Universidad Pontificia Comillas

Doctorate of Business Administration in Management and Technology

Using Big Data techniques to enhance Auditors' procedures and to analyse Auditor's report

Autor: Cleber Beretta Custodio
Director: Dr. Jose Portela Gonzalez

MADRID | Mayo y 2022

"Oh God thy sea is so great and my boat is so small"

Old Breton prayer given to President Kennedy by Admiral Hyman Rickover

Table of Contents

Acknowledgement:

First of all I must thank the Creator, source of all inspiration and the light in my destiny as He has given me the strength, health and has helped to bring me together with all those that I had the pleasure to learn from and work with.

Secondly, I want to thank my family for their support during this journey and for their contribution of the most valuable asset in the world – Family Time.

I would also like to thank my co-writers in the different papers being presented, Dr. Sheng-Feng Hsieh from National Taiwan University, Dr. Miklos Vasarhelyi from Rutgers University, Dr. Jose Portela Gonzalez from Universidad Pontificia de Comillas and Yu Gu, Doctorate Student from Rutgers University.

Finally, I would like to thank to people that made a significant contribution in inspiring and helping me during this journey, Dr. Maria Teresa Corzo, Dr. Antonio Rua, Dr. Carlos Bellón Núñez-Mera, Dr. Javier Morales Mediano, Dr. Aurora Garcia Domonte all from Universidad Pontificia de Comillas, Dr. Chanyuan (Abigail) Zhang from Rutgers University, Dr. Érica Custódio Rolim from IESE Business School, Kelly Deannine Langdon, Ignacio Blanco Gonzalez, Juan Jose Gil Rivas, Silvia Lopez Magallon, German de la Fuente and Manuel Arranz from Deloitte for their help, insights, edits and comments to this paper.

Introduction

Motivation

The audit of financial statements is a regulated process. The work that auditors must perform are clearly defined in auditing standards (ie. audits conducted in accordance with international auditing standard, audits conducted in accordance with the standards of the Public Company Accounting Oversight Board ("PCAOB") from the United States, etc.). The main objective of an audit of financial statements is to enable the auditor to express an opinion whether the financial statements being audited are prepared, in all material respects, in accordance with an applicable financial reporting framework. For auditors to reach such conclusion, they must plan and perform the audit to obtain reasonable assurance about whether the financial statements are free from material misstatement (information in such financial statements that is considered sufficiently incorrect impacting the economic decisions). To obtain such reasonable assurance, auditors must evaluate if sufficient and appropriate audit evidence was obtained during the work. Also, the auditor must evaluate if any uncorrected misstatements that could be material, individually or in aggregate exists and, also, if the requirements of the applicable financial reporting framework were fulfilled considering a qualitative assessment of the aspects of the company's accounting practices and potential risk indicators of bias in management's judgments. Evidently, considering these procedures, one can say that auditing is a complex process, that auditors are subsumed in a large quantity of data (evidence) and are, also, constantly applying judgment to assess management estimates. This process of gathering data/evidence and assessing estimates has not changed since the very inception of audits back in 1845 were law in England required audits to protect shareholders from "improper actions by promoters and directors." although it really expanded globally with the enactment of the U.S. Securities Acts of 1933 and 1934 after the 1930 stock market crash.

In an attempt to explain the importance of the audit procedures and the specific responsibilities of new entry auditor, in the introduction of the June 1991 edition of the Arthur Andersen Audit Approach Book there was a letter written by James G. Hooton dated on December 7, 1990 where he indicated that a part of the responsibility of the audit process fell (and still falls) particularly in the testing phase of an audit, which requires auditors to perform tests, review controls, request confirmations, perform recalculations, conduct inspections, etc. It was mentioned that all such work should be documented in the audit workpapers so that a more experienced auditor could access the evidence gathered and conclude about the adequacy of the results obtained. Such audit evidence/documentation also provides large amounts of data about the company being audited and, also, about auditors' assessment.

Technology and data availability has changed significantly since 1990 although the requirements about gathering data/evidence and its evaluation and have remained substantially the same. Nowadays auditors must consider such data availability as well as the technological developments (hardware and software) to perform an analysis of entire populations, not only samples, to reach a conclusion about the lack of error and achieve a high quality and robust audit work and working papers. Innovation, Big Data, and the use of technology should be part of auditors' day to day activity.

The incorporation of technology and innovative techniques within the audit procedures has also captured a lot of attention from regulators worldwide. For instance, in March 2020, the UK Financial Reporting Council (FRC) issued a comprehensive 30-page Audit Quality Review (AQR) Thematic Review entitled "The use of technology in the audit of financial statements" to update their 2017 review, which had become obsolete considering the rapid implementation of technology and innovation by the different Audit Firms.

In the updated review, the FRC summarized their work and concluded:

> *"We conclude that technology has much to offer the auditor in terms of efficiency and effectiveness. It may be used increasingly to support the assessment of the reasonableness of estimates made by management. The potential for the use of technological resources to enhance audit quality is clear. However, it is no replacement for the skills and informed judgement of an experienced auditor."*

Considering all the data available by companies being audited that now are accessible and new software applications available to perform a data analysis, visualization, querying, etc. there is a clear area of interest about efficiency and effectiveness to be evaluated ("knowledge gap"). Also, there is very little research used in real practice to assist auditors in evaluation the data obtained and how to use such data specifically for auditors' conclusion in an area that significant judgment could have to be exercised.

For instance, accounting standards require that financial information must be prepared under the assumption that there is no intention for liquidation due to bankruptcy/failure or cessation of business activities. In relation to such accounting requirement auditors must obtain sufficient appropriate audit evidence to conclude the appropriateness of management's use of the going concern basis of accounting in the preparation of their financial statements. In this respect, if auditors conclude that a material uncertainty still exists, they are required to include a specific paragraph in their audit report (ASU 2014-15; ISA 570).

Many academic research papers provide some indicators, ratios, or metrics that could be used to assist the users of financial information in their direct assessment of potential going concern uncertainties risk without regard of the entity own assessment. One of the metrics most commonly used until today is the Z-Score Model (Altman, 1968). Such Z-Score have been tested for many years and it have been proven that it worked reasonably well in the prediction of going concern uncertainties (Risk of Failure) although, obviously, with some imperfections that arises in some specific geography/situations (Altman, Iwanicz-Drozdowska, Laitinen, & Suvas, 2017). In addition to the pure academic literature, some authors in narrative books have published with an interesting analysis of past bankruptcy situations and thus providing some additional indicators that could be also helpful for the user of financial information in its independent evaluation of a potential risk of failure of an entity (Steer, 2018).

Additionally, in the past years, given the availability and volume of financial data, some authors have also tried to stablish a data-based-going-concern prediction model that could be helpful for the auditors' assessment based on quantitative techniques such as "random forest" approach (black-box approach) that outperforms the baseline methods in terms of the accuracy rate (Hsu & Lee, 2020) and others have introduced more advanced data mining techniques, such as support vector machines and rule-based classifiers (Martens, Bruynseels, Baesens, Willekens, & Vanthienen, 2008). Such black box models could suppose an additional difficulty for auditors in real practice as their mechanism would not be transparent in order to provide evidence in the auditors working papers that could be reperformed by a more experienced auditor. Finally, in relation to quantitative parameters, although researchers like those mentioned above concluded they could be a useful tool for auditors' assessment, others have considered that neither auditors' own qualitative assessment nor the Z-score bankruptcy prediction model are good predictors, as auditors also consider facts unrelated to a potential bankruptcy in addition to their natural conservatism and knowledge (Hopwood, McKeown, & Mutchler, 1994), therefore suggesting that qualitative and, therefore, auditors' knowledge are also important factors.

Considering all the above, it is evident that an investigation about a model based on simple machine learning white-box (as opposed to black box) tool combining both qualitative and quantitative factors could become a helpful tool. It is considered that such tool could eventually be very useful if it provides additional information to understand if the use of machine learning in an area or complex judgment could be useful in practice and help auditors in their assessment of entity risk of failure and, consequently, in their analysis for the decision of a potential inclusion or not of a going concern warning in their report. The result of this investigation could provide some helpful tools to increase audit quality and efficiency and, therefore, it will be treated in detail in Chapter I – Decision Tree Tool for Auditors' Going Concern Assessment in Spain.

Additionally, technological developments in tools available to review textual data provide another area of interest for an investigation about the auditors' report basically because auditors' reports nowadays provide a significant amount of data of some key audit subjects. This change in auditors' report began in 2018 in Spain when auditors were required to include a section in their reports about Key Audit Matters ("KAM") - defined in the audit standards as "those matters that, in the auditor's professional judgment, were of most significance in the audit of the financial statements of the current period.". Given the importance of such disclosure in auditors' reports, KAMs became critical to serve as a beneficial mechanism for enhancing financial reporting quality (Gold et al., 2020). Also, there was an expectation with the introduction of such requirements that such information shouldn't be "boilerplates" rather it was expected to provide relevant information to the users of the reporting (IAASB 2015b; Pelzer 2016; PwC 2015; Segal 2017; Velte & Issa 2019). However, there are no academic papers investigating and documenting the boilerplate issue of KAM disclosures. Evidently this also constitutes a knowledge gap that could be filled with the use of textual data analytics to evaluate and provide empirical evidence in the textual similarity of KAM disclosures. Also, considering the details embedded in each KAM, such evaluation could be more granular and consider three different perspectives: (1) KAM topic, (2) auditor, and (3) company segment/industry. This will be further treated in Chapiter II – Using Big Data Techniques to Analyse Key Audit Matters in the Auditors' Report.

Objectives

Considering data used by auditors and data available in their auditor's report has grown exponentially in the last few years, it is clear that there is a need for research to provide evidence that fills the existing knowledge gap in relation to the use of data to assist auditors in documenting their conclusion in a significant area of judgment such as the inclusion or not of a going concern paragraph and also using textual data analytics technique to evaluate the level of similarity in the auditors' report.

Therefore, the objective of this thesis is to use big data techniques in trying to enhance and assist auditors' procedures and documentation and, also, in analysing the results of the auditors' report.

More specifically, the first objective is to perform an experiment based on entities' financial historical data and auditors' going concern conclusion and evaluate if an efficient decision tree predicted model can be created and then tested on a subsequent period to access the reliability of such model. The research it incorporates both qualitative indicators based upon the individual auditors' and its Firm knowledge and experience, and quantitative indicators based upon entity's historical financial ratios and indicators, therefore a two-fold (qualitative and quantitative) model strategy differentiating it from the extant literature.

If the results of such investigation were to be positive, it could enable auditors to enhance the quality of their working papers by providing additional evidence about auditors' decision for the inclusion of a going concern paragraph in their independent auditor report. In addition, it would be highly valuable to test such predicted model with subsequent year real data evaluating the precision on the decision of the inclusion of a going concern paragraph.

The second objective of this essay it to investigate textual similarity of key audit matter (KAM) disclosures by analysing KAM items in auditor's reports of Spanish companies in fiscal years 2017 and 2018 with the objective to understand how similar KAMs they were from one year to another. To achieve this objective, the cosine similarity technique could be a technique to be applied that would provide empirical evidence about how similar KAM items were in terms of word usage (a Big Data Strategy). Also, the investigation could have some granularity to analyse KAM items for two consecutive years based on the following three combinations: (1) KAM topic, (2) KAM topic and auditor, and (3) KAM topic, auditor, and industry of the client being audited. The results could provide empirical evidence if auditors from the same accounting firm tend to have a recurring textual similarity under each KAM topic, and such similarity increases for clients within the same industry. Therefore, the results could answer the fundamental question of: "*What is the textual similarity of KAM disclosures from Spanish public companies classified by KAM topic, auditor, and industry?*"

Outline

The remainder of this thesis is organized as follows:

Chapter I is devoted to an experiment related to the first objective indicated previously, which relates to performing an experiment based on entities' financial historical data and auditors' going concern conclusion and evaluate if an efficient decision tree predicted model can be created then and then tested on a subsequent period to access the reliability of such model. Therefore, the main purpose of this chapter is to apply Big Data Techniques (Decision Tree Predictive analytics) in trying to create a model to assist auditors in their procedures to evaluate the inclusion of formal warning in their auditors' report about the existence of a significant uncertainty about the Company's ability to apply the going concern principles, that is, a warning that would indicate the company could be potentially face a bankrupt situation or liquidation. Also, this Chapter provides an evaluation of the application of the constructed model in real practice by comparing the result of the model with the final auditors' report.

The contents of this chapter is an integral copy of the paper named "Decision Tree Tool for Auditors' Going Concern Assessment in Spain" submitted in May 2022 to The International Journal of Digital Accounting Research (IJDAR). The authors of such paper were Cleber Beretta Custodio from Universidad Pontificia de Comillas, Spain, Yu Gu, Doctorate Student from Rutgers Business School, USA, and Dr. José Portela González from Universidad Pontificia de Comillas, Spain. Also, the paper acknowledged the comments and edits received from Dr. Miklos Vasarhelyi from Rutgers University, Dr. Chanyuan (Abigail) Zhang from Rutgers University, Dr. Érica Custódio Rolim from IESE Business School, and Kelly Deannine Langdon from Deloitte.

Chapter II is devoted to the second objective indicated above, which is to investigate textual similarity of key audit matter (KAM) disclosures by analysing KAM items in auditor's reports of Spanish companies in fiscal years 2017 and 2018 with the objective to understand how similar KAMs they were from one year to another. Therefore, in such chapter Big Data Techniques (Cosine similarity in Text Data) were applied to evaluate the textual similarity of key audit matters paragraphs in the auditor's report Spanish Companies to enable a conclusion about the level of similarity among different auditors, different industries and different accounting subjects

The content in this chapter is an integral copy of the paper with the title "The Textual Similarity of KAM Disclosures for Spanish Companies" submitted in June 2021 and accepted in September 2021 by The International Journal of Digital Accounting Research (IJDAR) and published in its Volume 21, 2021 Edition with the DOI: 10.4192/1577-8517-v21_7. The authors of this paper were Dr. Sheng-Feng Hsieh from National Taiwan University, Taiwan, , Cleber Beretta Custodio from Universidad Pontificia de Comillas, Spain, and Dr. Miklos A. Vasarhelyi from Rutgers Business School, USA.

Finally, the last part of this dissertation elaborates on the conclusions reached by both investigations, the archived contributions to the expansion of the knowledge in these subjects and the applicability in the practice, the limitations and, finally, areas where future investigations could help to expand the achieved knowledge.

