DOI: https://doi.org/10.37075/EA.2024.3.04 Economic Alternatives, 2024, Issue 3, pp. 529-540

Improving the Quality of Financial Information Through Machine Learning

Received:
Available online:

Georgi Hristov[*]

Abstract

This paper reviews previous research in order to emphasize the importance of financial information and its effects both on companies and stakeholders (owners, managers, investors, and creditors). It outlines the problems with one of the key financial reporting assumptions – the going concern assumption which is equalized to bankruptcy for the purposes of the analysis. The empirical analysis includes the creation of several machine learning models which classify companies as either "going concern" or "non-going concern" based on four financial indicators. The aim of the analysis is to provide insight on how machine learning approaches can improve financial information quality.

Keywords: financial information quality; financial reporting; going concern; bankruptcy; machine learning

JEL: C45, M41, M42, G33, D80, D91

1. Introduction

The purpose of this paper is to add to the existing research which deals with improving financial information quality. We present a machine learning approach to improving one of the main financial reporting assumptions – going concern, where lack of research is found.

The importance of financial information of high quality has been discussed in numerous research articles. The rationale behind this active pursuit to understand what constitutes high information quality and how to measure it can be explained through the fact that managerial decision-making is virtually impossible without information.

Figure 1 presents a simple visualization of the decision-making process. Identification of the business problem, analysis of the alternatives and evaluation of effectiveness all require financial information which must be relevant and faithful. Financial information is used as a benchmark of whether and how the entity achieves its goals. Naturally, first we should establish what the business strategy, critical success factors (CSFs), and key performance indicators (KPIs) of the entity are. All three are affected by financial information provided to managers.

2. Relevant literature

2.1. Importance of useful financial information

A business strategy is developed by the management body so that the entity can achieve its goals. Bhimani & Langfield-Smith (2007) show that financial information is used

[*] Department of Information Technologies and Communications, University of National and World Economy

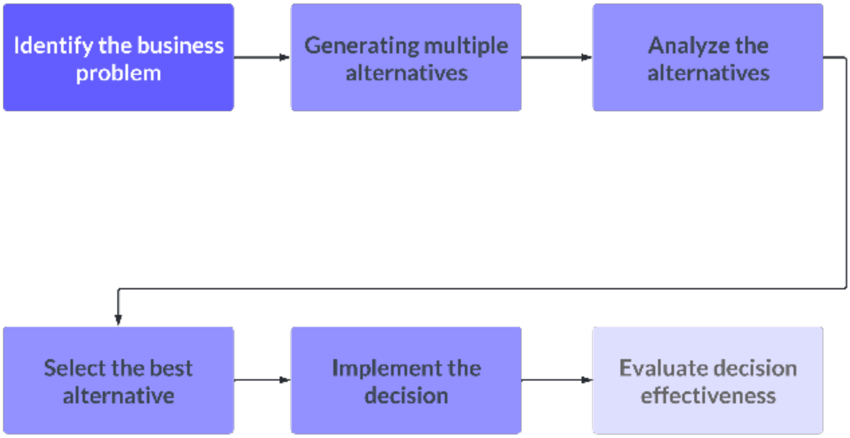


Figure 1. The decision-making process

both in strategy development and strategy implementation (Bhimani & Langfield-Smith, 2007). However, financial reporting irregularities can mislead managers and lead to undesired consequences for all stakeholders. This is why the Conceptual Framework for Financial Reporting (CFFR) issued by the International Accounting Standards Board (IASB) poses qualitative requirements for financial information to be considered useful. A major problem in this area is how to assess informational quality quantitatively. Beest, Braam, & Boelens (2009) attempt to construct a measurement tool to assess the quality of financial reporting in terms of relevance, faithful representation, understandability, comparability, verifiability, and timeliness (Beest, Braam, & Boelens, 2009). Measuring information quality is an ongoing debate (Tang, Chen, & Lin, 2016; Mbobo & Ekpo, 2016; Kythreotis, 2014), which aims at establishing what high quality actually is and how it could be improved with respect to individual users of the information.

Additionally, Xing & Yan (2019) found that accounting information quality is significantly and negatively correlated with systemic risk (Xing & Yan, 2019). In other words, poor accounting information quality causes higher costs of capital for the companies. Additionally, Lambert, Leuz, & Verrecchia (2007) demonstrate that quality of accounting information can affect the the cost of capital of the entity both directly and indirectly. As is well-known, costs of capital is one of the main benchmarks to be used in managerial decision making.

