Genetic Algorithm-based Feature Selection for Auditing Decisions

Tao Yang
University of Wisconsin - Milwaukee, taoyang@uwm.edu

Derek Nazareth
University of Wisconsin - Milwaukee, derek@uwm.edu

Genetic Algorithm-based Feature Selection for Auditing Decisions

Completed Research

Tao Yang
University of Wisconsin-Milwaukee
taoyang@uwm.edu

Derek L. Nazareth
University of Wisconsin-Milwaukee
derek@uwm.edu

Abstract

When examining a firm's financial statements, independent auditors seek to render opinions on their fairness, accuracy, presence of fraud, and going concern, among others. This research focuses on the going concern, and the ability to predict when the going concern is flagged based on an array of accounting measures. It seeks to determine a parsimonious set of measures that can accurately predict when the going concern is raised, when using a linear kernel support vector machine for prediction. A genetic algorithm is employed to effectively reduce the set of measures without compromising accuracy of prediction. Using data from audits of public firms, a parsimonious model is created utilizing only 8 measures from a set of 35 available measures. The model exhibits 98.6% accuracy, and outperforms several other machine learning techniques.

Keywords

Feature selection, genetic algorithms, going concern, auditing.

Introduction

Auditing a firm's financial statements and records involves multiple tasks and outcomes. Aside from gauging whether US Generally Accepted Accounting Principles (GAAP) were followed correctly, auditors seek to uncover any material misrepresentation of the firm's financial status, whether they be through error or fraud. In addition to recommendations to the firm about financial reporting and internal controls, auditors also issue opinions about the fairness and accuracy of the firm's financial statement, and whether there exists any going concern in the near future. These opinions are important for a number of external stakeholders including creditors, investors, ratings firms, suppliers and customers.

This research focuses on the going concern as it has enormous implications for external stakeholders. The going concern status for publicly listed companies needs to be assessed by independent auditors under Statement of Auditing Standards No. 59. Auditors are responsible for evaluating whether there is any substantial doubt about the company's going concern ability for a reasonable period, based on their knowledge of all available information. In addition to clear implications for the firm itself, there are connotations for creditors, investors, suppliers, and related stakeholders. Furthermore, there are implications for auditors themselves, in that failure to identify a going concern issue may entail damages for stakeholders, and invite litigation, and will likely damage the reputation of the auditors.

Auditors have a vast array of measures and ratios at their disposal in assessing a firm's financial statements. Given the need to avoid misrepresentation, it is no surprise that the focus for most audits tends to be on fraud detection and prediction (Khan et al., 2021, Abbasi et al., 2012). However, the opinions on a firm's going concern are important, nonetheless. Forming this opinion requires careful consideration on the part of the auditors, using the available measures wisely. Clearly not all measures are relevant to arriving at an opinion on the firm's going concern. Determining the subset of measures that will accurately predict a going concern, while still not sacrificing predictive ability, has two-fold implications in this context. First, it reduces the cognitive burden on the auditors, in that there are fewer measures to consider. In addition, limited resources, time pressures, and cost constraints also work in favor of a reduced set of measures. Third, the use of a limited set of measures can be beneficial in terms of future prediction of other firm's going concern. Tagging a firm with a going concern opinion is not something that can be done lightly, in that an undeserved assessment can affect the firm negatively, while a missed assessment can affect its suppliers, customers, creditors, and investors.

Selecting a subset of measures can be a challenging task. A set of 30 measures will generate ($2^{30} - 1$) different combinations, or 1.07×10^9 potential solutions. Evaluating all these solutions requires considerable resources and time. A more effective approach is clearly needed. This paper employs a heuristic search approach using a meta-heuristic algorithm, viz genetic algorithms, to search for a parsimonious and effective set of measures. Genetic algorithms (GA) were initially proposed by Holland (1975) as a technique that mimics a natural biological process to select the most effective solution from a vast array of potential solutions. Genetic algorithms have been applied in a number of problems domains, and have proven effective in arriving at good solutions without having to explore the entire solution space. In the machine learning context, the selection of a set of measures from a larger set has been characterized as a feature selection problem. Feature selection represents an attempt to reduce the number of measures or variables so as to eliminate noise and confounding by irrelevant variables while still maintaining accurate prediction (Guyon & Elisseef, 2003). This research uses a genetic algorithm to address the feature selection problem in relation to identifying measures that are effective in predicting when an auditor will issue a going concern opinion on a firm's likelihood of survival. It should be borne in mind that this research does not seek to predict if a firm will fail or not; simply whether an auditor will issue an opinion to that effect. In addition, the incidence of issuing these opinions is relatively low, and that will have implications for the feature selection as well as its evaluation.

