic literature research, which involved understanding the firm context, interpreting analyses, and developing initial models that made the most sense given the data.

References

Alexander, J.A. & Mozer, M.C. (1995). Template-based algorithms for connectionist rule extraction. In G. Tesauro, D.S. Touretzky and T.K. Leen (eds.), *Advances in Neural Information Processing Systems 7*, pp. 609–616. Cambridge, MA: MIT Press.

Beneish, M.D. (1999). Incentives and Penalties Related to Earnings Overstatements That Violate GAAP. The Accounting Review, 74(4).

Bower, J.M. & Beeman, D. (1995). *The Book of GENESIS: Exploring Realistic Neural Models with the GEneral NEural SImulation System.* New York: TELOS/Springer–Verlag.

Blackburne, T. P., Kepler, J. D., Quinn, P. J., & Taylor, D. (2020). Undisclosed SEC Investigations. *Management Science* 67(6):3321-3984.

Choi, J, & Gipper, B. (2019). Fraudulent financial reporting and the consequences for employees. Working paper, available at https://ssrn.com/abstract=3346759.

Feroz, E., Kwon, T., Victor, P., & Park, K. (2000). The efficacy of red flags in predicting the SEC's targets: an artificial neural networks approach. *International Journal of Intelligent Systems in Accounting, Finance and Management*

Hasselmo, M.E., Schnell, E. & Barkai, E. (1995). Dynamics of learning and recall at excitatory recurrent synapses and cholinergic modulation in rat hippocampal region CA3. *Journal of Neuroscience* **15**(7):5249-5262.

Mohammed, J., Mahmud, M.J. (2020). Selection of a machine learning algorithm for OSHA fatalities. *2020 IEEE Technology & Engineering Management Conference (TEMSCON)*, 1-5.

Skousen, C. J. (2004). An Empirical Investigation of the Relevance and Predictive Ability of the SAS 99 Fraud Risk Factors. *Doctor of Philosophy*, Oklahoma State University, UMI Dissertation Services.