Problems with information arise not only when information quality is considered, but from the point of view of the users of the information as well. Information asymmetry is one such problem – entities possess far more information than market participants which leads to a substantial advantage on their side. Cohen (2003) suggests that firms with high quality financial reporting policies have reduced information asymmetries (Cohen, 2003). Byard, Li, & Weintrop (2006) reaserch the importance of quality of information provided to financial analysts. They found that better quality corporate governance is associated with the quality of financial disclosures provided by the entity (Byard, Li, & Weintrop, 2006). The importance of information asymmetry has been enhanced by the fact that it has been studied not only in the context of “entity-market participants” relationship. Wolff, Schell, & Moog (2022) study the effects of information asymmetry in intrafamily business succession where, once again, the importance of financial information is emphasized – “The financial transfer is crucial for the future successor’s position in the firm, as ownership is associated with power (finance).” (Wolff, Schell, & Moog, 2022, p. 3).

Another instance of the effects of declining fininancial reporting is presented by Wang & Wu (2011) – many of the firms they studied managed their ernings via below-the-line items to overstate their value (Wang & Wu, 2011). They pose another problem when information quality is considered – managers and owners could use information asymmetries and ambiguities to their unfair advantage.

2.2. Going concern assumption

Financial reporting is based on principles, concepts and assumptions which aim at improving financial information quality and ensuring that information provided by companies is not misleading. Going concern assumption is one of the principles set out in CFFR the purpose of which is to establish a correct expectation in users of financial statements with respect to the future existence of the company. “Financial statements are normally prepared on the assumption that the reporting entity is a going concern and will continue in operation for the foreseeable future.” (International Accounting Standards Board, 2018, p. 22). Three main criteria can be derived from the CFFR which an entity should meet in order to be considered a going concern:

- the entity will continue to operate in the foreseeable future. In accounting practice “foreseeable future” usually means one reporting period (financial year);
- the entity does not intend to initiate a liquidation procedure voluntarily;
- the entity will not be forced to initiate a liquidation procedure.

If a company does not meet these criteria, it should disclose the fact that it is not a going concern, the reasons why it’s not and the new basis on which its financial statements are prepared. The basis is changed in order to reflect the fact the entity will most probably cease to exist in the next reporting period. Such disclosure is, of course, vital for investors, creditors and employees of the company. Such important managerial judgment should be supported by a vigorous statistical analysis of key industry financial metrics, operating results, liquidity, solvency, forecasted net cashflows, capital expenditure commitments and many more financial and non-financial factors.

2.3. Predicting bankruptcy with data mining

Due to the specificity of the topic, it is hard to find literature which deals with the problems of the going concern assumption through the application of machine learning technologies. For the purposes of the literature review, we assume that bankruptcy analysis can be equalized to going-concern analysis.

Anandarajan & Anandarajan (1999) compare machine learning approaches (neural network and expert systems) with the classical statistical approach (multiple discriminant analysis). They found that neural networks achieve higher classification

Table 1. Formulae for input variables

FINANCIAL RATIOS	FORMULAE
TLTA (Total Liabilities to Total Assets)	$\sum liabilities_{as\ of\ 31.12.2020} \Big/ \sum assets_{as\ of\ 31.12.2020}$
RETA (Retained Earnings to Total Assets)	$retained\ earnings_{as\ of\ 31.12.2020} \Big/ \sum assets_{as\ of 31.12.2020}$
QACL (Quick Assets to Current Liabilities)	$\frac{\sum cash, cash\ equiv., receivables_{as\ of\ 31.12.2020}}{\sum current\ liabilities_{as\ of\ 31.12.2020}}$
NITA (Net Income to Total Assets)	$\frac{profit/loss_{for\ FY\ 2020}}{(\sum assets_{as\ of\ 31.12.2020} + \sum assets_{as\ of\ 01.01.2020})/2}$

accuracy when predicting whether a company is a going concern or not (Anandarajan & Anandarajan, 1999).

Koh (2004) also compares neural networks, decision trees and logistic regression. The author confirms that machine learning/data mining techniques are better at going-concern prediction (Koh, 2004).