The remainder of the paper is organized as follows. Literature relating to GA, feature selection, and auditing decision prediction is reviewed in the next section. Formulation of the genetic algorithm to select features for a firm's going concern predict is then presented. Results of running the GA on a set of auditing data are presented. Implications for auditors as well as interested stakeholders are discussed. Future research in this domain rounds out the paper.

Review of related literature

This section briefly reviews relevant literature in the areas of feature selection, genetic algorithms, feature selection using genetic algorithms, and applications of feature selection in the auditing context.

Feature Selection

Feature selection is a topic that has gained a lot of attention in the machine learning discipline over the last few decades. Initial attempts at feature selection were motivated by the need to reduce the number of inputs used by machine learning algorithms, with a goal of improving performance, improving comprehensibility, and increasing prediction (Cai et al., 2018). Part of the motivation for this was that overlapping features would likely generate noise and lead to poor predictive accuracy. During the same period, several statistical techniques were developed to reduce dimensionality, though in truth this is different from feature selection, in that dimension reduction often combined features to yield new dimensions, while feature reduction seeks to eliminate features that introduce noise and impair predictive ability. The goals for feature selection were to identify powerful predictive features, remove irrelevant features, reduce the dimensionality of the feature space, use cleaner and better data quality, and generate simpler models with improved predictive ability (Ladla & Deepa, 2011).

Early approaches were applied on problems with a limited number of features, usually between 50 and 100. However, the application of machine learning to problems involving genetic data and text classification quickly changed the nature of feature selection to a problem where only a small fraction of features are to be retained from the original set (Guyon & Elisseeff, 2003). A number of newer approaches were developed, which can be summarized as filter methods, wrapper methods, and embedded methods (Chandrashekar & Sahin, 2014). Filter methods usually employ a ranking technique to eliminate poorly ranked features, and are relatively simple to use. Machine learning techniques are then applied to the filtered set of features chosen. Wrapper methods, on the other hand seek to evaluate subsets of features by running them through the machine learning algorithm, and then selecting the set with the best performance. A complete search using wrapper methods is prohibitively expensive, and some heuristic search approaches are generally employed. Embedded methods include the machine learning algorithms as part of the search process, and can be effective in reducing the time to arrive at a good solution. This research uses genetic algorithms as the heuristic search strategy in an embedded approach to feature selection.

Genetic Algorithms

Genetic algorithms are a form of evolutionary algorithm, which uses a population based stochastic search approach inspired from biology (Mirjalili, 2018). The approach was first proposed by Holland (1975) to locate high quality solutions using a natural selection and evolution process. It simulates natural inheritance and biological evolution processes to select the candidates which are the fittest to produce next generation offspring. The evolutionary processes involved in genetic algorithms include selection, crossover and mutation. This metaheuristic search method is capable of finding near-optimal solutions within a reasonable time. On account of their simplicity and power, genetic algorithms have been widely used to address problems in multiple domains, including economics, engineering, natural sciences, business management, among others.

The first phase in a genetic algorithm is selecting the initial population. Possible solutions from the search space are randomly selected and encoded as individual chromosomes, using a binary representation. The length of the chromosome is determined by the specification of each solution, usually measured as the number of genes or alleles. Each instance in the initial population is assessed using an evaluation function, which is akin to the objective function in an optimization problem. This assessment is termed the fitness value for the solution. New solutions are generated by combining existing solutions from the population. Using the principles of natural selection, chromosomes with higher fitness values are mated, using crossover operations to yield new, and potentially betters solutions. In an effort to introduce diversity in the solutions, and prevent being trapped in local optima regions, mutation principles are applied, wherein individual genes are randomly flipped in the binary representation. After the creation of all the individuals in the next generation, the process repeats until a predetermined stopping rule is reached. Within the structure of the genetic algorithm, there are several parameters that can be adjusted to control the operation and the effectiveness of the algorithm. The first consideration is the size of the population. Small populations suffer from lack of diversity, and results in limited exploration of the search space, and hence poor overall solutions. Large populations make for slow performance (Lobo & Lima, 2005). Recommendations for size include suggested numbers, between 50 and 100, and chromosome size consideration, between L and 2L, where L is the number of genes in a chromosome (Alander, 1992).