References

Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, 23(4), 589-609. https://doi.org/10.2307/2978933

Altman, E. I., Iwanicz-Drozdowska, M., Laitinen, E. K., & Suvas, A. (2017). Financial distress prediction in an international context: A review and empirical analysis of Altman's z-score model. *Journal of International Financial Management & Accounting*, 28(2), 131-171. https://doi.org/10.1111/jifm.12053

Gold, A., Heilmann, M., Pott, C., & Rematzki, J. (2020). Do key audit matters impact financial reporting behavior? *International Journal of Auditing*, 24(2), 232-244. https://doi.org/10.1111/ijau.12190

Hopwood, W., McKeown, J. C., & Mutchler, J. F. (1994). A reexamination of auditor versus model accuracy within the context of the going-concern opinion decision. *Contemporary Accounting Research*, 10(2), 409-431. https://doi.org/10.1111/j.1911-3846.1994.tb00400.x

Hsu, Y., & Lee, W. (2020). Evaluation of the going-concern status for companies: An ensemble framework-based model. *Journal of Forecasting*, 39(4), 687-706. https://doi.org/10.1002/for.2653

International Auditing and Assurance Standards Board (IAASB). (2015b). Reporting on audited financial statements – New and revised auditor reporting standards and related conforming amendments. *International Federation of Accountants.* https://www.iaasb.org/system/files/publications/files/Basis%20for%20Conclusions%20-%20Auditor%20Reporting%20-%20final.pdf

Martens, Bruynseels, Baesens, Willekens, & Vanthienen, 2008. Predicting going concern opinion with data mining. *Decision Support Systems*, 45(4),765-777. https://doi.org/10.1016/j.dss.2008.01.003

Pelzer, J. R. E. (2016). Understanding barriers to critical audit matter effectiveness: A qualitative and experimental approach. *[Doctoral dissertation, Florida State University].* FSU's Digital Repository. http://fsu.digital.flvc.org/islandora/object/fsu%3A360417

PricewaterhouseCoopers (PwC). (2015). Delivering the value of the audit New insightful audit reports. https://www.pwc.com/gx/en/audit-services/publications/assets/pwc-auditing-report-new-insightful.pdf

Segal, M. (2017). ISA 701: Key audit matter – An exploration of the rationale and possible unintended consequences in a South African. *Journal of Economic and Financial Sciences*, 10(2), 376-391. https://doi.org/10.4102/jef.v10i2.22

Steer, T. (2018). In Profile Books (Ed.), The signs were there: The clues for investors that a company is heading for a fall. Profile Books; Main edition (November 29, 2018). doi: ASIN: B07CMDKVJ3

Velte, P., & Issa, J. (2019). The impact of key audit matter (KAM) disclosure in audit reports on stakeholders' reactions: A literature review. *Problems and Perspectives in Management*, 17(3), 323-341. http://dx.doi.org/10.21511/ppm.17(3).2019.26

Chapter I – Using Big Data Techniques to create a predictive analysis model to enhancing Auditors' procedures

1. INTRODUCTION

The basic characteristics of relevance and faithful representation of transactions are the fundamental qualitative characteristics for financial reporting in the International Accounting Standards Board (IASB) Conceptual Framework for Financial Reporting, and they are essential for proper financial risk assessment. Accounting standards require that financial information must be prepared under the assumption that there is no intention for liquidation due to bankruptcy/failure or cessation of business activities. There is an assumption that the entity will continue in operation for the foreseeable future, and therefore, there is no uncertainty about whether the firm will continue to be a "going concern". However, paragraph 25 of the International Accounting Standards (IAS) 1 – *"Presentation of Financial Statement"* states that *"When management is aware, in making its assessment, of material uncertainties related to events or conditions that may cast significant doubt upon the entity's ability to continue as a going concern, the entity shall disclose those uncertainties. When an entity does not prepare financial statements on a going concern basis, it shall disclose that fact, together with the basis on which it prepared the financial statements and the reason why the entity is not regarded as a going concern."* Thus, given the importance of this matter, many regulators have issued guidance for auditors about situations that should be considered in their assessment of potential going concern uncertainty and some key elements and best practices for financial reporting disclosures (FRC 2016; IASB 2021; PCAOB 2012; AICPA 2021). Under the ISA and the Generally Accepted Auditing Standards (GAAS), auditors must obtain sufficient appropriate audit evidence to conclude the appropriateness of management's use of the going concern basis of accounting in the preparation of their financial statements. If auditors conclude that a material uncertainty still exists, they are required to include a specific paragraph in their audit report (ASU 2014-15; ISA 570).

The onset of the COVID-19 pandemic has had an uneven impact on different industries in the world economy. It has triggered a wave of distress and bankruptcy all over the world.[1] Corporate bankruptcy in the U.S. reached a 10-year high in 2020.[2] Among those industries most affected, such as entertainment companies and oil and gas companies, 6,917 companies filed the required information of their status under Chapter 11 of the U.S. Bankruptcy Code, at least 30 percent higher than in any of the four years preceding 2020.[3] Other than wholesale and retail, service industries in the U.K. appear to be doing substantially worse than they were at the beginning of 2020.[4] In general, transportation, automotive, electronics, and retail were hit harder than other industries.[5] Confronted with different risks, such as decreasing customer demand or volatile financial market, different industries require different actions to respond the impact of COVID-19 (EY 2020).

Meanwhile, regulators released numerous alerts to auditors and management, emphasizing the importance of informing stakeholders of uncertainty in a company's continuity (AICPA 2020; FRC 2020; PCAOB 2020; IAASB 2020; CEAOB 2020; ESMA 2020). The auditor's role as protector of the capital markets has never been more critical (AICPA & CIMA 2020). However, as the importance of that role increases, so does the enormous pressure for auditors to find alternative ways to collect audit evidence and complete their work during the pandemic. Working remotely forced auditors to make full use of remote access to relevant files, workflow,

---

[1] https://www.washingtonpost.com/business/2021/02/26/pandemic-economy-bankruptcies/
[2] https://www.globest.com/2021/01/07/corporate-bankruptcies-end-2020-at-10-year-high/?slreturn=20220305200240
[3] https://www.washingtonpost.com/business/2021/02/26/pandemic-economy-bankruptcies/
[4] https://blogs.lse.ac.uk/businessreview/2020/05/07/how-is-covid-19-affecting-businesses-in-the-uk/
[5] https://www.allianz-trade.com/en_global/news-insights/economic-insights/no-stone-unturned-how-covid19-is-disrupting-every-industry.html

and file sharing solutions. Even still, they were less likely to be present to observe how companies were handling the crisis, posing an unprecedented challenge to auditors issuing going concern opinions (Wilson 2020).

There is no unique method in the literature for evaluating a company's going concern situation that could provide absolute assurance about the inclusion of a going concern paragraph in the auditors' opinion (going concern opinion [GCO]). Some researchers indicate that only quantitative indicators of a company would be sufficient to assess a potential going concern situation, and academic literature suggests that Z-Score is still a very reliable indicator to predict possible bankruptcy (Altman 1968; Altman et al. 2014). However, auditor judgement also influences the assessment of going concern status, which suggests that qualitative factors are necessary components of going concern assessment (Hopwood et al. 1994). Auditor judgment could be influenced by an auditor's Firm culture, experience, training, and size of the auditor's firm (Tagesson and Öhman 2015; Svanberg and Öhman 2016).

Machine learning, which is now often used in a variety of fields, could improve data analysis and lead to more evidence-based decision-making (Jordan and Mitchell 2015). This paper aims to provide a two-fold strategy to prepare a machine learning-based automated predictive model that would consider both quantitative and qualitative indicators in order to assist auditors in their going concern assessment. The low explainability of opaque models (i.e., "black-box") is the critical challenge that holds back auditors from adopting complex machine learning algorithms for their decision making (AICPA 2020; CPAB 2021). The rule-based Decision Tree is a white-box method that provides a transparent step-by-step procedure and, therefore, more clear audit evidence than impenetrable black-box machine learning algorithms.

Data used to build the model was obtained from the 2019 audit results of an auditing Firm in Spain. Based on research about the use of Z-Scores in the determination of potential bankruptcies (Altman 1968) and some other specific variables related to recent accounting scandals (Steer 2018), 22 variables relating to 2,909 companies were obtained for building a machine learning-based automated predictive model for the inclusion of a Going Concern paragraph (the main variable to be predicted – YES/NO). The qualitative variables from the auditor assessment were based on: 1) auditors' knowledge about the entity's risk, according to their experience and expertise, and 2) the auditing firm's risk assessment of the industry of the entity being audited. The quantitative variables from the company's financial ratios include widely used indicators, such as calculated Z-Scores, revenue/assets, and working capital/assets.

Classification differences between the model and the final auditors' report in 2020 were also investigated. In more than half of all cases, three main types of circumstances (subsequent financing, group financial support, or subsequent improvements to cash flow) explained the inconsistencies between the model results and the final outcome in the auditors' report. This study provides empirical evidence that automated predictive models can be beneficial for assisting auditors in concluding in a critical area, such as evaluating the necessity to include a going concern paragraph in their auditors' report. It also suggests that no model can provide absolute assurance. Recalibration of the model is critical and must be an ongoing analysis year after year to account for future economic changes.

The proposed mixed qualitative and quantitative model contributes to the literature by offering an efficient predictive Decision Tree white-box model. In addition, any Audit Firm elsewhere could benefit from using the methodology. All partners who have used the provided data shared the same cultural values, as they belong to the same firm and have a common perception of going concern risks in relation to the entities being audited.

The resulting proposed model is highly effective, with an accuracy of 0.83, and demonstrates significant benefits in practice. It can also be potentially used by auditors in other firms or jurisdictions since the qualitative factors could be adapted to any environment. Consequently, this paper provides valuable information about the use of machine learning in auditors' going concern assessments and how to improve future Decision Tree models' overall audit quality, and thus enhances the protection of the public interest.

The remainder of this paper is organized as follows. Section 2 reviews the literature on the going concern opinions and their prediction model and variables. Section 3 introduces the data and methodology. Section 4 shows the empirical results. Section 5 discusses current limitations and opportunities for future research. Section 6 concludes the paper.

2. LITERATURE REVIEW

2.1. Importance of GCO

Management should undertake the going concern evaluation first when preparing the financial statements for the entity (ASU 2014-15; IAS 1). A diligent going concern assessment made by management is critical for the public interest and economic stability since potential uncertainties could change the decision of investors and other market participants and, consequently, could impact other market participants such as banks and corporate or public services (Zéman & Lentner, 2018).

In those jurisdictions where an independent audit is required, or for those entities that have engaged an independent auditor to perform an audit of their financial information under auditing standards (such as IAS and GAAS), the auditor must follow specific standards and procedures to evaluate management's assessment of the entity's ability to continue as a going concern covering at least the same period as the one used by management. According to auditing standards, the auditor shall verify, based on the auditor's knowledge and all evidence obtained during the audit, whether management has made a proper assessment and considered all relevant information. Investors consider the going-concern opinion relevant for valuing a company's common stock and, therefore, relevant to pricing stocks (O'Reilly 2010). In addition, going concern opinions are helpful in predicting bankruptcy, and they can provide some explanatory power in predicting bankruptcy resolution (Chen & Church 1996; Bessell et al. 2003). Ajona et al. (2012) find that the inclusion of the auditor's paragraph in going concern situations in Spain is critical, as many companies have gone bankrupt in Spain after the inclusion of such paragraph.

2.2. Prediction models

Several researchers have focused on going-concern prediction after McKee (1976). Bellovary et al. (2007) identified 27 statistical models that have been developed for predicting the going concern opinion, such as multivariate discriminant analysis (MDA), logit analysis, probit analysis, neural networks (Serrano-Cinca 1996), and a random forest model (Hsu and Lee 2020). Statistical methods help assist auditors in issuing going concern opinions (Koh 1991). But auditors' opinions are inferior bankruptcy indicators relative to the predictions of statistical models (Hopwood et al. 1994). The more complex algorithms usually have better performance (Zhang et al. 2021).

However, there is a tradeoff issue between machine learning algorithm performance and explainability (Zhang et al. 2022; Virág and Nyitrai 2014). As the machine learning model includes more variables, increases dimensionality, and uses more sophisticated calculations, predictability improves, but explainability decreases (Zhang et al. 2022; DARPA 2016; Baryannis et al.2019). The low explainability of opaque models (i.e., "black-box" models) is a critical challenge that holds back auditors from adopting complex machine learning algorithms for their decision making (AICPA 2020; CPAB 2021). The existing standards (e.g., AS 1105 and AS 1215) require an auditor to explain and document the result of any machine learning models used, a task black-box models make difficult (Zhang et al. 2022; AICPA 2020; CPAB 2021). Using an explainable white-box model would alleviate this issue. A white-box model, like a Decision Tree, has two key characteristics: 1) the features must be understandable, and 2) the machine learning process must be transparent.[6] (Zhang et al. 2022; Hall and Gill 2019; Molnar 2021). Therefore,

[6] https://www.siliconrepublic.com/enterprise/white-box-machine-learning#:~:text=There%20are%20two%20key%20elements,and%20decision%2Fregression%20tree%20models.

one of this paper's contributions is using such a white-box technique, providing a transparent method that could be easily applied in practice.

2.3. Prediction variables

The decision to issue a GCO is complex and involves the issuing auditor's judgment. Determinants of GCOs include client factors, auditor factors, auditor-client relationships, other environmental factors, etc. (Carson et al. 2013; Brunelli 2018). They can be divided into quantitative (objective) and qualitative (subjective) variables. Therefore, audit quality and good audit judgment are based upon various qualitative factors (Francis 2004).

Quantitative analysis for going concern assessment is necessary. Financial distress, debt default, and leverage have a significant influence on the auditor's going concern opinion (Achyarsyah 2016). Most archival studies focus on quantitative variables to measure the distress level of firms. There are two ways to measure distressed firms: 1) market value for distance to distress (Merton 1974), calculated by the firm's market value minus the value of debt, and then divided by the volatility of the asset, and 2) financial ratios. Altman's Z-Scores dominate the area. Using a combination of several quantitative financial ratios, a score could be calculated (Z-Score) and used with a high degree of accuracy to determine the risk of a potential bankruptcy (Altman 1968; Altman 1983; Altman 2013; Altman et al. 2017). It could also be adopted as a combined model of accounting and auditing data (Muñoz-Izquierdo et al. 2020). The Altman model predicts bankruptcy in a significant majority of companies (Salimi 2015) in the international context (Altman et al. 2014), in the U.S. (Altman et al. 2017), in Europe (Chieng 2013), and in Spain (Fitó Bertran et al. 2018). However, Carreras Peris (2017) argue that Z-Scores may not be helpful in construction companies in Spain compared to Ohlson's model (1980) and Ismail's model (2014). Additional quantitative KPIs could be useful to enhance any proposed predictive bankruptcy model (Steer 2018).

The qualitative factors of GCO are essential but ignored when building many automated models. In other words, the auditors' going concern decision is very subjective, which may influence audit quality (Harris and Harris 1990; Haron et al. 2009; Lipe 2008). Matsumura et al. (1997) analyze a game-theoretic model in which a client can potentially avoid a going-concern opinion and find that the auditor's forecast of entity viability impacts the auditor's tendency to express fewer going-concern opinions. Also, auditors' experience in audit and a client's industry are key to producing high quality work and accurately assessing potential going concern (Blandón et al. 2020). Auditor characteristics may also influence the output of the GCO assessment (Carson et al. 2013). Studies examining the effects of audit size have had mixed results. Tagesson and Öhman (2015) find a positive relationship between audit fee amounts and the likelihood of including a going concern paragraph in the audit report, and demonstate that Big 4 (used to refer collectively to the four largest professional auditing service networks in the world: Deloitte, EY, KPMG, and PwC) auditors are more likely to issue such warnings than other auditors. However, Gallizo Larraz and Saladrigues Solé (2016) find that a small-scale auditor is more likely to issue a going concern audit opinion. In addition, a skeptical audit culture is more likely to maintain auditor objectivity than less supportive cultures, emphasizing the importance of office culture in the assessment of going concern judgment (Svanberg and Öhman 2016). Spain's audit quality (measured by auditors' independence and knowledge) also affects the probability that a financially distressed company will receive a going-concern opinion (Ruiz-Barbadillo et al. 2004).