The research of Rafiei, Manzari, & Bostanian (2011) shows that neural networks achieve substantially higher classification accuracy than multiple discriminant analysis when predicting bankruptcy (Rafiei, Manzari, & Bostanian, 2011).

Chi & Shen (2022) outline the importance of the going concern assumption. They construct eight models out of which CHAID–C5.0 achieves the highest prediction accuracy.

The importance of going concern/ bankruptcy prediction is well-supported by the research community but not enough due to specific domain knowledge requirements. Further in the paper, we present the results of different machine learning models which aim at going concern prediction.

2.4. Predicting bankruptcy with text mining

The current dynamic business environment poses many challenges to managers – one of which is working with non-numerical (qualitative) data. Nowadays, professionals have to extract insights from unstructured data (e.g. text, video, audio, social media data). When predicting bankruptcy, analysts can use the disclosers made from companies in their financial reports which are in the form of free-form text. Shirata & Sakagami (2008) analyze different components of disclosures made by Japanese companies. Their goal is to find keywords which can be used as indications for bankruptcy (Shirata & Sakagami, 2008).

Shirata, Takeuchi, Ogino & Watanabe (2011) upgrade the research of Shirata & Sakagami (2008) by extracting indicative key phrases (Shirata C. , Takeuchi, Ogino, & Watanabe, 2011).

3. Methodology

3.1. Sample

The sample consists of 100 entities operating in the construction sector in Bulgaria. 50 of them were declared bankrupt in 2021 (non-going concern or NGC) and 50 of them are financially stable (going concern or GC). Therefore, the sample is balanced from the point of view of the target variable CLASS (GC or NGC). For each entity 4 financial ratios are calculated: RETA (Retained Earnings/Total Assets) as a representative of efficiency; NITA (Net Income/Total Assets) also as a representative of efficiency; QACL (Quick Assets/Current Liabilities) as a representative of liquidity; and TLTA (Total Liabilities/Total Assets) as representative of indebtedness. Table 1 presents the formulae used to calculate the input variables.

In addition to those four input variables, the sample consists of three more features – CLASS which is going to be the target variable (it can either be GC or NGC); ID_COM_REGISTER – a unique identification number of each company in the commercial register – it cannot be used as an input variable due to its high entropy; and COMPANY – which is the name of the respective company – also ignored in the analysis due to the same reason.

3.2. Data preprocessing

Features are fitted to normal distribution through log transformation. None and INF values are replaced with 0 which should be taken into account for the final conclusions and analysis. Additionally, features are normalized, and the target variable is binarized. Figure 2 presents the distribution of input variables before and after the log transformation.

The Spearmen correlation coefficient is calculated in order to establish which of the input variables are of highest importance for the model. Figure 2 presents the correlation matrix – NITA, QACL and RETA and is highly correlated with the target variable CLASS.

Since the number of input variables is low (four variables) all variables are included into the modelling phase of the analysis.

3.3. Modelling

Due to the nature of the analysis, we aimed to create a model which provides the highest classification accuracy. The choice of algorithms was driven by the small sample size – for example, neural networks were not implemented due to the high probability of overfitting. However, using algorithms such as the decision tree allows us to observe how the model made the decision as well.

4. Results

The first algorithm which was implemented was k-Nearest-Neighbors (kNN). Grid search method was used to determine what number of neighbors yields the highest classification accuracy. As with every algorithm in the research, a close look at both training and testing accuracy was maintained to avoid overfitting (Figure 3). Because of the small sample size, i.e. low resource requirements, 10-fold cross-validation was implemented. kNN was tested among 1 to 24 number of neighbors – the highest accuracy of 0.88 was yielded when 6 neighbors were used.

Figure 4 represents the confusion matrix of the kNN model. False positive misclassifications are predominant which naturally leads to higher recall and lower precision score.