Generating new solutions involves selection of the fittest individuals. This can be accomplished by selecting from the top fraction of the ordered set, or using a roulette-wheel selection (Man et al. 1996), where the probability of selection is proportional to the fitness value – the larger the fitness value, the greater the probability of being selected. Two chromosomes from the selected set form the parents for the new chromosomes in the next generation through a crossover operation, combining the genetic information from the two coupled parents, and exchanging genetic information to form the offspring. Crossover points can be selected using a variety of options. Fixed crossovers are easily implemented, but lead to degradation over many iterations. Random crossovers within a range around the mid-point generate diversity over several generations, and are preferred (Gomez & Bielza, 2004). Over several iterations, the algorithm begins to zero in on good solutions within a specific region. However, this region may be representative of a local optimum. To counter this, a mutation operation is sometimes applied, so as to introduce additional diversity and force exploration of other regions (Hassanat et al. 2019). If the new region is unpromising, it is quickly abandoned; if promising, it will be explored further. The rate and degree of mutation can be varied to control the exploration of new areas. Mutation rates are usually set at a low value (between 0 and 0.05), so as to introduce controlled diversity without causing wild swings in the areas explored (Gomez & Bielza, 2004). The stopping criteria for genetic algorithms vary widely. In some cases, the fitness values are bounded, e.g. reliability may be upper bounded by 1.0. Solutions that fall with a predetermined neighborhood of this bound can be used as a stopping criterion for termination of the algorithm. Other stopping criteria are based on fitness improvement, e.g. no improvement in overall fitness for the last *x* iterations. Other implementations of genetic algorithms use a fixed number of iterations, based on the size of the population and the search space (Safe et al., 2004).

Embedded Feature Selection using Genetic Algorithms

Genetic algorithms have proven to be adaptive and effective in this context, and have been widely used in feature selection (Babatunde et al., 2014; Raymer et al., 2000; Yang & Honavar, 1998). They have been used to efficiently search for the optimal feature subset in high–dimensional feature spaces to achieve superior predictive accuracy (Amini & Hu, 2021). The preferred subset will exhibit high predictive performance with a parsimonious set of features. Clearly some tradeoff is implied – the number of features can be reduced but predictive accuracy will degrade. It is important to recognize that this approach is an expensive one, in that the evaluation function for each chromosome is an application of a machine learning technique to the set of features that the chromosome represents. Nonetheless, it has been used considerably in the feature selection process.

Murthy & Koolagudi (2018) use a GA-based feature selection approach for classification of vocal and non-vocal segments, employing ANN, SVM and RF classifiers for comparison. They found improvement in classification accuracy through the use of GAs. Labani et al. (2020) used a multi-objective genetic algorithm in classification performance of text feature selections, find it outperformed other methods of feature selection. Hoglund (2017) develops a GA-based decision support tool to select important information in predicting tax payment defaults with almost 75% predictive accuracy. Other uses of GA-based feature selection include applications to seed classification (Chtioui et al., 1998), corporate bankruptcy prediction (Gordini, 2014), and credit risk evaluation (Oreski, 2014).

Feature Selection in the Auditing Context

Auditing remains a critical aspect of business operations. As per US Generally Accepted Auditing Standards (GAAS), all US public listed companies must include an audit report issued by independent auditors with company's annual report. The auditor's decision about whether the company comply with US GAAP and the opinions issued about the financial statements have a critical impact on the decision making of other stakeholders such as suppliers, investors, creditors, and regulators.

Due to resource limitations, auditors cannot audit all business transactions, recorded numbers, and ledger accounts. While sampling remains an important aspect of auditing, there has been increased interest in identifying the critical measures that would play a role in the assembly of quality audit opinions. Though still at a preliminary stage, there is increasing research interest in applying data mining techniques in auditing settings. Financial fraud represents the area where the efforts are most noticeable. This includes the following research on feature selection in the auditing context.

Financial fraud has far-reaching consequences for the firm and its stakeholders, in the firm's supply chain as well as its investment spheres. Early detection of financial fraud is a major objective among auditors. Applying business intelligence to help auditors' decision making in whether the company is engaged in any fraudulent activities has been intensively investigated using data mining techniques. Researchers have employed different data mining techniques to identify fraudulent financial statements. Kirkos, et al. (2007) identify factors related to fraudulent statement and achieve approximately 70% accuracy using the full set of features available. Pediredla et al. (2011) use six data mining techniques to perform feature select and predict financial fraud and find PNN outperform all other techniques with and without feature selection. Hoglund (2017) developed a genetic algorithm-based feature selection to provide information about the factors in tax payment default detection. Alden et al. (2012) conduct fraud classification experiments in 458 balanced dataset using GA and MARLEDA models without feature selection and also achieve around 70% accuracy. Classification research in other auditing setting also includes internal control audit decision making and audit quality evaluations (Boskou, 2019; Boritz, 2013). Kirkos et al. (2008) used decision trees, neural networks and support vector machines to predict auditor selection with feature selection based on prior research and ANOVA. Nasir et al. (2021) developed a decision support system for internal control weakness detection. Genetic algorithm is used to select most important variables and perform best among others methods.