3. DATA AND METHODOLOGY

3.1. Data

During the COVID-19 pandemic at the beginning of 2020, auditors have been warned by regulators that they must perform a careful analysis of material uncertainties about companies' ability to apply the going concern basis of accounting because many industries would be facing unfavorable future outcomes caused by the pandemic (AICPA 2020; FRC 2020; PCAOB 2020; IAASB 2020; CEAOB 2020; ESMA 2020).

The auditing firm in this paper used a scorecard checklist that asked partners to include some financial variables about the entity being audited to classify cases that would require careful analysis to conclude whether the inclusion of a GCO would be necessary. In addition, some qualitative industry risk evaluations were included. An example of this checklist is shown in Appendix A. It contains a going concern risk score for each engagement to assist auditors and their firm in developing a decision-making system to evaluate whether to include a GCO. The data obtained in the scorecard checklist was analyzed to set up a simple machine learning white-box tool based on data that could assist auditors in deciding about the potential inclusion of a GCO. The timeline for data obtained is shown in Figure 1.

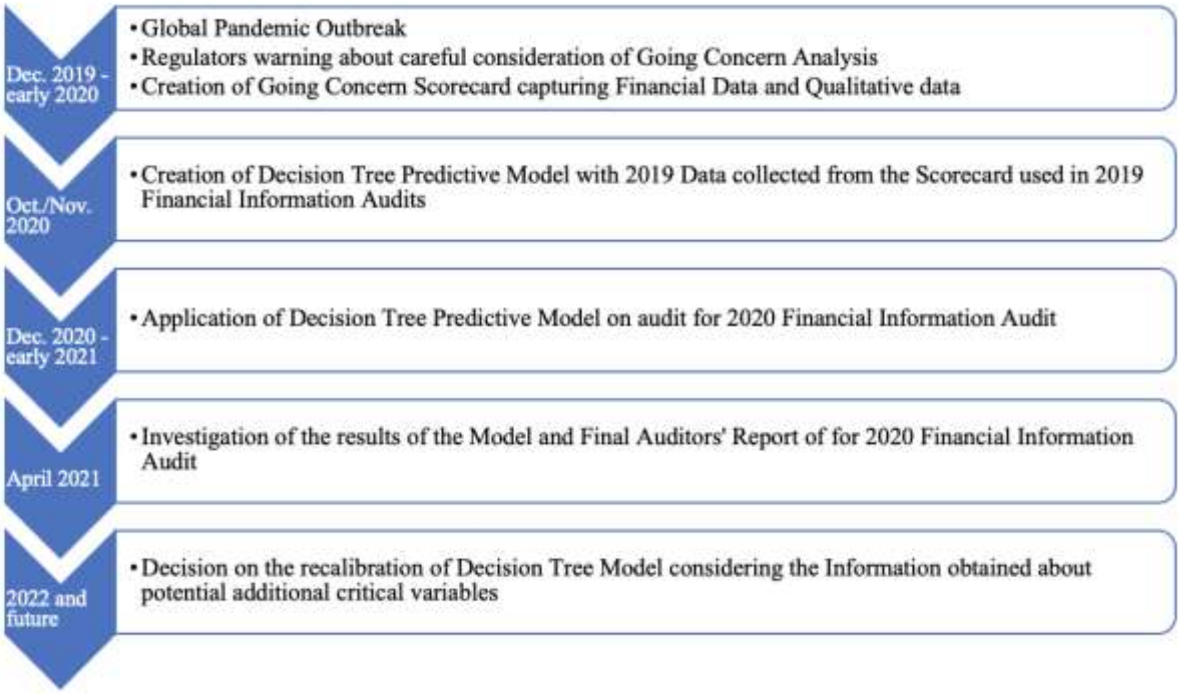


Figure 1. Timeline for Data Obtaining

Table 1 displays the sample composition of 2,909 cases in the auditing firm in 2019. Out of the total 2,909 cases, 133 (4.6%) cases were issued with going concern opinion paragraphs in the audit reports, and 2,776 (95.4%) cases were without going concern paragraphs in the audit reports. There were 22 variables obtained for each case, and there is no missing data.

Table 1. Sample Composition

Year	2019
The number of cases	2,909 (100%)
The number of real GCO - NO	2,776 (95.4%)
The number of real GCO- YES	133 (4.6%)
The number of variables per case	22

The data of 22 variables, as shown in Appendix B, is obtained for all audited entities for the 2019 audits with the following three key categories: the response variable is a dichotomous dummy variable representing whether the auditor's report had included a going concern paragraph (1) or not (0); the predictor variables include qualitative variables of partner risk assessment; and the predictor variables include quantitative variables representing the financial results of audited entities. To give their score, partners must consider financial, operating, and other circumstances for risk assessment (ISA 570 paragraph A3-A6).

3.2. Using a Decision Tree to model GCO

The data obtained in 2019 was used to build the Decision Tree model for estimating the inclusion of GCO. Then the Decision Tree, based on 2019 data, was applied in the 2020 audit process, shown in Figure 2.

Figure 2. Data Flow Design

The 2019 dataset was split into two parts: 80% for training, and 20% for testing. In order to optimize the pruning of the decision tree, the Complexity Parameter (CP) is used as a hyperparameter (Therneau et al. 2013). Ten-fold cross-validation was used on the training data to select the optimum hyperparameter value, and the *Area Under the ROC Curve* value was used as the primary performance metric. The number of positive going concern cases in the 2,909 companies used for training the model was limited (133 YES and 2,776 NO), which could cause the model to give biased results. In fact, following the aforementioned process, a tree model with high accuracy was obtained (up to 96% of correctly classified cases) but with a low level of sensitivity[7] (0.22 in the test set – i.e., 22% of companies in the test set with 'YES' CGO were correctly identified). This result implies that the initial model would not have been useful for predicting GCO because auditors' decisions would misclassify "YES" cases.

Consequently, for an adequate calibration of the model, some techniques for handling class imbalance problems using oversampling techniques were implemented (Gosain and Sardana 2017). All training data cases with YES have been "over-sampled" to obtain an equal number of YES and NO cases, penalizing the misclassification of YES samples. In addition, the length of the tree had to be adjusted by an appropriate selection of its Complexity Parameter (CP) to provide an appropriate balance between precision and complexity. Based on the data obtained, a simple and intuitive Decision Tree without too many decision nodes could provide a high level of accuracy, thereby helping auditors assess a going concern uncertainty situation on their audits. Figure 3 shows the diagram of CP and the performance of the model via ROC.

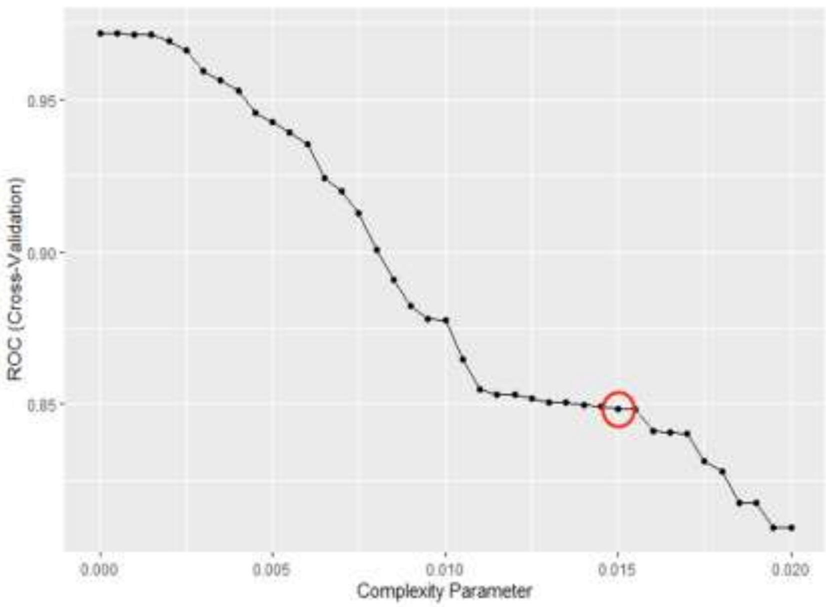


Figure 3. Complexity Parameter and ROC

[7] Same as True Positive Rate. Refers to the percentage of 'YES' CGO companies the model was able to correctly identify.

In this case, we constructed the Decision Tree with a Complex Parameter of 0.015, which results in an equilibrium of high accuracy and efficiency. The model's overall accuracy is high at 0.83, with sensitivity levels up to 0.82707 (enabling an almost 83% prediction of YES GOC Cases in the final model), as demonstrated in Table 2. We used the R-based Rpart library to build the model. Once the optimal complexity was selected, the final Decision Tree model, using all samples in the dataset, was constructed to maximize the amount of data available for the tree to learn. Cross-validation performance for this final tree showed results similar to its earlier incarnation, indicating that the model was not overfit. The main resulting figures of the model are as follows:[8]

Table 2. Final model Performance Metrics on All Samples

Accuracy	0.83
Sensitivity	0.82707
Specificity	0.82745
Pos Pred Value	0.1867
Neg pred Value	0.99
Prevalence	0.045
Detection Rate	0.037
Detection Prevalence	0.202
Balanced Accuracy	0.827

[8] The script is available to public. https://github.com/yugu431/Decision-Tree_Going-Concern and in Appendix C

4. THE EMPIRICAL RESULTS

Figure 4 demonstrates the Decision Tree model output from R. The deeper the orange of the nodes and leaves, the more likely the result is "YES." The Decision Tree goes through at most seven and at least three decisions to determine the issuance of GCO. Figure 5 illustrates the importance of each variable to the final decision.