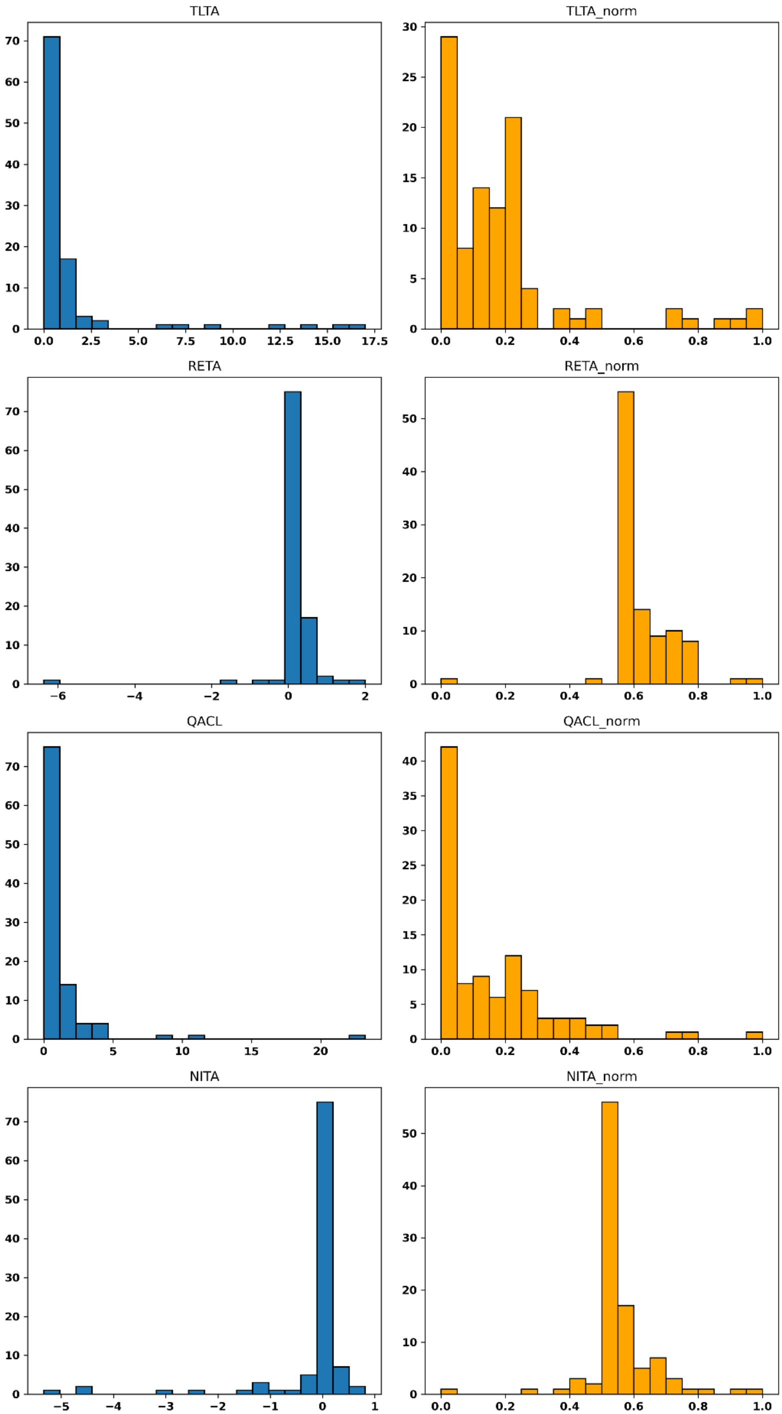


Figure 2. Features before and after preprocessing

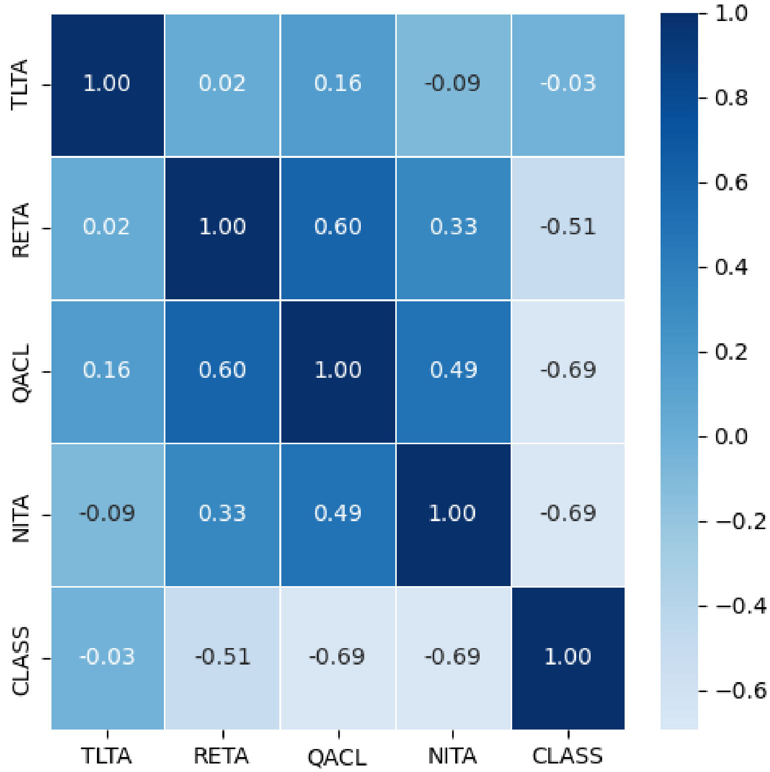


Figure 3. Correlation matrix

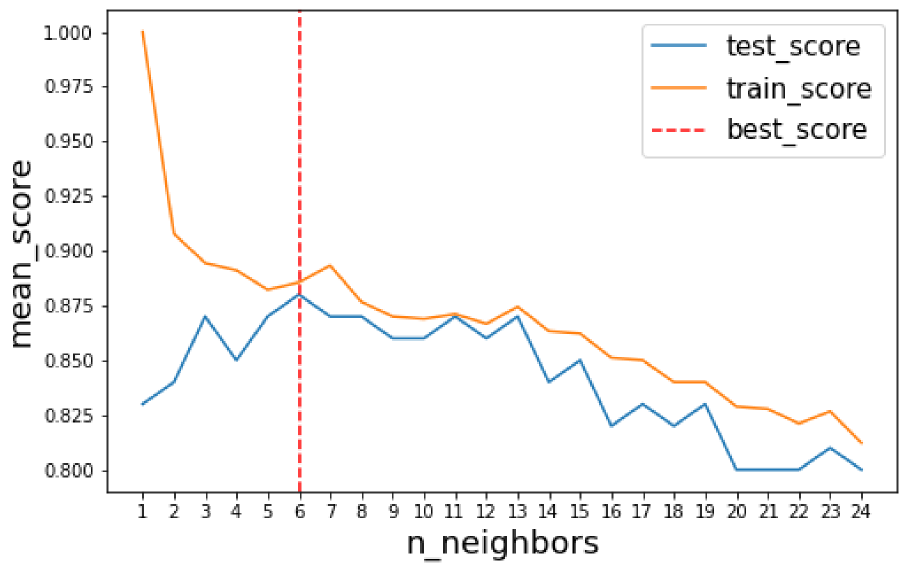


Figure 4. Train-test accuracy comparison

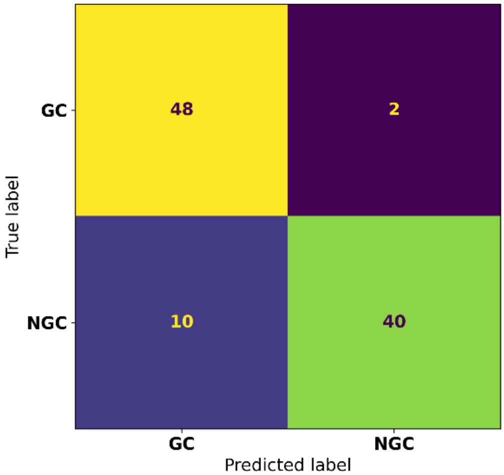


Figure 5. kNN confusion matrix

The next model was built upon the ridge regression algorithm. The model was trained using different levels of regularization (alphas). The highest accuracy (0.85) is achieved at alpha = 0.001. Unlike the kNN model, ridge regression yields more false negatives than false positives. Therefore, precision is higher than recall.

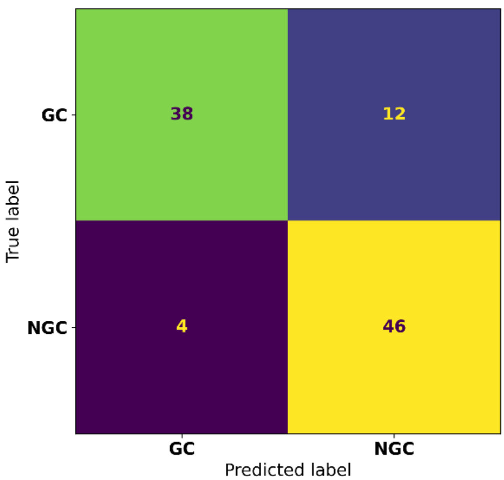


Figure 6. Ridge regression confusion matrix

The last algorithm applied was decision tree (Appendix A). The model is trained by implementing different levels of the different hyperparameters (minimum samples in leaf, max depth, and max features to consider when looking for the best split). The best decision tree yields 0.89 accuracy and the confusion matrix shown on Figure 6.