A different aspect about the auditor's opinion that is also important but has received comparatively less attention. According to US GAAS, auditors must also indicate whether there is any substantial doubt about the firm's going concern within the next year, based on the evidence obtained. Failure to detect potential going concern issue can result in losses to stakeholders, as well as have a negative impact on the auditor's reputation, and the confidence in the accuracy of their opinions. Raising this issue needs to be done carefully, since it can cause damage to the firm if improperly flagged, while it can have adverse effects for investors and creditor if overlooked. Like other audit decisions, it is subject to noise and obfuscation in the presence of a large number of accounting measures and indicators. The ability to distill the essential factors relevant to this opinion is crucial for effective specification of this concern. Feature selection is a natural strategy to achieve this, and this research examines the effectiveness of a GA-based feature selection approach to help predict when auditor opinions include information about a firm's going concern.

Feature Selection for Going Concern using Genetic Algorithms

The assessment of a firm's going concern remains a difficult task for auditors. Appropriate selection of features will simplify the task for auditors. Selection of the features needs to be driven by data on actual audits. We employ audit data extracted from COMPUTSTAT and WRDS. The dataset contains a total of 12,870 publicly listed US companies spanning the period 2010 to 2019. Complete data was available for 21,438 audits. Each observation included a total of 35 measures. The feature selection problem in this case encompassed a total of 3.44×10^{10} different combinations. Clearly, an evolutionary algorithm to effectively search the solution space would be helpful. The list of measures appears in Table 1. All measures are continuous in nature, except the three financial reporting measures, which are binary. The last three indicators are nominal and ordinal in nature.

Profitability measures	Financial reporting quality measures
After-tax Return on Average Common Equity	Auditor's financial reporting opinion
After-tax Return on Total Stockholders' Equity	Auditor's internal control opinion
Gross Profit/Total Assets	Restatement
Pre-tax return on Net Operating Assets	Valuation measures
Pre-tax Return on Total Earning Assets	Enterprise Value Multiple
Return on Assets	P/E
Leverage measures	Price/Cash flow
Current Liabilities/Total Liabilities	Price/Operating Earnings
Interest/Average Total Debt	Liquidity measures
Short-Term Debt/Total Debt	Cash Balance/Total Liabilities
Total Debt/Capital	Cash Flow/Total Debt
Total Debt/Equity	Cash Ratio
Total Debt/Total Assets	Current Ratio
Total Debt/Total Assets	Interest Coverage Ratio
Total Liabilities/Total Tangible Assets	Inventory/Current Assets
Operating efficiency measures	Operating CF/Current Liabilities
Accruals/Average Assets	Quick Ratio (Acid Test)
Payables Turnover	Other indicators
Receivables/Current Assets	Financial reporting year
	Global company key
	PERMNO

Table 1: Accounting Measures Available for Selection

The following structure was adopted for the genetic algorithm. Each combination of features was encoded as a chromosome. The evaluation function in this case is a little more involved than a simple expression. Since the objective was to better predict the presence of a going concern in the auditor's opinion, the evaluation function needs to be a performance measure of a classification task. Given that the data set included information on whether a going concern was raised or not, we opted to use a supervised learning technique that would work effectively with a small number of features. We selected Support Vector Machines to accomplish this. Support vector machines represent supervised learning when observations are classified in binary classes, and seek to assign observations to either category though the use of an n-dimensional hyperplane (Cortes and Vapnik, 1995). Since support vector machines produce a nonprobabilistic classification, they are well suited to the task at hand. Support vector machines can produce good classification results with a small number of dimensions. We opted to use a linear kernel for the support vector machine, given their effective performance. For the performance measure, we opted to use classification accuracy. Using a standard confusion matrix, where TP denotes true positive, TN denotes true negative, FP denotes false positive and FN denotes false negative, the accuracy is given by:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

The evaluation of each chromosome in the population requires the application of a support vector machine computation to that set of features and subsequent computation of accuracy, making the running of the GA rather compute-intensive. Accordingly, we elected to cap the number of features at 20, yielding a maximum of 1.05×10^6 combinations – a more manageable number, but still beyond what an auditor can assess.

The rest of the GA parameters were set as follows. The GA population was set at 100. The population comprised individuals that were randomly selected from the solution space, with the caveat that the features were capped at 20. The crossover point was set at 0.5. In order to avoid being trapped in a local area, the mutation probability was set at 0.04. The evaluation function was classification accuracy using a linear kernel support vector machine. The GA was run for 50 iterations.