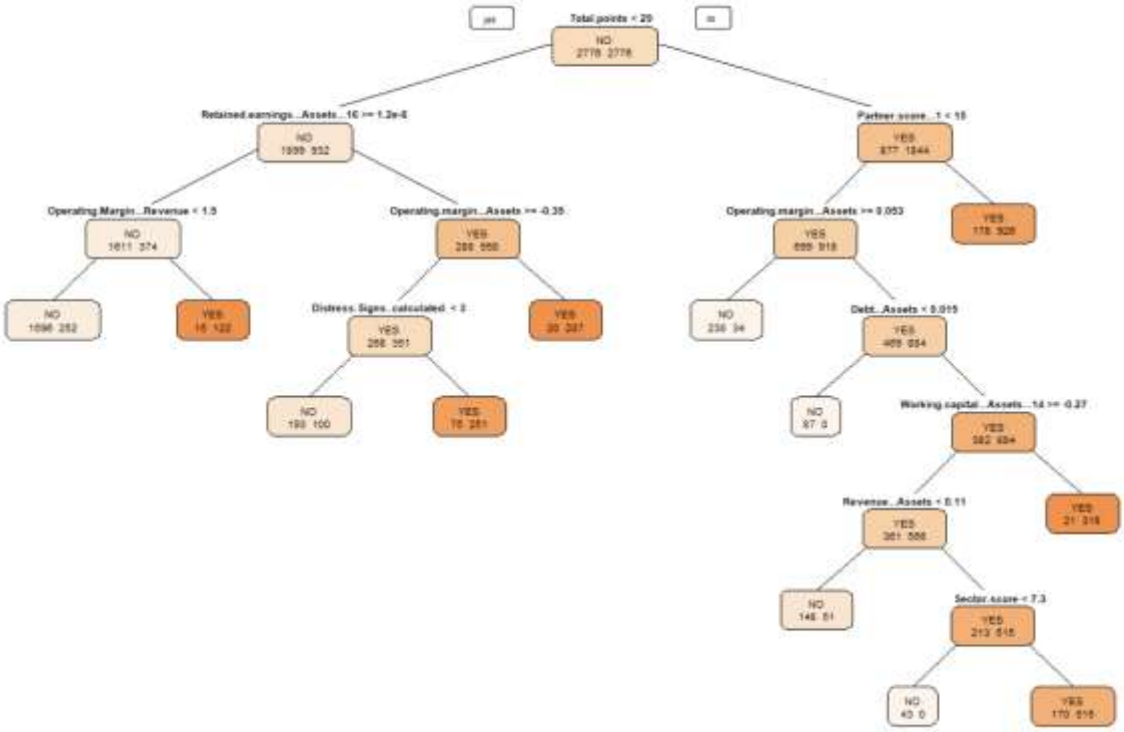


Figure 4. The Decision Tree

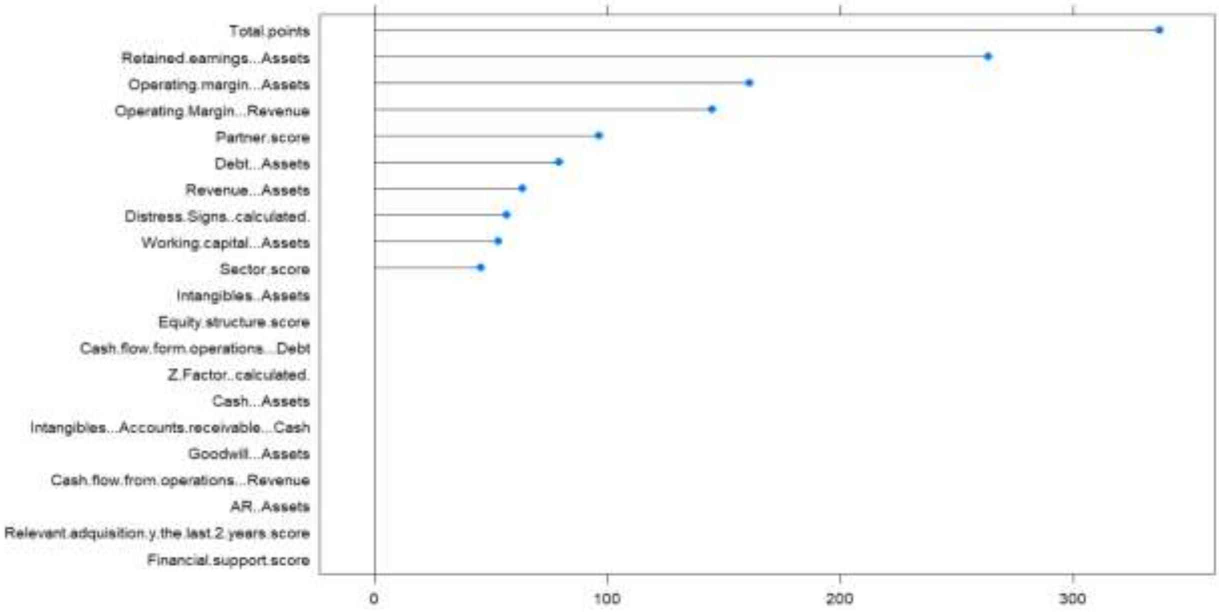


Figure 5. The Main Variables that the Final Decision Tree Considers Relevant

Figure 6 depicts the conceptual Decision Tree model. The green boxes identify qualitative factors: 1) *Partners Score for the Company*; 2) *Firm Score for the Industry*. The blue boxes are the quantitative factors: 1) *Total Points the Going Concern Score Card Checklist* 2) *Operating Profit/Assets*; 3) *Debt/Total Assets*; 4) *Working Capital/Assets*; 5) *Sales Revenues/Assets*; 6) *Other Equity Reserves/Assets*; 7) *Operating Profit/Sales Revenues*; 8) *Operating Profit/Assets*; 9) *Checklist Distress Sign Score*. *Operating Profit/Assets* appears twice in the conceptual Decision Tree model in Figure 6. The blue arrows pointing to grey output areas above and below indicate recommendations for whether to include a going concern paragraph in the report based on the values of the relevant boxes. For example, the first decision, and the most consequential factor, checks whether an entity's *Total points in the Going Concern Scorecard Checklist* value is greater than or equal to 29. If it is, then a value for *Partners' Score for the Company* greater than or equal to 14 would indicate that auditors should consider including a Going Concern paragraph in the auditors' report. If the *Partner Score* value is below 14, the next decision considers the *Operating Profit/Asset* ratio. If the ratio is greater than or equal to 0.053, the model indicates that auditors should not issue a GCO. If the *Operating Profit/Asset* ratio is less than 0.053, and the *Debt/Total Assets* ratio is less than 0.015, there would be no GCO issued. If the *Working Capital/Assets* ratio is less than -0.27, which means the entity does not have working capital to support the business operations, the model would suggest including a going concern special section in the auditor's report. Otherwise, the Decision Tree goes to the next decision that considers Sales Revenue/Assets. If the ratio is lower than 0.11, there is no GCO concern. If not, the model will continue to the final decision, a subjective indicator of the Firm Score for the Industry. If the score is greater than or equal to 7, which is identified as a high-risk industry, the model would suggest including a going concern special section in the auditor's report. Oppositely, the final result of the model would be no such a going concern special section in the auditor's report.

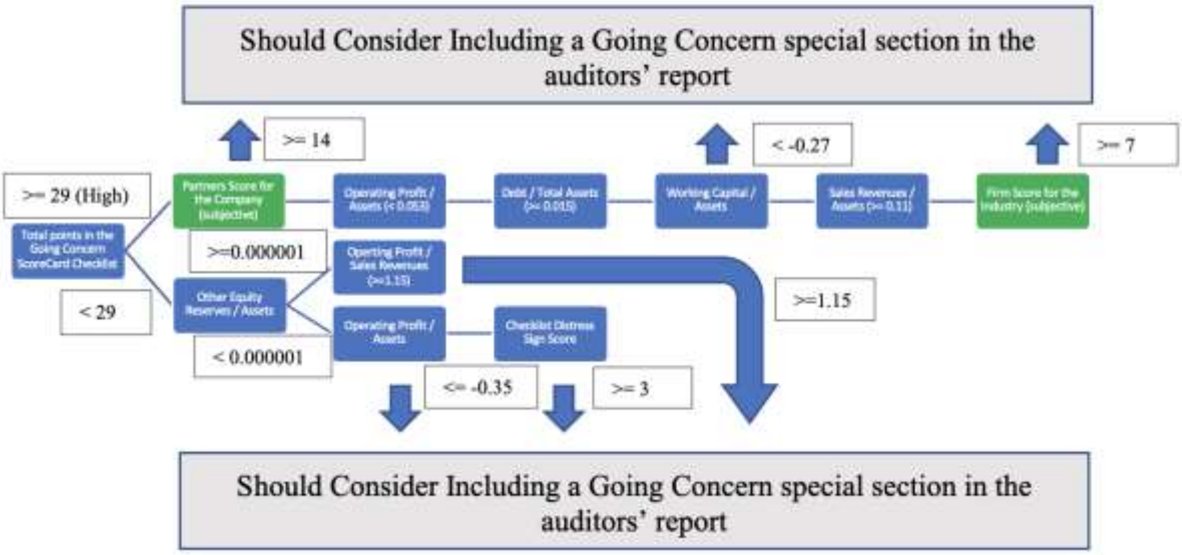


Figure 6. The Conceptual Decision Tree Model

This Decision Tree model constructed with 2019 data was utilized for the audit of 2020 to assist auditors performing a going concern assessment in an auditing Firm in Spain. Auditors considered the result generated by the Decision Tree model. As shown in Table 3, 93% of the final decision to include a going concern was consistent with the auditor' final GCO conclusions.

Table 3. The Confusion Matrix

	Real = No	Real = Yes	
Model = No	2,215	11	
Model = Yes	243	146	93.0%[9]
	9.3%[10]		

However, the model was not perfect. As seen in Table 3, there are 243 cases where the Decision Tree model indicated a GCO but auditors chose not to include the GCO in the audit report. Similarly, there are 11 cases where the Decision Tree model did not indicate a GCO but auditors decided to include the GCO in the audit report. It is worth noting that these results follow a similar distribution of errors as the training results obtained during the fitting process.

A root cause analysis was performed in a random selection of 150 cases (more than half of the population of inconsistencies, or 61.7% of 243 cases), which provided some insight: additional variables might have had a significant impact and would require further evaluation to determine whether they would need to be considered in future investigations. Figure 7 shows a survey of these 150 cases. There are three reasons why the model indicated "YES," but the auditors opted for "NO." First, in 36% of the cases, there is evidence that the company could obtain group financial support. Second, 21% of cases are a result of seeking subsequent external financing. Third, in 43% of the cases, there is subsequent positive evidence of cash flows. Companies need to show their effective mitigation plans that can bring cash flows to keep them afloat for at least twelve months and alleviate substantial doubt (ESMA 2020; Dohrer and Mayes 2020; Wang 2021). Wang (2021) also finds that raising debt, restructuring debt, growing revenue, and selling assets extracted from financial reports might all help to mitigate the unfavorable market reaction. FRC (2020) emphasize mitigation actions in their review of the

[9] $93.0\% \approx 146/(11+146)$
[10] $9.3\% \approx 243/(2215+11+243+146)$

financial reporting effects of COVID-19. They expect sufficiently granular plans in their financial reports that can help auditors and other users understand clearly whether the company will survive for at least twelve months.

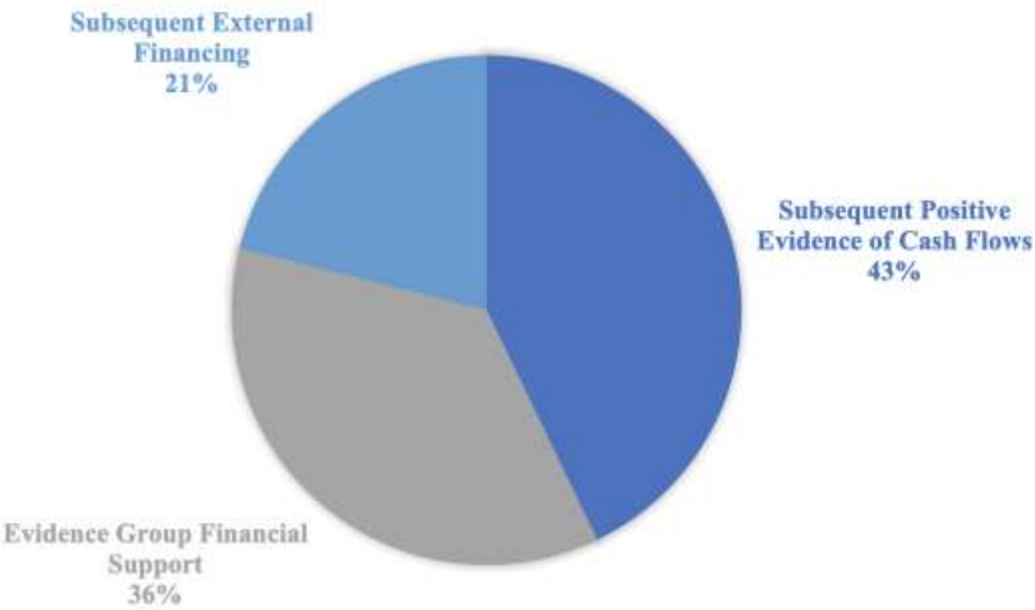


Figure 7. Investigation of 243 cases that Model Indicated “YES” and Auditor Opted “NO”

Similarly, we also examined where the Decision Tree model indicated there should be no GCO while the auditors chose to include a going concern opinion in their audit report. The results of 10 cases (90.9%) of the total 11 cases are shown in Figure 8. Most inconsistencies are due to random factors rather than the model itself. For example, inconsistencies in 30% of cases are due to changes in circumstances between the date when the model was applied and the conclusion date of the final report (e.g., because new lines of credit have been obtained by the time auditor was issuing their report that where not availbe before). Another 30% of cases are due to going concern in parent-level subsidiary contamination, whereby the parent company of the group had going concern opinions that could affect subsidiaries. Two cases are due to a human error response in the checklist. One is a borderline case, in which the number is close to the threshold of the going concern decision, and the last one is for other reasons not considered relevant for the investigation.

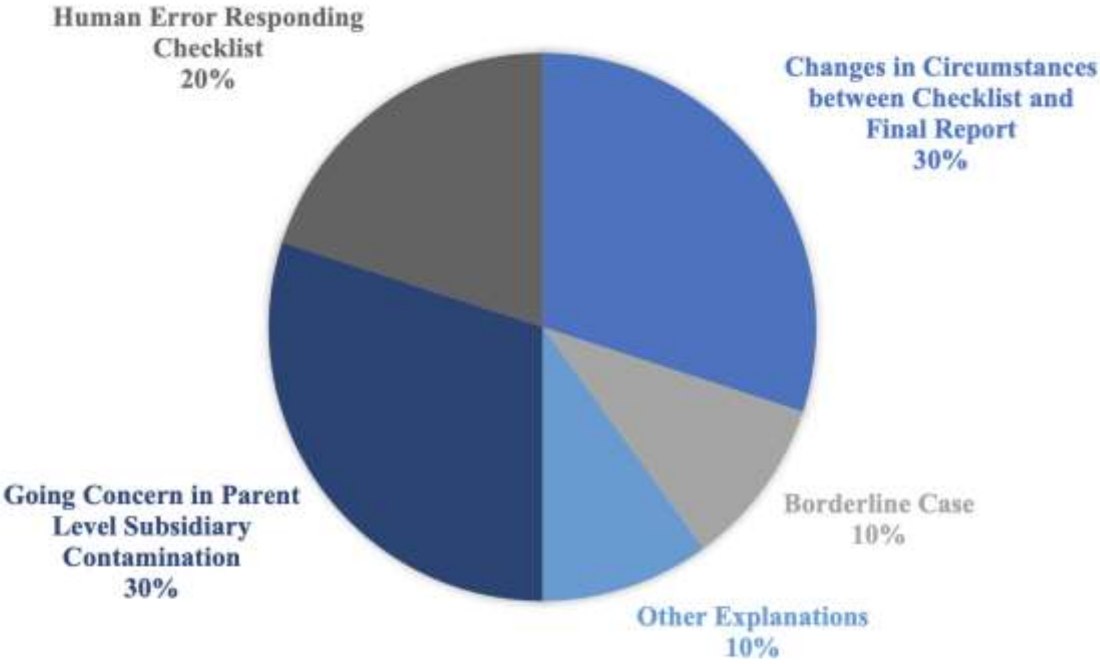


Figure 8. Investigation of 10 cases that Model Indicated “NO” and Auditor Opted “YES”

5. DISCUSSION

A Decision Tree automated model is proposed as an aid to improve and enhance the evaluation and documentation of auditors' GCO assessment, in part as a response to increasing concern about the GCO decision-making process because of the COVID-19 pandemic. The model produced highly accurate predictions after being validated and employed by an auditing firm in Spain. This methodology assisted auditors in documenting their assessment by embedding their auditors' judgment. It could also be beneficial for regulators when considering the white-box machine learning's capacity to depict auditors' decision-making process.

Explanations for inconsistencies between the model and auditor behaviors are explored. As indicated above, the reasons are clear for the False Negative cases (Model indicating "NO" and auditors opt "YES") and False Positive cases (Model indicating "YES" and auditors opt "NO"). Therefore future work and future Decision Tree model should consider the new variables, such as mitigation plans (whether firms have external financing, or the coverage ratio of the liability), in order to increase efficiency. Also, the model should be continuously updated with new data each year, as Figure 9 displays. More future research based upon the results obtained and further investigation could be performed, aiming for a model recalibration each year.

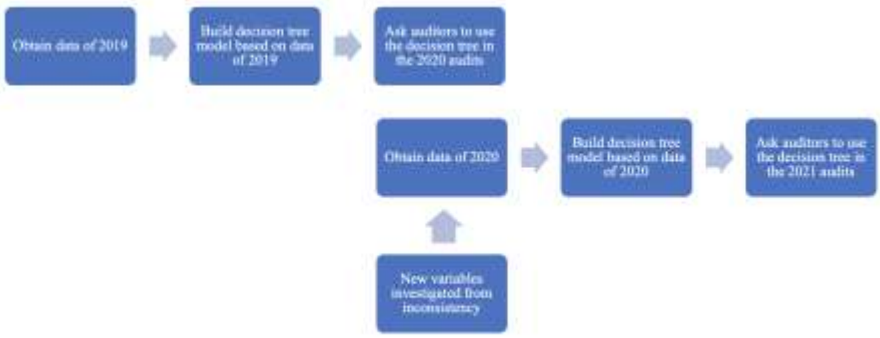


Figure 9. Future Research Directions

This paper has three limitations. First, this work has been performed on a set of Spanish data for one audit firm. As firm cultures vary among firms and countries, the final Decision Tree predictive model may not be entirely applicable for any other firm or country. Each firm should consider its own data and variables, especially subjective variables, and create its own Decision Tree. However, the procedure established in this study could be easily extrapolated and could benefit firms in setting up their predictive model based upon white-box machine learning technology. Second, the model has been built considering the available data for a short period (one year, 2019). Therefore, the model could have been different or even more accurate if more periods were to be included to build the model. Third, this Decision Tree model was not designed to predict real business failure. Rather it was built to aid auditor's decision-making processes by formalizing auditors' past judgment history with qualitative and quantitative data. To compare with the real business failure could be an avenue for future research.

6. CONCLUSION

COVID-19 posed considerable pressures and difficulties for auditors to assess the going concern situations of the audited entities. The standards (ISA 570; ASU 2014-15) require auditors to apply professional judgment on going concern issues. This study provides a tool to aid auditors in their assessment of the risk of entity failure and, consequently, their analysis of whether or not to include a going concern paragraph in their auditor report as required by the applicable auditing standards. The automated tool is a Decision Tree that would help auditors decide if their report should include a going concern opinion. The resulting predictions are significantly similar to actual auditors' decisions, suggesting the model is effective for its purpose in providing additional evidence about conclusion that has been reached by auditors.