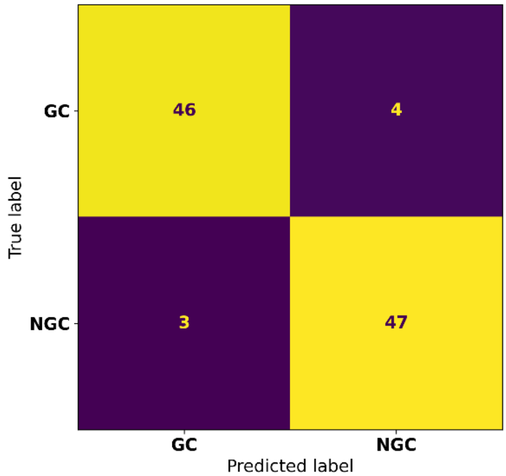


Figure 7. Decision tree confusion matrix

4.1. Model comparison

Table 2 presents all the metrics used to evaluate the performance of the models. Being the most complex model, the decision tree achieves the highest classification accuracy, followed by kNN and Ridge regression.

Table 2. Score comparison

MODEL	ACCURACY	PRECISION	RECALL
kNN	0.88	0.80	0.95
Ridge Regression	0.85	0.92	0.79
Decision Tree	0.89	0.94	0.92

However, in the process of selecting the most appropriate model in a real business environment, managers should take into account the availability of resources at companies, since high performing models usually are time-, space-, and maintenance-consuming.

4.2. Guidelines on improving the empirical analysis

There are several ways of improving the analysis which can act as guidelines for future work.

In the first place, the number of input variables can be increased to accommodate more complex models. This would give an opportunity to discover useful and unexpected insights into what drives companies to bankruptcy.

Furthermore, sample size should be increased. This would improve the representativeness of the sample. Additionally, with large samples, neural networks can be employed, which would increase the accuracy of the classification. Higher resource requirements should be taken into account when considering larger sample size and deep learning approaches.

In the third place, the analyzed period can be extended due to specificities in which the analyzed companies have filed for bankruptcy. Time series analysis could be included as well.

In addition to the aforementioned improvements, a text mining approach can also be applied in order to determine keywords and key phrases which distinguish going concern from non-going concern companies. We should take into account that not all companies in Bulgaria are required to prepare disclosure notes which are the main source of text data for such analysis.

5. Conclusion

Machine learning models are constantly implemented with the idea of automating manual tasks and giving opportunities to managers, accountants, and auditors to engage in more creative and not so labor-intensive work. This paper goes to show that machine learning can be applied with the purpose of improving subjective managerial judgments which in turn improves the quality of the decision made not only by internal users of financial information (such as managers and owners) but also for external users – creditors and potential investors. The results from the empirical analysis prove that a successful model could be built if domain specific knowledge and machine learning expertise are combined efficiently enough.

Authors' Note

The code of the implementation in Python (Jupyter Notebook) is provided in the Supplementary Material to this paper.