Results

The GA was implemented using the sklearn-genetic Python package, which is dedicated to the selection of features using a GA meta-heuristic. Performance of the GA in terms of classification accuracy as well as number of features was tracked across the iterations. Performance data for the GA appears in Figure 1.

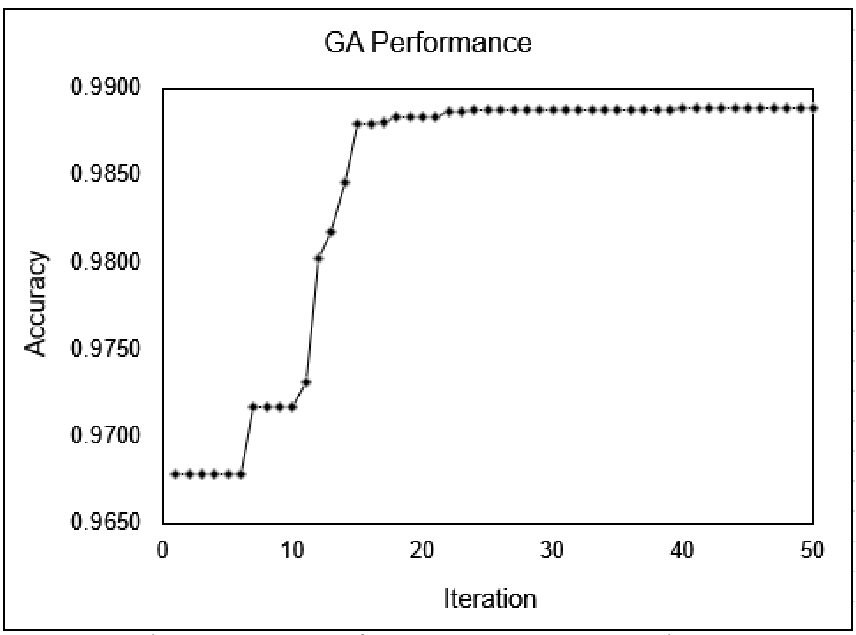

Figure 1: GA Performance across Iterations

The graph indicates that the GA is performing as expected, with the evaluation function improving over the iterations. It should be noted that the initial level of accuracy is relatively high, but it does improve as the GA progresses. The improvement appears to tail off after about 25 iterations, indicating that the selection of 50 iterations for the stopping rule is acceptable. Since one of the objectives was to reduce the number of features needed for classification, we also compared feature count and accuracy across the best results for each iteration, as depicted in Figure 2.

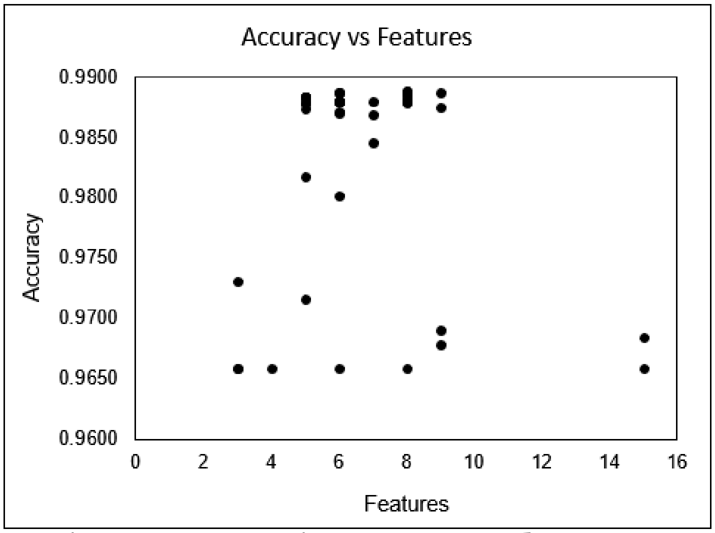

Figure 2: Contrasting Accuracy and Features

The top left corner represents the ideal solution, i.e. fewest features needed for classification, with the best accuracy possible. Clearly that is an unattainable ideal, and so we explore regions near it. In the ultimate analysis, the tradeoff between features and accuracy will determine which solution is the preferred one. The top cluster spans solutions that vary between 5 and 9, with a maximum accuracy observed with a solution involving 8 measures. This set is depicted in Table 2.