The contribution in this paper is to prove a two-fold strategy to prepare an intuitive, white-box, and easy-to-use predictive model based upon simple Decision Tree questions that incorporate qualitative and quantitative data to assist auditors in their going concern assessment. Quantitative indicators would consider an entity's financial figures based on Z-Score and supported by other quantitative indicators. The qualitative financial indicators would also add important information about: 1) auditors' knowledge about the entity's risk considering their experience and expertise, and 2) the auditing firm's risk assessment of the company's industry being audited. Considering that all data used in this study come from one audit firm in one country, there is a risk that different audit quality/culture could impact results, as qualitative scores may be perceived differently. The data used for this study was obtained from a single firm. It can be assumed that the risk perception of its partners is consistent because they share a common training experience and audit methodology. This research could be easily replicated in other firms/cultures by using their risk assessment (qualitative indicators) methods. Auditors should consider making use of this model by inputting their data and elaborating their prediction model, as it can be a beneficial tool in evaluating whether a going concern paragraph is needed in their auditors’ report. This paper also serves as an example for regulators on how to apply machine learning for better quality audits.

References

Achyarsyah, P. (2016). The analysis of the influence of financial distress, debt default, company size, and leverage on going concern opinion. *International Journal of Applied Business and Economics Research*, 14(10), 6767-6783. http://repository.unas.ac.id/id/eprint/62 Accessed 27 April 2022.

Adler, P., Falk, C., Friedler, S. A., Nix, T., Rybeck, G., Scheidegger, C., ... & Venkatasubramanian, S. (2018). Auditing black-box models for indirect influence. *Knowledge and Information Systems*, 54(1), 95-122. https://doi.org/10.1007/s10115-017-1116-3

AICPA. (2020). The Data-Driven Audit: How Automation and A.I. are Changing the Audit and the Role of the Auditor. https://www.aicpa.org/content/dam/aicpa/interestareas/frc/assuranceadvisoryservices/dow nloadabledocuments/the-data-driven-audit.pdf Accessed 27 April 2022.

AICPA&CIMA. (2020). As auditors work with clients throughout the pandemic, there is a real possibility that scope limitations will occur. https://www.journalofaccountancy.com/issues/2020/may/auditing-during-coronavirus-pandemic.html Accessed 27 April 2022.

Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, 23(4), 589-609. https://doi.org/10.2307/2978933

Altman, E. I. (1983). A complete guide to predicting, avoiding, and dealing with bankruptcy John Wiley & Sons Inc; 4th edición (30 abril 2019).

Altman, E. I. (2013). Handbook of research methods and applications in empirical finance. Handbook of research methods and applications in empirical finance (pp. 428-456) *Edward Elgar Publishing*.

Altman, E. I., Iwanicz-Drozdowska, M., Laitinen, E. K., & Suvas, A. (2017). Financial distress prediction in an international context: A review and empirical analysis of Altman's z-score model. *Journal of International Financial Management & Accounting*, 28(2), 131-171. https://doi.org/10.1111/jifm.12053

Appelbaum, D., Budnik, S., & Vasarhelyi, M. (2020). Auditing and accounting during and after the COVID-19 crisis. *The CPA Journal*, 90(6), 14-19.

AS 1105, Audit Evidence.

AS 1215, Audit Documentation.

Baryannis, G., Dani, S., and Antoniou, G. (2019). Predicting supply chain risks using machine learning: The tradeoff between performance and interpretability. *Future Generation Computer Systems*, 101, 993-1004. https://doi.org/10.1016/j.future.2019.07.059

Bellovary, J. L., Giacomino, D. E., & Akers, M. D. (2007). A review of going concern prediction studies: 1976 to present. *Journal of Business & Economics Research*, 5(5). https://doi.org/10.19030/jber.v5i5.2541

Brunelli, S. (2018). *Audit Reporting for Going Concern Uncertainty: Global Trends and the Case Study of Italy.* Springer.

Canadian Public Accountability Board (CPAB). (2021). Technology in the audit. https://www.cpab-ccrc.ca/docs/default-source/thought-leadership-publications/2021technology-audit-en.pdf?sfvrsn=f29b51ce_14 Accessed 27 April 2022.

Carreras Peris, B. (2017). Business bankruptcy prediction models: Application to companies in the construction sector in Spain.

Carson, E., Fargher, N. L., Geiger, M. A., Lennox, C. S., Raghunandan, K., & Willekens, M. (2013). Audit reporting for going-concern uncertainty: A research synthesis. *AUDITING: A Journal of Practice & Theory*, 32(Supplement 1), 353-384. https://doi.org/10.2308/ajpt-50324

Chieng, J. (2013). Verifying the Validity of Altman's Z" Score as a Predictor of Bank Failures in the Case of the Eurozone (Doctoral dissertation, Dublin, National College of Ireland).

DARPA. (2016). Broad Agency Announcement Explainable Artificial Intelligence (XAI). https://www.darpa.mil/attachments/DARPA-BAA-16-53.Accessed 27 April 2022.

Dohrer, B., & Maye, C. (2020). Auditing during COVID-19: 6 areas to focus on. https://www.journalofaccountancy.com/issues/2020/may/auditing-during-coronavirus-pandemic.html Accessed 27 April 2022.

EMSA. (2020). ESMA's COVID-19 Related Recommendations for Financial Market Participants. https://www.jonesday.com/en/insights/2020/03/esmas-covid19-recommendations Accessed 27 April 2022.

EY. (2020). How can your industry respond at the speed of COVID-19's impact? Available at: https://www.ey.com/en_es/covid-19/how-can-your-industry-respond-at-the-speed-of-covid-19s-impact Accessed 27 April 2022.

Fitó Bertran, M. À, Pla Erta, D., & Llobet, J. (2018). Usefulness of Z scoring models in the early detection of financial problems in bankrupt Spanish companies. *Intangible Capital*, 14(1), 162-170. http://dx.doi.org/10.3926/ic.1108

Francis, J. R. (2004). What do we know about audit quality? The British Accounting Review, 36(4), 345-368. https://doi.org/10.1016/j.bar.2004.09.003

Gallizo Larraz, J. L., & Saladrigues Solé, R. (2016). An analysis of determinants of going concern audit opinion: Evidence from Spain stock exchange. Intangible Capital, vol.12, núm.1, p.1-16. http://dx.doi.org/10.3926/ic.683

García-Blandon, J., Argilés-Bosch, J. M., & Ravenda, D. (2020). Learning by Doing? Partners Audit Experience and the Quality of Audit Services: Learning by Doing? La experiencia en auditoría del socio auditor y la calidad de los servicios de auditoría. *Revista de Contabilidad-Spanish Accounting Review*, 23(2), 197-209. https://doi.org/10.6018/rcsar.366921

Gosain, A., & Sardana, S. (2017, September). Handling class imbalance problem using oversampling techniques: A review. In 2017 *international conference on advances in* *computing, communications and informatics* (ICACCI) (pp. 79-85). IEEE. https://doi.org/10.1109/ICACCI.2017.8125820

Hall, P., and Gill, N. (2019). An introduction to machine learning interpretability. O'Reilly Media, Incorporated.

Haron, H., Bambang, H., Ansari, M., & Ishak, I. (2009). Factors influencing auditors' going concern opinion. *Asian Academy of Management Journal*, 14

Harris, C. R., & Harris, W. T. An expert decision support system for auditor going concern evaluations. Paper presented at the Proceedings of the 1990 Symposium on Applied Computing, 155-160.

Hopwood, W., McKeown, J. C., & Mutchler, J. F. (1994). A reexamination of auditor versus model accuracy within the context of the going-concern opinion decision. *Contemporary Accounting Research*, 10(2), 409-431. https://doi.org/10.1111/j.1911-3846.1994.tb00400.x

Hsu, Y., & Lee, W. (2020). Evaluation of the going-concern status for companies: An ensemble framework-based model. *Journal of Forecasting*, 39(4), 687-706. https://doi.org/10.1002/for.2653

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science,* 349(6245), 255-260. https://doi.org/10.1126/science.aaa8415

Koh, H. C. (1991). Model predictions and auditor assessments of going concern status. *Accounting and Business Research*, 21(84), 331-338. https://doi.org/10.1080/00014788.1991.9729848

Lipe, M. G. (2008). Discussion of "judging audit quality in light of adverse outcomes: Evidence of outcome bias and reverse outcome bias". *Contemporary Accounting Research*, 25(1), 275-282. https://doi.org/10.1506/car.25.1.11

Matsumura, E. M., Subramanyam, K. R., & Tucker, R. R. (1997). Strategic auditor behavior and going-concern decisions. *Journal of Business Finance & Accounting*, 24(6), 727-758. https://doi.org/10.1111/1468-5957.00131

Merton, R. C. (1974). On the pricing of corporate debt: The risk structure of interest rates. The Journal of Finance, 29(2), 449-470. https://doi.org/10.2307/2978814

McKee, T. (1976, August). Discriminant prediction of going concern status: A model for auditors. In Selected Papers of the AAA Annual Meeting.

Molnar, C. (2020). Interpretable machine learning. Lulu. com.

Muñoz-Izquierdo, N., Laitinen, E. K., Camacho-Miñano, M., & Pascual-Ezama, D. (2020). Does audit report information improve financial distress prediction over altman's traditional Z-Score model? *Journal of International Financial Management & Accounting*, 31(1), 65-97. https://doi.org/10.1111/jifm.12110

Ruiz-Barbadillo, E., Gómez-Aguilar, N., De Fuentes-Barberá, C., & García-Benau, M. A. (2004). Audit quality and the going-concern decision-making process: Spanish evidence. *European Accounting Review*, 13(4), 597-620. https://doi.org/10.1080/0963818042000216820

Putrayasa, I. M. A., Saputra, M. D., & Dewi, N. W. K. (2022, March). Risk Assessment With Audit Tool and Linked Archive System (ATLAS) Methods on Going-Concern During The Covid-19 Pandemic. In International Conference on Applied Science and Technology on Social Science 2021 (iCAST-SS 2021) (pp. 507-510). *Atlantis Press*. https://doi.org/10.2991/assehr.k.220301.083

Salimi, A. Y. (2015). Validity of Altman's z-score model in predicting bankruptcy in recent years. *Academy of Accounting and Financial Studies Journal*, 19(2), 233.

Serrano-Cinca, C. (1996). Self-organizing neural networks for financial diagnosis. Decision Support Systems, 17(3), 227-238. https://doi.org/10.1016/0167-9236(95)00033-X

Steer, T. (2018). In Profile Books (Ed.), The signs were there: The clues for investors that a company is heading for a fall Profile Books; Main edition (November 29, 2018). doi:ASIN : B07CMDKVJ3

Sun, T. (2019). Applying deep learning to audit procedures: An illustrative framework. *Accounting Horizons*, 33(3), 89-109. https://doi.org/10.2308/acch-52455

Svanberg, J., & Öhman, P. (2016). Does ethical culture in audit firms support auditor objectivity? Accounting in Europe, 13(1), 65-79. https://doi.org/10.1080/17449480.2016.1164324

Tagesson, T., & Öhman, P. (2015). To be or not to be–auditors' ability to signal going concern problems. *Journal of Accounting & Organizational Change*. https://doi.org/10.1108/JAOC-04-2013-0034

Thernau, T., Atkinson, B., & Ripley, B. (2013). Rpart: Recursive Partitioning. R Package 4.1-0. Available at: http://CRAN.R-project.org/package=rpart

Virág, M., and Nyitrai, T. (2014). Is there a tradeoff between the predictive power and the interpretability of bankruptcy models? The case of the first Hungarian bankruptcy prediction model. *Acta Oeconomica,* 64(4), 419-440. https://doi.org/10.1556/aoecon.64.2014.4.2

Wang, J. (2021). Management Going Concern Disclosure, Mitigation Plan, and Failure Prediction-Implications from ASU 2014-15Management Going Concern Disclosure, Mitigation Plan, and Failure Prediction-Implications from ASU 2014-15. *The Accounting Review.* https://doi.org/10.2308/TAR-2019-0432

Wilson, J. (2020). How the COVID-19 crisis is leading to lasting change. *Journal Of Accountancy*, 8(2), 87-89.

Zhang, C. A., Cho, S., & Vasarhelyi, M. (2021). Explainable Artificial Intelligence (XAI) in Auditing: A Framework and Research Needs. Soohyun and Vasarhelyi, Miklos, Explainable Artificial Intelligence (XAI) in Auditing: A Framework and Research Needs (October 1, 2021).

Zhang, C. A., Gu, Y., & Vasarhelyi, M. (2021). Error, Noise, and Bias of Auditors' Going Concern Opinions and the Role of Machine Learning. Yu and Vasarhelyi, Miklos, Error, Noise, and Bias of Auditors' Going Concern Opinions and the Role of Machine Learning (December 2021).

Zhang, C. A., Thomas, C., & Vasarhelyi, M. (2021). Attended Process Automation in Audit: A Framework and A Demonstration Attended Automation in Audit. *Journal of Information Systems.* https://doi.org/10.2308/ISYS-2020-073

Appendix A Scorecard Going Concern Checklist - 2019

1. Z Factor

1.1. Working capital / Assets

1.2. Retained earnings / Assets

1.3. Operating margin / Assets

1.4. Cash flow from operations / Debt

1.5. Revenue / Assets

Total Z Factor Score ->Formula (Working capital / Assets*1,2) + (Retained earnings / Assets* 1,4) + (Operating margin / Assets *3,3) + (Cash flow from operations / Debt*0,6) + Revenue / Assets

I) Points: If Z Factor Score equals 0, then 10 points

II) Points: If Z Factor Score is greater than 0 to 1.82, then 5 points

III) Points: If Z Factor Score is greater than 1.81 to 3, then 0 points

2. Distress signs

2.1. Intangibles / Accounts receivable + Cash

2.2. Cash flow from operations / Operating margin

2.3. Goodwill / Assets

2.4. Relevant acquisition in the last 2 years score: (Yes or No)

Distress Signs Score-> Formula (Intangibles / Accounts receivable + Cash) + (Cash flow from operations / Operating margin) + (Goodwill / Assets) + (Relevant acquisition y the last 2 years score)

I) Points: If the Distress Signs Score equals 0, then 0 points

II) Points: If the Distress Signs Score is greater than 0 to 1, then 6 points

III) Points: If the Distress Signs Score is greater than 1 to 1.5, then 7 points

IV) Points: If the Distress Signs Score is greater than 1.5 to 2, then 8 points

V) Points: If the Distress Signs Score is greater than 2 to 2.5, then 9 points

VI) Points: If the Distress Signs Score is greater than 2.5 to 2.99, then 10 points

3. Sector Score: Score based on entity sector

4. Partner score: Score based on the partner's knowledge of the entity, more points worst situation (0-20)

5. Financial Support score: Qualitative based upon Partners' assessment - more points more Risk of not receiving Financial Support (0-20)

6. Solvency risk score: Qualitative based upon Partners' assessment Good = 0 Points, Enough = 5 Points, Poor= 10 Points

7. Equity structure score: Shareholders with problems or low solvency = 10 points Shareholders without problems = 5 Points Solvent unipersonal Company = 0 Points

8. Total Points: Final Score of the checklist.

Total Points-> Formula Partner score + Sector score + Financial support score + Solvency risk score + Equity structure score + Z Factor + Distress Signs

I) Points: If Total Points are smaller than 30, then "Partners' decision."