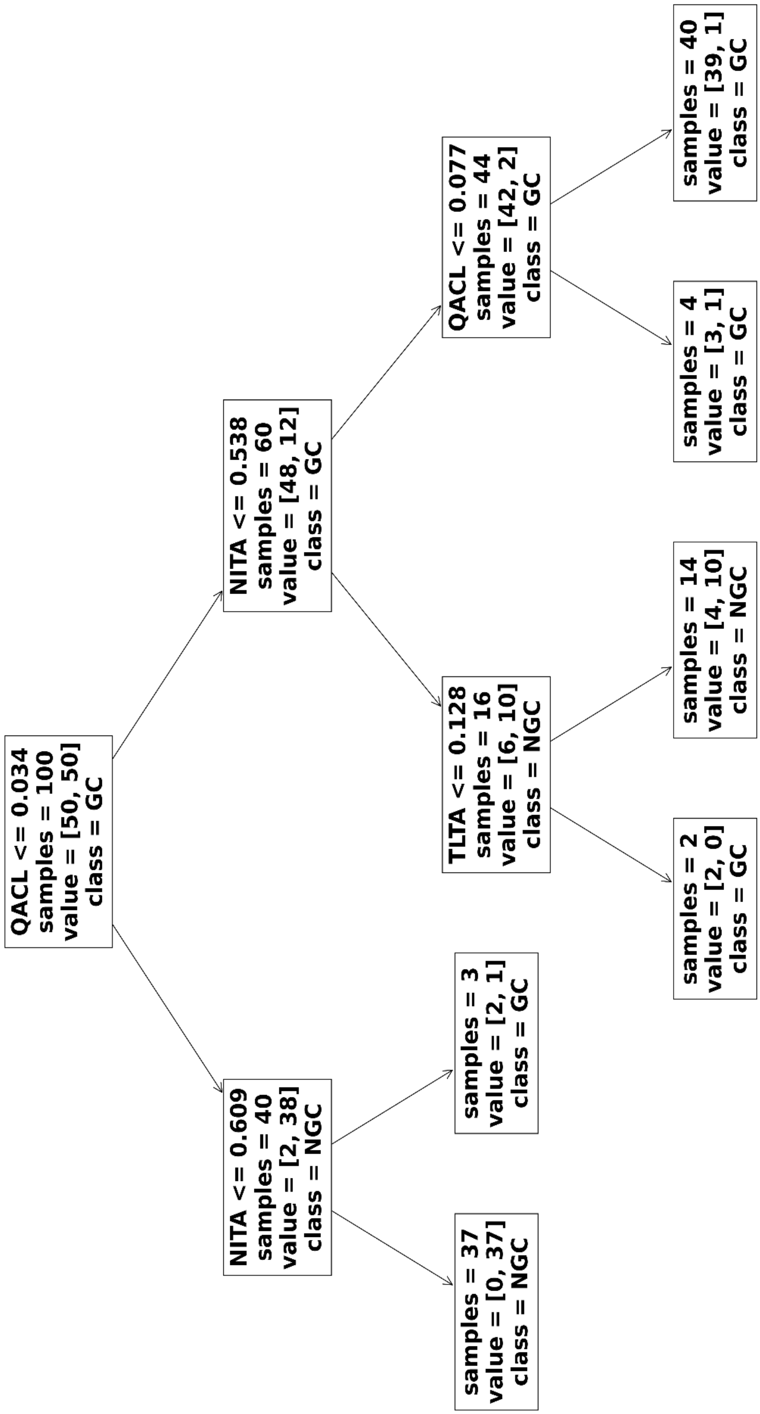
Funding

The author received no financial support for the research, authorship, and/or publication of this article.

Supplemental material

Supplemental material for this article is available online.

Appendix A. Decision tree



References

Anandarajan, M., & Anandarajan, A. (1999). A comparison of machine learning techniques with a qualitative response model for auditor's going concern reporting. *Expert Systems with Applications, 16,* 385-392.

Beest, F. v., Braam, G., & Boelens, S. (2009). Quality of Financial Reporting: measuring qualitative characteristics. *Nijmegen : Economie*.

Bhimani, A., & Langfield-Smith, K. (2007). Structure, formality and the importance of financial and non-financial information in strategy development and implementation. *Management Accounting Research*, 3-31.

Boghean, F., Boghean, C., & Morosan-Danila, L. (2011). Approaches and Advances in the Use of the Financial-Accounting Information in the Decision-Making and Innovative Process. *Revista Economica*, 38-44.

Byard, D., Li, Y., & Weintrop, J. (2006). Corporate governance and the quality of financial analysts' information. *Journal of Accounting and Public Policy*, 609–625. doi:https://doi.org/10.1016/j.jaccpubpol.2006.07.003

Chi, D.-J., & Shen, Z.-D. (2022). Using Hybrid Artificial Intelligence and Machine Learning Technologies for Sustainability in Going-Concern Prediction. *Sustainability*, 1-18. doi:https://doi.org/10.3390/

Cohen, D. A. (2003, December). Quality of Financial Reporting Choice: Determinants and Economic Consequences. *SSRN Electronic Journal*. doi:http://dx.doi.org/10.2139/ssrn.422581

Grigoras-Ichim, C.-E., & Morosan-Danila, L. (2020). Analysis of the Correlation Managerial Decisions - Interim Financial Statements. *1st International Conference Global Ethics - Key of Sustainability (GEKoS)* (pp. 203-211). Iasi, Romania: LUMEN Publishing House. Retrieved 12 22, 2022, from https://www.ceeol.com/search/chapter-detail?id=884665

Herath, S. K., & Albarqi, N. (2017). Financial Reporting Quality: A Literature Review. *International Journal of Business Management and Commerce, 2*(2), 1-14.