Return on Assets
Auditor's Internal Control Opinion
Pre-tax Return on Total Earning Assets
Inventory/Current Assets
Current Liabilities/Total Liabilities
Operating CF/Current Liabilities
Auditor's Financial Reporting Opinion
Current Ratio

Table 2: Measures Selected for Maximum Accuracy

The presence of internal controls and financial reporting are expected. Prior research has demonstrated internal controls to be an important factor in business operations, designed to safeguard the operation of the company, address business risks sufficiently, and promote operational efficiency. Companies with weak internal controls are more likely to receive going concern opinions from auditors (Jiang, 2010), and to have increased business risks and uncertainty of future corporate operations (Goh et al., 2013).

The accuracy of the support vector machine with all 35 features, i.e. no feature selection was 0.9660. For the selected 8 features, the best accuracy was 0.9888. Ordinarily, one would expect that an increase in features will lead to improved classification accuracy. However, inclusion of additional features can lead to overfitting, which will degrade classification accuracy on a holdout sample. The high values of accuracy prompted a closer look at the data. It transpired that of the set of 21,438 observations, only 440 cases involved an opinion of going concern. While this is representative of most audits, it does lead to a situation of imbalanced data. As a consequence, classification accuracy will be high, and not serve as a very discriminating performance measure. Instead, an F1-measure, which is the harmonic mean between precision and recall, will be more appropriate. Upon closer inspection, the F1-measure, which assigns equal weight to precision and recall may not be the preferred selection. False negatives, where a going concern opinion is present and the model predicts that it is not applicable, are a far greater concern. As such, greater emphasis should be placed on recall, as opposed to precision. Therefore, the F2-measure is preferred in this case. Table 3 illustrates the accuracy and F2-measures, with and without feature selection.

	No Feature Selection	Best Features Selected
Accuracy	0.9660	0.9888
F2-measure	0.1096	0.7486

Table 3: Performance Measures with and without Feature Selection

It can be seen that appropriate feature selection results in improved prediction performance. While the accuracy improves marginally, echoing the imbalanced data set, the F2-measure improves dramatically. This indicates that appropriate feature selection is able to eliminate noise from irrelevant variables, and the confounding effects of other measures. In an effort to assess the effectiveness of support vector machines for feature selection, 10 additional classification models were applied to the data, using the full set as well as the reduced set of features identified through the genetic algorithm. These results appear in Table 4.

Classification Model	No Feature Selection		Best Features Selected	
	Accuracy	F2-Measure	Accuracy	F2-Measure
Ada Boost Classifier	0.9845	0.7343	0.9858	0.7657
Decision Tree Classifier	0.9812	0.7385	0.9807	0.7159
Extra Trees Classifier	0.9879	0.7592	0.9876	0.7928
Gradient Boosting Classifier	0.9878	0.7805	0.9876	0.7974
K Neighbors Classifier	0.9646	0.1702	0.9861	0.7642
Light Gradient Boosting Machine	0.9888	0.7962	0.9875	0.8008
Linear Discriminant Analysis	0.9807	0.6917	0.9833	0.7392
Logistic Regression	0.9640	0.0400	0.9841	0.7067
Naive Bayes	0.9414	0.3392	0.8226	0.5474
Random Forest Classifier	0.9867	0.7102	0.9877	0.7839

Table 4: Performance of Alternative Classification Models

The results are illuminating in several respects, though the comparison may not be entirely fair. It is readily apparent that some classification models perform rather poorly when working with the entire set of features. This is understandable given that they were created to address different classification tasks, using potentially different assumptions. However, when limited to a small set of features, they all perform rather well. In fact, the variation across the models is far less pronounced. This indicates that appropriate feature selection will improve the classification performance. The improvement in accuracy is marginal, due to the presence of an imbalanced data set. However, the F2-measure, that places greater emphasis on false negatives, clearly demonstrates the improvement due to appropriate feature selection for predicting any going concern issues.

Research and Managerial Implications

Feature selection has often been employed when working with data sets that include hundreds or thousands of features, as in words in a document or corpus. In those circumstances, it is a very necessary part of the analysis, often to gauge sentiment, or underlying bias or intent. Feature selection will result in a dramatic reduction of relevant features. However, when working with a much smaller set of measures, it is unclear that feature reduction will produce similar results. Our research demonstrates that even when each measure has material intrinsic value for decision makers, judicious feature reduction can generate a parsimonious set of measures that possess good predictive ability. This research focused on the firm's going concern for prediction using the audited financial statement data. There are several other concerns that can bubble up in an auditor’s opinion, including fraud, error, GAAP compliance, among others. Similar research can be conducted to determine which set of measures are most germane for predicting those other concerns.