II) Points: If Total Points are equal to or greater than 30 and smaller than 60, "Review with the second partner."

III) Points: If Total Points are equal to or greater than 60 and smaller than 80, "Review with Expert Network."

IV) Points: If Total Points are equal to or greater than 80, then "Review with the technical team."

Appendix B Variables Definitions

Category	Variables included in Modeling (R Script)	Definitions
Going concern opinion	GCO	Whether the auditor's report had a going concern or not
Partners risk assessment (Qualitative)	Partner score	Engagement partner evaluation about going concern situation (see Appendix A)
	Financial support score	Engagement partner evaluation about financial support from group or entity owner (see Appendix A)
	Sector score	Score based on entity sector (see Appendix A)
	Equity structure score	Score based on entity structure
Financial results (Quantitative)	Z Factor (calculated)	The Z-score is a linear combination of four or five common business ratios, weighted by coefficients. The coefficients were estimated by identifying a set of firms that had declared bankruptcy and then collecting a matched sample of firms that had survived, with matching by industry and approximate size (assets) (see Appendix A)
	Distress Signs (calculated)	Score based on the result of a calculation (see Appendix A)
	Total points (Scorecard GCO Checklist)	Calculation based on other columns (see Appendix A)
	Operating margin / Assets	
	Cash flow from operations / Debt	
	Revenue /Assets	

	Intangibles / Accounts receivable + Cash	
	Goodwill / Assets	
	Relevant acquisition in the last 2 years score	
	Working capital / Assets	
	Retained earnings / Assets	
	Debt / Assets (1)	
	Intangibles/ Assets (1)	
	Accounts Receivables (AR) /Assets	
	Cash / Assets (1)	
	Operating Margin / Revenue (1)	
	Cash flow from operations / Revenue (1)	

(1) calculated from the data included in Scorecard Going Concern Checklist – Appendix A

Appendix C R Script

```
library(readxl)

library(ggplot2)

library(caret)

library(e1071)

library(rpart)

library(rpart.plot)

library(pROC)

#IMPOTAMOS LOS DATOS

Entidades <- read_excel("EntidadesGC_Limpio_Completo.xlsx")


#Show data

View(Entidades)

summary(Entidades)


#How many cases are there of each type

table(Entidades$`¿Incluye GC?`)


#factor the variables

Entidades$`¿Incluye GC?`<- factor(Entidades$`¿Incluye GC?`)


#fix the seed

set.seed(150)


#split the population into training and testing

trainIndex<- createDataPartition(Entidades$`¿Incluye GC?`,

                        p= 0.8,

                        list = FALSE,

                        times = 1)
```

```
fTR<-Entidades[trainIndex,]

fTS<-Entidades[-trainIndex,]


#Since the population of Going concern = 'Si' is too small, we simulate data to be
representative,

#to be the same number as Going concern = 'No'

set.seed(100)

fTR_UP <- upSample(  x = Entidades[,-which(colnames(Entidades)=='¿Incluye GC?')],   #Input
variables

              y = Entidades$`¿Incluye GC?`)     #Output variable


fTR_UP$`¿Incluye GC?` <- fTR_UP$Class

fTR_UP$Class <- NULL

table(fTR_UP$`¿Incluye GC?`)


fTR <- fTR_UP

fTR_eval <- fTR

fTS_eval <- fTS


#Configuration of CROSS Validation, Make 10 partitions


ctrl <- trainControl(method = "cv",

              number = 10,

              summaryFunction = twoClassSummary,

              classProbs = TRUE)


#We standardize the names of the variables

colnames(fTR)<- make.names(colnames(fTR))

colnames(fTS)<- make.names(colnames(fTS))
```

```
#To determine the degree of complexity of the tree (cp) we make an estimate through a sequence

set.seed(150)

tree.fit = train(form = `X.Incluye.GC.` ~  .,

             data = fTR,

             method = "rpart",

             control = rpart.control(maxcompete = 0),

             parms = list(split = "gini"),   # impuriry measure


             #tuneGrid = data.frame(cp = 0.02),

             tuneGrid = data.frame(cp = seq(0,0.02,0.0005)),

             trControl = ctrl,

             metric = "ROC")   # <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<


tree.fit


plot(tree.fit)



#We make the tree with CP = 0,010 the best without losing efficiency


tree.fit = train(form = `X.Incluye.GC.` ~  .,

             data = fTR,

             method = "rpart",

             control = rpart.control(maxcompete = 0),

             parms = list(split = "gini"),

             tuneGrid = data.frame(cp = 0.015),

            # tuneGrid = data.frame(cp = seq(0,0.02,0.0005)),

             trControl = ctrl,

             metric = "ROC")
```

```
tree.fit

#Table with all possible combinations
tree.fit$finalModel
rules<-rpart.rules(tree.fit$finalModel)

#Simple Image of the tree
plot(tree.fit$finalModel, uniform = TRUE, margin = 0.1)
text(tree.fit$finalModel, use.n = TRUE, all = TRUE, cex = .8,xpd = TRUE)

#image of the tree with visual characteristics
rpart.plot(tree.fit$finalModel, type =1, fallen.leaves = FALSE,
        extra=1, tweak=1.2, box.palette = "Oranges", gap=0, space=0)

prp(tree.fit$finalModel,type = 2, extra=102,box.palette = "Greens")
plot(varImp(tree.fit,scale = FALSE))


#Predictions

#Trainning
fTR_eval$tree_prob <- predict(tree.fit, type="prob", newdata = fTR) # predict probabilities
fTR_eval$tree_class <- predict(tree.fit, type="raw", newdata = fTR) # predict classes

#TEST
fTS_eval$tree_prob <- predict(tree.fit, type="prob", newdata = fTS) # predict probabilities
fTS_eval$tree_class <- predict(tree.fit, type="raw", newdata = fTS) # predict classes
```

```
#confusion matrix Training data

confusionMatrix(data = fTR_eval$tree_class, #Predicted classes

            reference = fTR_eval$`¿Incluye GC?`,# observations

            positive = "YES") #Class labeled as Positive


#confusion matrix test data

confusionMatrix(fTS_eval$tree_class,

            fTS_eval$`¿Incluye GC?`,

            positive = "YES")


#Distribution Histogram

ggplot(fTS_eval)+geom_histogram(aes(x=tree_prob$YES,fill=`¿Incluye GC?`),bins = 10)+facet_grid(.~`¿Incluye GC?`)



#Evaluation of the model on the entire data set

fdata_eval <- Entidades

colnames(fdata_eval) <- make.names(colnames(fdata_eval))

fdata_eval$`¿Incluye GC?` <- fdata_eval$X.Incluye.GC.

fdata_eval$tree_prob <- predict(tree.fit, type="prob", newdata = fdata_eval) # predict probabilities

fdata_eval$tree_class <- predict(tree.fit, type="raw", newdata = fdata_eval) # predict classes


confusionMatrix(data = fdata_eval$tree_class, #Predicted classes

            reference = fdata_eval$`¿Incluye GC?`,# observations

            positive = "YES") #Class labeled as Positive


ggplot(fdata_eval)+geom_histogram(aes(x=tree_prob$YES,fill=`¿Incluye GC?`),bins = 10)+facet_grid(.~`¿Incluye GC?`)


###Prediction list Going Concern= YES

ListaYES <- fdata_eval[fdata_eval$tree_class=='YES',]
```

```
ListaYES$ID <- rownames(ListaYES)
#Excel with prediccion Going Concern= YES


#Prediction list Going Concern= No
ListaNO <- fdata_eval[fdata_eval$tree_class=='NO',]
ListaNO$ID <- rownames(ListaNO)
write.table(ListaNO,'ListaNO.csv',sep = ';',row.names = F)
```

Chapter II – Using Big Data Techniques to Analyse Key Audit Matters in the Auditors' Report

1. INTRODUCTION

Key audit matters (KAMs) are "*matters that, in the auditor's professional judgment, were of most significance in the audit of the financial statements of the current period* (IAASB 2015a)." Auditors need to determine, communicate, and disclose KAM item(s) for audits of clients' financial statements. The KAM communication is designed to provide users of auditor's reports with incrementally more useful information than the prior standardized auditor's reports. Moreover, auditors are expected to identify and disclose KAM(s) reflecting the specific circumstance to the client. Whether auditors communicate distinct KAMs for each individual client becomes an important issue that should be addressed (Cordoş & Fülöp 2015; Moroney et al., 2021). Given the importance of such disclosure in auditors' reports, it is critical for KAMs to serve as a beneficial mechanism for enhancing financial reporting quality (Gold et al., 2020). There should be information embedded in its contents and not "boilerplates" which are expected but provide less relevant information to the users of the reporting (IAASB 2015b; Pelzer 2016; PwC 2015; Segal 2017; Velte & Issa 2019). However, to our best knowledge, there are few academic papers investigating and documenting the boilerplate issue of KAM disclosures. Therefore, this study aims to provide the empirical evidence in the textual similarity of KAM disclosures by analyzing the difference of KAM disclosures in terms of word usage from one fiscal year (FY) to the preceding one.

We collect KAM items in auditor's reports from available Spanish companies in FYs 2017 and 2018, then classify those KAM items based on the following three combinations: (1) KAM topic, (2) KAM topic and auditor, and (3) KAM topic, auditor, and industry. Textual analyses are performed based on these three dimensions, respectively. Consistent with other accounting research (Bozanic & Thevenot 2015; Brown & Tucker 2011; Hoberg & Phillips 2016; Lang & Stice-Lawrence 2015; Loughran & McDonald 2016; Peterson et al., 2015), we utilize cosine similarity (CS) to measure the textual similarity between two KAM items. Words in a KAM item are transformed into a vector representing the frequency of each word appearing in the item, and the CS value is calculated between any two vectors (two items). The values of CS vary between zero and one; the higher the CS values, the more similar the two KAM items are.

The empirical results reveal that, generally, KAM items are more similar under the combinations of (1) KAM topic and auditor, therefore, suggesting a recurring textual similarity for the same client; and (2) KAM topic, auditor, and industry in FY 2018 than in FY 2017, therefore suggesting a recurring textual similarity from one year to another. Furthermore, we identify that auditors from one accounting firm had highly similar word usage in those 8 (12) KAM items for the topic provisions for litigation and procedures for their client in the financial services – banking (financial services – capital markets) industry in FY 2017 (2018). Such results might suggest a strong level of textual similarity, delivering a more standardized wording and, consequently, reducing the potential benefit from auditor's reports.

This research contributes to the accounting literature by offering empirical evidence of the textual similarity of KAM disclosures. This evidence is beneficial to researchers, practitioners, regulators, and, especially, standard setters in their review of the auditing standard. Although the study only covers two consecutive fiscal years, further analysis with more FYs can be done to understand if the similarity of KAM disclosures is gradually increased. Moreover, researchers can also examine the association of the KAM disclosure similarity with the market reaction.

This paper is organized as follows. The literature review about KAM research is provided in Section 2. We summarize the data and the methodology utilized in Section 3, and the empirical results are presented in Section 4. Additional analyses are performed in Section 5 and discussions of the results are in Section 6. Lastly, the research is concluded in Section 7.

2. LITERATURE REVIEW

The International Auditing and Assurance Standards Board (IAASB) issued International Standard on Auditing (ISA) 701: *Communicating key audit matters in the independent auditor's report* in 2015. Auditors need to communicate key audit matters (KAMs) in auditor's reports for audits of financial statements whose ending period lies on or after December 15, 2016 (IAASB 2015a). The KAM disclosures are designed to provide users of auditor's reports with more information specific to the client and conveyed directly from the auditor's perspective. Specifically, paragraph A44 states that "*[r]elating a matter directly to the specific circumstances of the entity may also help to minimize the potential that such descriptions become overly standardized and less useful over time* (IAASB 2015a)." Moreover, "*it may be useful for the auditor to highlight aspects specific to the entity (…) in order to make the description more relevant for intended users. This also may be important in describing a key audit matter that recurs over periods* (IAASB 2015a)." Therefore, to empirically and comprehensively compare KAM disclosures among all available auditor's reports and understand their textual similarity would become a critical issue to perceive the specific fact and circumstances in each disclosure. This comparison is the objective of this research.

Many papers investigated the impact of the implementation of expanded auditor's reports in different countries. For instance, Gutierrez et al. (2018) found little evidence that the requirement to issue expanded auditor's reports in the U.K. had effects on market reaction, audit fees, and audit quality. Bédard et al., (2019) also found that the justifications of assessments (JOAs) in expanded auditor's reports[11] had no significant impact on the market reaction, audit fees, audit quality, and audit report lag in France. Hollie (2020) documented early

[11] The justifications of assessments (JOAs) in expanded auditor's reports in France share a similar objective with the KAMs, intending to increase the information content embedded in expanded auditor's reports (Bédard et al., 2019).

evidence for the implementation of critical audit matter (CAM) communication for large accelerated filers in the U.S. All of such studies may suggest the low importance of KAMs for market participants. Such outcomes, perhaps, are from the lack of sufficient precision in the KAM disclosures. However, little prior literature investigated and documented the empirical evidence for the across-year consistency of KAM disclosures by different KAM topics, auditors, and industries. Therefore, this research intends to fill this gap and contribute to the literature by comparing the KAM disclosures in terms of word usage and raises the research question (RQ) as follows:

RQ: What is the textual similarity of KAM disclosures from Spanish public companies classified by KAM topic, auditor, and industry?

3. DATA AND METHODOLOGY

A KAM disclosure, typically, contains three major sections, including (1) the title of the KAM item, (2) the description about the KAM item to elaborate why the auditor would identify this specific issue as a KAM, and (3) the procedures to address this KAM item. We collect the original PDF files of auditor's reports for consolidated financial statements from all Spanish public companies in fiscal years (FYs) 2017 and 2018, convert into DOC files, and manually extract individual KAM item one by one.

We conducted this study in Spain since it is a market that has a unique characteristic as the regulator has required that all auditors' reports (both public interest entities and private owned entities – all the entities) to should include KAM(s). Therefore, it is a market that potentially could have a higher textual similarity as auditors could following a more similar description of risks and audit approach to avoid inconsistency risk for similar matters among their client bases. However, such hypothesis would have to be tested in a further study of a different market where KAM are not being used for all entities to conclude if the Spanish requirement increase textual similarity or not. We also have collected only public companies' auditors report as they were the most readily available data. Differences of textual similarity between private own companies and public companies could exist but it is an area that further investigation would be required.