International Accounting Standards Board. (2018). Conceptual Framework for Financial Reporting. Retrieved from https://www.ifrs.org/content/dam/ifrs/publications/pdf-standards/english/2022/issued/part-a/conceptual-framework-for-financial-reporting.pdf

Koh, H. C. (2004). Going concern prediction using data mining techniques. *Managerial Auditing Journal, 19*(3), 462-476.

Kythreotis, A. (2014). Measurement of Financial Reporting Quality Based on IFRS Conceptual Framework's Fundamental Qualitative Characteristics. *European Journal of Accounting, Finance and Business, 2*, 4-29. doi:https://EconPapers.repec.org/RePEc:scm:ejafbu:v:2:y:2014:i:3:p:4-29

Lambert, R. A., Leuz, C., & Verrecchia, R. E. (2007, May). Accounting Information, Disclosure, and the Cost of Capital. *Journal of Accounting Research*, 385-420. Retrieved 12 22, 2022, from https://www.jstor.org/stable/4622036

Mbobo, M. E., & Ekpo, N. B. (2016). Operationalising the Qualitative Characteristics of Financial Reporting. *International Journal of Finance and Accounting*, 184-192. doi:10.5923/j.ijfa.20160504.03

McDaniel, L., Martin, R. D., & Maines, L. A. (2002). Evaluating Financial Reporting Quality: The Effects of Financial Expertise vs.

Financial Literacy. *THE ACCOUNTING REVIEW*, 139-167. doi:10.2308/accr.2002.77.s-1.139

Rafiei, F. M., Manzari, S., & Bostanian, S. (2011). Financial health prediction models using artificial neural networks, genetic algorithm and multivariate discriminant analysis: Iranian evidence. *Expert Systems with Applications, 38*, 10210–10217.

Rico, M., & Puig, F. (2021, April). Successful turnarounds in bankrupt firms? Assessing retrenchment in the most severe form of crisis. *BRQ Business Research Quarterly*, 114-128. doi:https://doi.org/10.1177/2340944421994117

Shirata, C., & Sakagami, M. (2008). An Analysis of the "Going Concern Assumption": Text Mining from Japanese Financial Report. *Journal of Emerging Technologies in Accounting*, 1-16.

Shirata, C., Takeuchi, H., & Watanabe, H. (2011). Extracting Key Phrases as Predictors of Corporate Bankruptcy: Empirical Analysis of Annual Reports by Text Mining. *Journal of Emerging Technologies in Accounting, 8*, 31-44.

Shirata, C., Takeuchi, H., Ogino, S., & Watanabe, H. (2011). Extracting Key Phrases as Predictors of Corporate Bankruptcy: Empirical Analysis of Annual Reports by Text Mining. *JournalL of Emerging Technologies in Accounting*, 31-44.

Tang, Q., Chen, H., & Lin, Z. (2016). How to Measure Country level Financial Reporting Quality? *Journal of Financial Reporting and Accounting, 14*(2), 230-265. doi:https://doi.org/10.1108/JFRA-09-2014-0073

Wang, X., & Wu, M. (2011). The quality of financial reporting in China: An examination from an accounting restatement perspective. *China Journal of Accounting Research*, 167-196. doi:https://doi.org/10.1016/j.cjar.2011.09.001

Wolff, S., Schell, S., & Moog, P. (2022, April). Information asymmetries in intrafamily business succession. *BRQ Business Research Quarterly, 25*(2), 173-192. doi:https://doi.org/10.1177/2340944420942525

Xing, X., & Yan, S. (2019). Accounting information quality and systematic risk. *Review of Quantitative Finance and Accounting, 52*, 85-103. doi:https://doi.org/10.1007/s11156-018-0703-z