Given the large number of measures available to auditors, the cognitive processing load can be significant if all measures are to be considered when issuing opinions. Different measures are likely to be of varying importance to opinions concerning various aspects about the firm’s performance and viability. This research demonstrates that not all measures are necessary when assessing a firm’s going concern. In fact, elimination of more than three-fourths of the measures results in better predictive ability. This has substantial implications for a variety of stakeholders. For auditors, it indicates that they can narrow their analysis to a small set of select measures when attempting to gauge the firm's going concern. In addition, focusing on this select set has the potential to result in more accurate indications for the firm's going concern. For investors, lenders, and analysts, access to more accurate assessments of going concern is invaluable. In addition, the benefit of this approach is that one does not have to wait until an audit is conducted. Instead, a review of the firm's relevant measures can serve as an early warning signal that something is amiss. The analysis can also be performed for all firms in an industry or sector, for those individuals seeking to invest or manage portfolios within an industry or sector.

Conclusions, Limitations, and Future Research

A firm's stakeholders rely heavily on the professional opinions of auditors about the business to make decisions. Firms with going concern issues can present substantial risks to the stakeholders and auditors alike, particularly if they result in adverse outcomes. If going concern problems are not effectively identified by auditors, they face potential litigation risks, and damage to their reputation as well. Relying on a large set of measures that can be contradictory in nature is likely to make the assessment of the going concern harder. In addition, the decision to issue a going concern audit opinion is not an easy one, in that it can have negative consequences for the firm or its stakeholders, particularly if done incorrectly. While there are many options for selecting features from a large set, including enumeration, sequential selection, statistical-based selection, this research adopts a heuristic search strategy. Through the use of a genetic algorithm-based approach, a number of feature combinations are explored quickly and efficiently. The resulting set of features is parsimonious and has powerful predictive ability. There are several limitations associated with this research. First, it focuses on whether auditors issue a going concern opinion, and not whether the firm fails. Issuing a concern is no guarantee of failure, and the lack of a concern does not mean continued success. Auditor independence, conservatism, and ability affect the issuance of a going concern opinion (DeFond & Zhang, 2014). Predicting failure would require augmentation of the data set. Next, as with any machine learning approach, the results are sensitive to the data set employed. A different data set may result in different features being selected. Inclusion of periods of more widespread failure, e.g. financial crisis of 2008 and the Covid-19 pandemic, could very well alter the results. Lastly, if the feature set is made public, there exists the potential for firms to attempt to game the assessment. Nonetheless, given the results, this approach can be applied in a proactive manner to flag firms for potential going concerns as an early warning signal. The approach can also prove effective for selecting features that can be used to derive salient features to predict other concerns in an auditor's opinion, including fraud, error, GAAP compliance, among others.

References

Abbasi, A., Albrecht, C.C., Vance, A., and Hansen J. 2012. "Metafraud: A meta-learning framework for detecting financial fraud", *MIS Quarterly* (36:4), pp. 1293-1327.

Alander, J.T. 1992. "On optimal population size of genetic algorithms". *CompEuro 1992, Proceedings Computer Systems and Software Engineering*, pp. 65-70.

Alden, M.E., Bryan, D.B., Lessley, B.J., and Tripathy, A. 2012. "Detection of Financial Statement Fraud Using Evolutionary Algorithms", *Journal of emerging technologies in accounting* (9:1). pp. 71-94.

Amini F. and Hu G. 2021. "A two-layer feature selection method using Genetic Algorithm and Elastic Net", *Expert Systems with Applications* (166, 15 March 2021, 114072.

Babatunde, O. H., Armstrong, L., Leng, J., and Diepeveen, D. 2014. "A Genetic Algorithm-Based Feature Selection", *International Journal of Electronics Communication and Computer Engineering* (5:4). pp. 899-905.

Boritz, J.E., Hayes, L. and Lim, J.H. 2013. "A content analysis of auditors' reports on IT internal control weaknesses: the comparative advantages of an automated approach to control weakness identification", *International Journal of Accounting Information Systems* (14: 2) pp. 138-163.

Boskou, G., Kirkos, E., and Spathis, C. 2019. "Classifying internal audit quality using textual analysis: the case of auditor selection", *Managerial Auditing Journal* (34: 8). pp. 924-950.

Cai, J., Luo, J., Wang, S., and Yang, S. 2018. "Feature selection in machine learning: A new perspective". *Neuro computing* (300: 26). pp. 70-79.

Chandrashekar, G. and Sahin, F. 2014. “A survey on feature selection methods”, *Computers and Electrical Engineering* (40), pp. 16–28.

Chtioui, Y., Bertrand, D., and Barba, D. 1998. “Feature selection by a genetic algorithm. Application to seed discrimination by artificial vision”, *Journal of the Science of Food and Agriculture* (76: 1). pp. 77-86.

Cortes, C. and Vapnik, V. 1995. “Support-Vector Networks”, *Machine Learning* (20: 3), pp. 273–297.