Panel A The Codes for KAM Topics*

Topic Code	The Name of the Topic
01	Real estate assets: valuation and impairment
02	Acquisitions and business combinations
03	Capitalization of R&D expenses
04	Impairment of customers
05	Stocks: Other
06	Stocks: Valuation

07	Going Concern
08	Taxes: Other
09	Taxes: Recoverability of deferred assets
10	Taxes: Transfer pricing
11	Property, plant and equipment: Valuation and impairment
12	Derivative financial instruments
13	Intangibles: Assessment and deterioration
14	Listed financial investments: Valuation, existence and possession
15	Financial investments in group companies and associates
16	Unlisted financial investments: Valuation
17	Other provisions
18	Others
19	Presentation of relevant facts
20	Debt provisions
21	Provisions for litigation and procedures
22	Provisions for pensions
23	Provisions for insurance
24	Revenue recognition: Integrity
25	Revenue recognition: Occurrence
26	Revenue recognition: Several
27	Information security and control systems

*The KAM topics are derived from Audit Analytics KAM taxonomy (non-public available).

Panel B The Codes for Accounting Firms

Firm Code	The Name of the Firm
01	BDO
02	Deloitte
03	Joint Deloitte-PwC
04	E&Y
05	KPMG
06	Mazars
07	Others
08	PwC

Panel C The Codes for Industries of each Client Being Audited*

Industry Code	The Name of the Industry
01	Business and professional services – business services – industrial
02	Business and professional services – business services – other
03	Consumer business – consumer products
04	Consumer business – food and beverages
05	Consumer business – gamin and betting
06	Consumer business – hospitality and leisure
07	Consumer business – retail
08	Consumer business – travel and aviation
09	Energy and Resources – metals and mining
10	Energy and Resources – oil and gas
11	Energy and Resources – power and utilities
12	Financial services – asset owner/infrastructure funds
13	Financial services – banking
14	Financial services – capital markets
15	Financial services – insurance
16	Financial services – investment management
17	Healthcare and life sciences – healthcare and life science
18	Infrastructure services and real estate – business services – industrial
19	Infrastructure services and real estate – contracting/construction
20	Infrastructure services and real estate – real estate investment trust
21	Manufacturing – manufacturing – other industrial
22	Manufacturing – manufacturing – process and packaging
23	TMT – media
24	TMT – technology
25	TMT – telecommunications

*The codes for industries are derived based on the company industry classification in Spain for tax returns.

Table 1. The Codes for Key Audit Matter (KAM) Item Classifications

Considering this and the public companies auditors' report, the number of available KAM items is 358 (545) in FY 2017 (2018). These KAM items are classified according to the following three combinations: (1) KAM topic, (2) KAM topic and auditor, and (3) KAM topic, auditor, and the client's industry. The textual analyses of KAM items based on the above three combinations will be helpful to obtain more insights on their disclosure similarity. The codes and names for KAM topics, auditors, and industries are disclosed in Panel A, B, and C of Table 1, respectively. The KAM topics (Panel A of Table 1) are derived from Audit Analytics KAM taxonomy (non-public available) and the industry classification (Panel C of Table 1) is based on the company industry classification in Spain for tax returns. The number of possible observations among three combinations to classify KAM items in FYs 2017 and 2018 is disclosed in Table 2.

Combinations	**Fiscal Year 2017**	**Fiscal Year 2018**
(1) KAM topic	26	27
(2) KAM topic and auditor	100	102
(3) KAM topic, auditor, and industry	327	400

Table 2. The Number of Possible Observations among Three Combinations to classify KAM items in Fiscal Years 2017 and 2018

Cosine similarity (CS) has been used to measure the resemblance of documents in the accounting literature (Bozanic & Thevenot 2015; Brown & Tucker 2011; Hoberg & Phillips 2016; Lang & Stice-Lawrence 2015; Loughran & McDonald 2016; Peterson et al., 2015). Moreover, comparing to other textual documents, word usage in KAM disclosures is not dynamically changed. Therefore, we continue using the CS to measure the textual similarity of KAM items in this research.

It is necessary to preprocess textual documents before calculating CS values. Specifically, we perform the following preprocessing steps, including (1) tokenization, (2) removal of non-alphabets, (3) removal of stop words, (4) stemming, (5) removal of punctuations, and (6) conversion of all lower cases. Furthermore, we follow Loughran and McDonald (2011) and implement the term frequency-inverse document frequency (TF-IDF) weighting function to incorporate the weighting scheme during the calculation of the CS values. All of the above steps are performed in Python environment[12].

To obtain CS between two KAM items (K_1 and K_2), we split these KAM items into two vectors (X and Y), representing the frequency of N words appearing in each KAM item. The CS value between the two KAM items is calculated as (Loughran & McDonald 2016):

$$\text{cosine similarity } (K_1, K_2) = \frac{\sum_i^N X_i Y_i}{\sqrt{\sum_i^N X_i^2}\sqrt{\sum_i^N Y_i^2}} \qquad (1)$$

The CS values range from 0 to 1 because the frequency of words is always positive[13]. Higher CS value indicates the two KAM items are more similar; in the extreme case, the two KAM items are exactly the same in terms of the word usage if their CS value equals 1. The CS values, however, do not indicate the percentage of textual similarity between two KAMs. For instance, it does not imply that two KAM items are 95% similar in terms of word usage if their CS value is 0.9500.

[12] The Python codes used in the analysis are available by contacting the corresponding author.
[13] Typically, the cosine values of any two vectors are ranging from -1 to 1 when negative numbers are possible to be appeared in vectors. In this research, numbers in vectors represent the frequency of words used in KAM items and are always to be positive; therefore, the CS values are ranged from 0 to 1.

4. EMPIRICAL RESULTS

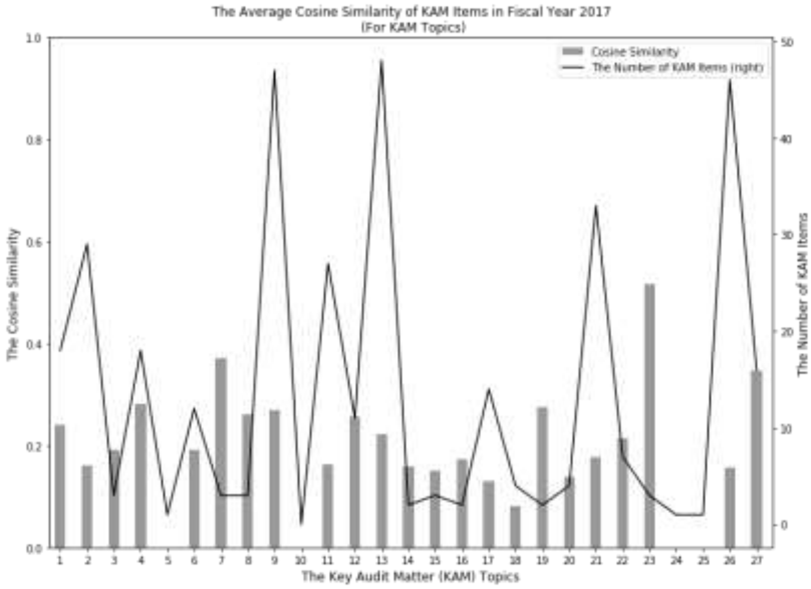
We present the *average* CS values for each KAM topic in FYs 2017 and 2018 in Panel A and B of Figure 1, respectively. For example, the KAM Topic 3 (Capitalization of R&D expenses) has three KAM items in FY 2017 – KAM 3_1, 3_2, and 3_3. Because the CS only measures the textual resemblance between two documents, we further calculate and present the *average* CS. The average CS of KAM Topic 3 in FY 2017, *0.1913*, is the average of 0.3011 (between 3_1 and 3_2), 0.1122 (between 3_1 and 3_3), and 0.1606 (between 3_2 and 3_3).

The three KAM topics with highest average CS are Topic 23 (CS = 0.5165 with 3 KAM items), 7 (CS = 0.3707 with 3 KAM items), and 27 (CS = 0.3471 with 16 KAM items) in FY 2017, and Topic 4 (CS = 0.3805 with 9 KAM items), 23 (CS = 0.3618 with 5 KAM items), and 20 (CS = 0.3305 with 24 KAM items) in FY 2018. KAM Topic 10 has no CS value is because there is no KAM item for the topic in FY 2017, and KAM Topic 5, 24, and 25 (14, 16, and 19) have only one KAM item that classified for each topic in FY 2017 (2018).

Panel A and B of Figure 2 present the top 20 average CS of combinations of KAM topic and auditor in FYs 2017 and 2018, respectively. The combination with the highest average CS in FY 2017 is 26_07[14] (CS = 0.9335 with 2 KAM items), meaning that the two KAMs related to *revenue recognition: several* (Topic 26) issued by the Accounting Firm 07 are highly similar. In FY 2018, 03_08, the code for KAM items related to *capitalization of R&D expenses* (Topic 03) issued by the Accounting Firm 08 has an average CS value equal to 1, indicating that the two KAM items in this combination are exactly the same.

[14] The codes for combinations of KAM topics and accounting firms have two parts with the order and separated by one underline. For instance, 03_02 represents the KAM Topic 03 (Capitalization of R&D expenses) issued by Accounting Firm 02.

Panel A Fiscal Year 2017



Panel B Fiscal Year 2018

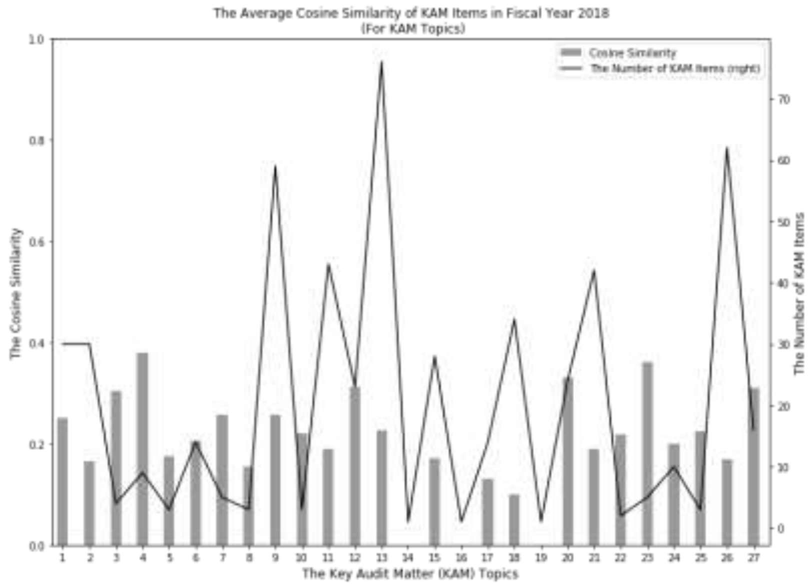
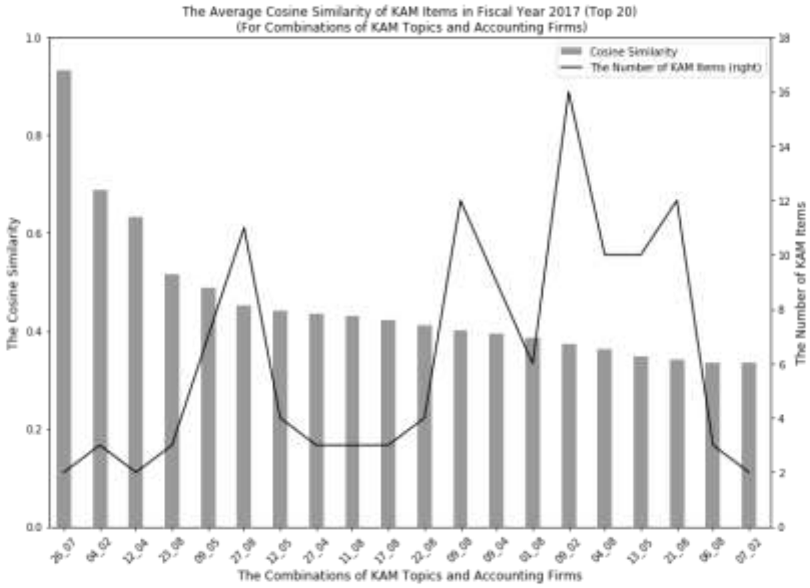


Figure 1. The Average Cosine Similarity of KAM Items for Each KAM Topics

Panel A Top 20 in Fiscal Year 2017



Panel B Top 20 in Fiscal Year 2018

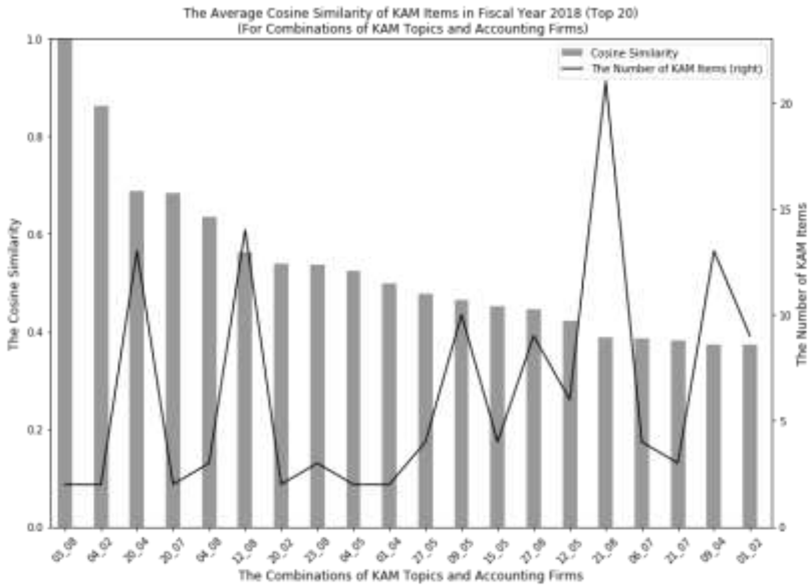


Figure 2. The Average Cosine Similarity of KAM Items for Each Combination of KAM Topics and Accounting Firms

One more interesting finding is that the combination of 04_02 has the second-highest average CS in both FYs 2017 and 2018, revealing that KAM items related to *impairment of customers* (Topic 04) issued by Accounting Firm 02 are consistently and highly similar in both FYs in terms of word usage.

We further classify KAM items into different combinations of KAM topic, auditor, and industry to which clients belong and present the average CS among the combinations in FYs 2017 and 2018 in Panel A and B of Figure 3, respectively. The combination 04_02_13[15] (CS = 0.6885 with 3 KAM items) has the highest average CS in FY 2017. It represents that the three KAM items related to *impairment of customers* (Topic 04), issued by Accounting Firm 02, and issued for clients in the industry of financial services – banking (Industry 13) are highly similar to each other.