DeFond, M. and Zhang, J. 2014. "A review of archival auditing research", *Journal of Accounting and Economics* (58: 2-3), pp. 275-326.

Goh, B.W., Krishnan, J.K., and Li, D. 2013. “Auditor Reporting Under Section 404: The Association between the Internal Control and Going Concern Audit Opinions”, *Contemporary Accounting Research* (30: 3) pp. 970-995.

Gomez, M. and Bielza, C. 2004. “Node deletion sequences in influence diagrams using genetic algorithms. *Statistics and Computing* 14. pp. 181–198.

Gordini, N. 2014. A genetic algorithm approach for SMEs bankruptcy prediction: Empirical evidence from Italy”, *Expert System with Applications* (41: 14). pp. 6433-6445.

Guyon, I. and Elisseeff, A. 2003. “An Introduction to Variable and Feature Selection”, *Journal of Machine Learning Research* (3), pp. 1157–1182.

Hassanat, A., Almohammadi, K., Alkafaween, E., Abunawas, E., Hammouri, A. and Prasath, V.B.S. 2019. “Choosing Mutation and Crossover Ratios for Genetic Algorithms—A Review with a New Dynamic Approach”, *Information* (10: 12), 390; doi:10.3390/info10120390

Holland, J. H. 1975. *Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence*. U Michigan Press.

Hoglund, H. 2017. “Tax payment default prediction using genetic algorithm-based variable”, *Expert Systems with Applications* (88), pp. 368 – 375.

Jiang, W., Rupley, K.H., Wu, J. 2010. “Internal control deficiencies and the issuance of going concern opinions”, *Research in Accounting Regulation* (22 :1). pp. 40-46.

Khan, A.T., Cao, X., Li, S., Katsikis, V.N., Brajevic, I. and Stanimirovic, P.S. 2021. “Fraud detection in publicly traded U.S firms using Beetle Antennae Search: A machine learning approach”, *Expert Systems with Applications* (191), 116148

Kirkos, E., Spathis, C., and Manolopoulos, Y. 2007. “Data Mining techniques for the detection of fraudulent financial statements”, *Expert Systems with Applications* (32: 4). pp. 995–1003.

Kirkos, E., Spathis, C., and Manolopoulos, Y. 2008. “Support vector machines, Decision Trees and Neural Networks for auditor selection”, *Journal of Computational Methods in Sciences and Engineering* (8: 3). pp 213-224.

Labani, M., Moradi, P., and Jalili, M. 2020. “A multi-objective genetic algorithm for text feature selection using the relative discriminative criterion”, *Expert Systems with Applications* (149), 113276.

Lobo, F. G., and Lima, C.F. 2005. “A review of adaptive population sizing schemes in genetic algorithms”. *Proceedings of the 7th annual workshop on Genetic and evolutionary computation* June. pp. 228–234.

Man, K.G., Tang, K.S., and Kwong, S.1996. “Genetic algorithms: concepts and applications [in engineering design]”, *IEEE Transactions on Industrial Electronics* (43: 5). pp. 519-534.

Mirjalili, S., 2018. *Evolutionary Algorithms and Neural Networks*. Springer Nature.

Murthy, Y.V.S., and Koolagudi, S.G. 2018. “Classification of vocal and non-vocal segments in audio clips using genetic algorithm based feature selection (GAFS)”, *Expert Systems with Applications* (106: 15), pp. 77–91.

Nasir, M., Simsek, S., and Ragothaman, S. 2021. “Developing a decision support system to detect material weakness in internal control”, *Decision Support Systems* (151: 2). 113631.

Oreski, S., Oreski, G. 2014. “Genetic algorithm–based heuristic for feature selection in credit risk assessment”, *Expert Systems with Applications* (41: 4). pp. 2052-2064.

Pediredla, R.S., Ravi, V., Rao, G.R., and Bose, I. 2011. “Detection of financial statement fraud and feature selection using data mining techniques” *Decision Support Systems* (50: 2). pp.491-500.

Raymer, M.L., Punch, W.F., Goodman, E.D., and Kuhn, L. 2000. “Dimensionality reduction using genetic algorithms” *IEEE Transactions on Evolutionary Computation* (4: 2), pp. 164–171.

Safe, M., Carballido, J., Ponzon, I., and Brignole, N. 2004. “On stopping criteria for genetic algorithms”, *Advances in Artificial Intelligence (SBIA)*. pp. 405-413.

Yang, J., and Honavar, V. 1998. “Feature subset selection using a genetic algorithm”, *IEEE Intelligent Systems* (13: 2), pp. 44–49.