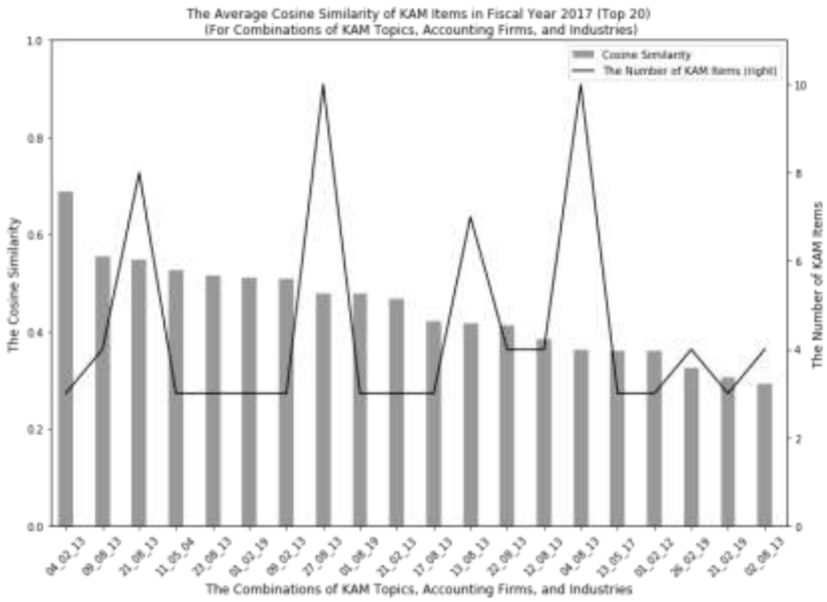
Two important findings are identified after comparing Panel A and B of Figure 3. First, the average CS values in FY 2018, generally, are higher than in FY 2017; even the number of KAM items of combinations with the top 5 highest average CS values are higher in FY 2018. This uncovers the trend that, generally, auditors in the same accounting firm tend to issue similar KAM items related to the same KAM topic for clients within the same industry in FY 2018 than in FY 2017.

Second, the combinations 21_08_13 (CS = 0.5477 with 8 KAM items) and 21_08_14 (CS = 0.9272 with 12 KAM items) occupy with the third (FY 2017) and the first (FY 2018) highest average CS values, respectively (indicated in Figure 3). It reveals that auditors in Accounting Firm 08 had

[15] The codes for combinations of KAM topic, auditor, and industry have three parts with the order and separate by two underlines. For instance, 04_02_13 represents KAM items of KAM Topic 04 (Impairment of customers) issued by Accounting Firm 02 (Deloitte) for clients within Industry 13 (Financial services – banking).

highly similar word usage in those 8 (12) KAM items for the KAM topic 21 *provisions for litigation and procedures* for their client within the Industry 13 financial services – banking (Industry 14 financial services – capital markets) in FY 2017 (2018). Those highly similar KAM items might deliver less incrementally useful information specific to the client for the users of auditor's reports.

Panel A Top 20 in Fiscal Year 2017



Panel B Top 20 in Fiscal Year 2018

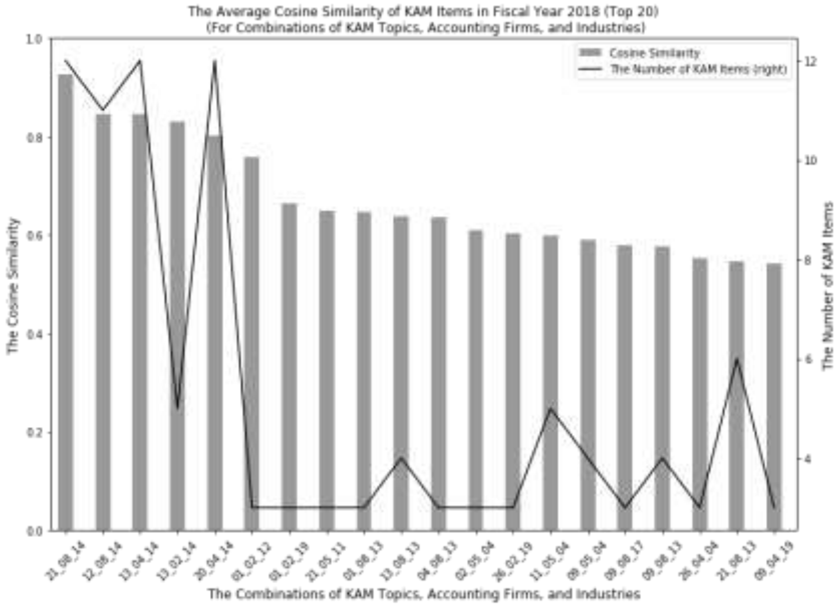


Figure 3. The Average Cosine Similarity of KAM Items for Each combination of KAM Topics, Accounting Firms, and Industries

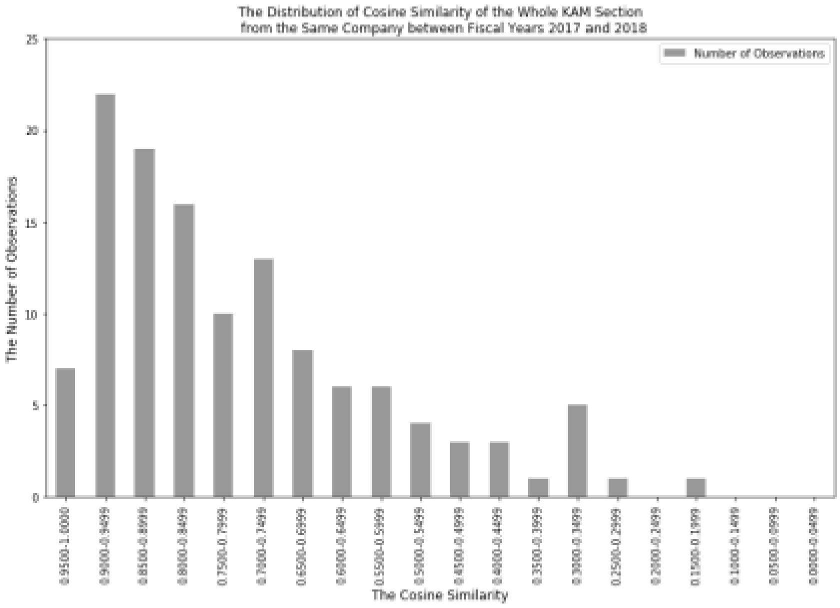
5. ADDITIONAL ANALYSES

We perform two additional analyses other than the three-level comparisons in the previous section. Specifically, we compare the KAM disclosure section in auditor's reports of the same company between FYs 2017 and 2018 and calculate their CS value. This comparison will provide a broad view of the consistency in KAM disclosures from the same company between different years. Furthermore, we calculate the CS for KAM items within the same KAM topic from the same company between FYs 2017 and 2018, narrowing down to understand the KAM disclosure difference in each KAM topic from the same company between different years.

We identify 125 companies existing in both FYs 2017 and 2018 and calculate the CS values of the whole KAM section from the same company between both years. Panel A of Figure 4 summarizes the distribution of the CS values. There are 48 (38.4% = 48/125) companies with CS values higher than 0.8500, indicating their KAM disclosure section between FYs 2017 and 2018 are relatively similar.

To obtain a more granular understanding of the similarity of KAM disclosures, we further identify KAM items within the same KAM topic from the same company between FYs 2017 and 2018 and calculate the CS values. 258 KAM item pairs are identified, and the distribution of the CS values is summarized in Panel B of Figure 4. There are 133 (51.6% = 133/258) KAM item pairs with CS values higher than 0.8500, revealing that auditors tend to use similar words when they discussed the same KAM topic for the same client between both FYs. This might alert the lack of precision in KAM disclosures, therefore, the potential use of boilerplates.

Panel A The Whole KAM Section



Panel B The KAM Items within the Same KAM Topic

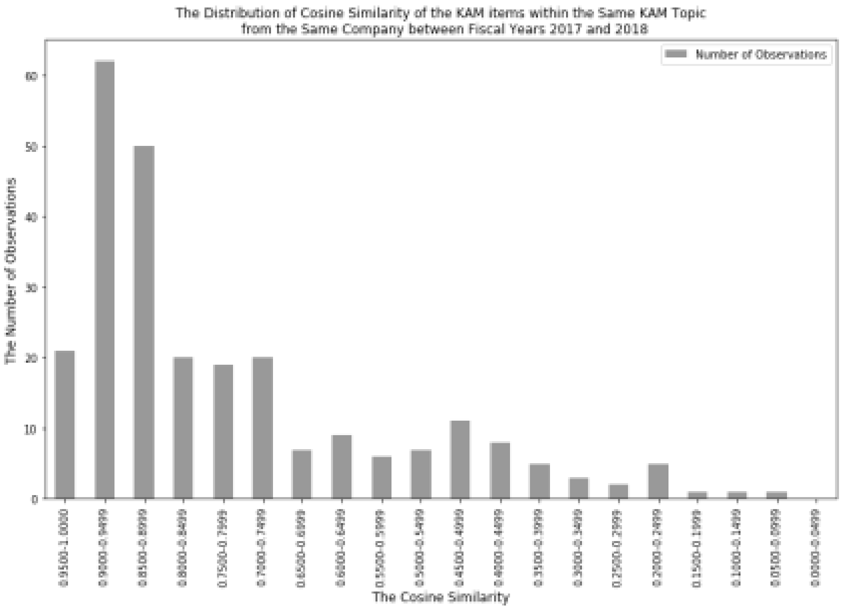


Figure 4. The Distribution of Cosine Similarity from the Same Company between Fiscal Years 2017 and 2018

6. DISCUSSIONS

This research documents the textual similarity of KAM disclosures in terms of word usage measured by CS values by using KAM items in auditor's reports from Spanish companies in FYs 2017 and 2018. We perform the analyses in three different levels (combinations). The empirical results, first, show that auditors from the same accounting firm generally tend to express a more similar KAM item for the same topic for clients within the same industry in FY 2017 than in FY 2018, as presented in Figure 3. This might be a result of accounting standards, auditors' risk assessment, and procedures not significantly changing from year to year. Therefore, they could be complying with the procedures established in paragraph 10 of ISA 701. Considering the evidence found about the KAM similarity, if market participants consider that a more granular or customized KAM is necessary, then specific guidance and examples about how to perform such customization would be helpful.

It is observed in the additional analyses that among 258 KAM item pairs from the same company discussing the same KAM topic between FYs 2017 and 2018, more than half of the pairs have CS values higher than 0.8500. The observation shows that there might be a strong similarity for such KAM items. One possible reason is due to the same lack of change of auditors' procedure(s). The use of highly similar KAM might not provide users of auditor's reports with the precise and detailed information that might be useful to support their decision-making, contrary to the spirit of expanded auditor's reports. Whether users of auditor's reports (financial statements) react to the textual similarity of KAM disclosures could be the next research topic that is worthy of investigation.

This research has limitations and potential additional questions that should be considered in ensuing works. First, researchers can include more KAM items in subsequent FYs to obtain more insights into the textual similarity of KAM disclosures with a longer time-series evolvement.

Second, the impact of signing partner changes within the same accounting firm on the KAM disclosure similarity for the same company can also be investigated. This investigation may be helpful to understand whether the individual signing partner has the ability to draft KAM disclosures even for the same client. Finally, it also could suggest that a broader discussion between regulators, auditors' and market participants to identify what type of precision would be useful in auditors' disclosure and the reasonable balance between such disclosure and confidentiality issues. This would avoid substantial competitive harm to the company being audited and prevent disclosure of information that might not be necessary for investors.

7. CONCLUSIONS

To fill the gap in the literature, we raise the research question intending to understand the textual similarity of KAM disclosures by using KAM items in auditor’s reports of Spanish companies in FYs 2017 and 2018. The CS is used and calculated to measure the KAM textual similarity. We document the empirical evidence that the average CS values in FY 2018, generally, are higher than in FY 2017, indicating the KAM disclosures get more similar. Furthermore, we also observe some accounting firms disclose relatively similar KAM items for specific KAM topics for clients within specific industries.

The observation in textual similarity of KAM disclosures would be helpful for standard setters to understand whether the boilerplate issue empirically exists after the issuance of extended auditor’s reports. From the users’ perspective, we encourage researchers can investigate whether the textual similarity of KAM disclosures impact on the reaction of users of auditor’s reports.

References

Bédard, J., Gonthier-Besacier, N., & Schatt, A. (2019). Consequences of expanded audit reports: Evidence from the justifications of assessments in France. *Auditing: A Journal of Practice & Theory*, 38(3), 23-45. https://doi.org/10.2308/ajpt-52339

Bozanic, Z., & Thevenot, M. (2015). Qualitative disclosure and changes in sell-side financial analysts' information environment. *Contemporary Accounting Research*, 32(4), 1595-1616. https://doi.org/10.1111/1911-3846.12123

Brown, S. V., & Tucker, J. W. (2011). Large-sample evidence on firms' year-over-year MD&A modifications. *Journal of Accounting Research*, 49(2), 309-346. https://doi.org/10.1111/j.1475-679X.2010.00396.x

Cordoş, G., & Fülöp, M. (2015). Understanding audit reporting changes: Introduction of key audit matters. *Journal of Accounting & Management Information Systems*, 14(1), 128-152. https://www.researchgate.net/publication/275032161_Understanding_audit_reporting_changes_introduction_of_Key_Audit_Matters

Gold, A., Heilmann, M., Pott, C., & Rematzki, J. (2020). Do key audit matters impact financial reporting behavior? *International Journal of Auditing*, 24(2), 232-244. https://doi.org/10.1111/ijau.12190

Gutierrez, E., Minutti-Meza, M., Tatum, K. W., & Vulcheva, M. (2018). Consequences of adopting an expanded auditor's report in the United Kingdom. *Review of Accounting Studies*, 23(4), 1543-1587. https://doi.org/10.1007/s11142-018-9464-0

Hoberg, G., & Phillips, G. (2016). Text-based network industries and endogenous product differentiation. *Journal of Political Economy*, 124(5), 1423-1465. http://dx.doi.org/10.1086/688176

Hollie, D. (2020). Early evidence on the AS 3101 critical audit matters disclosure. *Journal of Forensic and Investigative Accounting*, 12(1), 45-54. https://doi.org/10.2139/ssrn.3481403

International Auditing and Assurance Standards Board (IAASB). (2015a). *Communicating key audit matters in the independent auditor's report. International Standard on Auditing 701.* New York: International Federation of Accountants. https://www.ifac.org/system/files/publications/files/ISA-701_2.pdf

International Auditing and Assurance Standards Board (IAASB). (2015b). *Reporting on audited financial statements – New and revised auditor reporting standards and related conforming amendments.* New York: International Federation of Accountants. https://www.iaasb.org/system/files/publications/files/ Basis%20for%20Conclusions%20-%20Auditor%20Reporting%20-%20final.pdf

Lang, M., & Stice-Lawrence, L. (2015). Textual analysis and international financial reporting: Large sample evidence. *Journal of Accounting & Economics*, 60(2), 110-135. https://doi.org/10.1016/j.jacceco.2015.09.002

Loughran, T., & McDonald, B. (2016). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of Finance,* 66(1), 35-65. https://doi.org/10.1111/j.1540-6261.2010.01625.x

Loughran, T., & McDonald, B. (2016). Textual analysis in accounting and finance: A survey. *Journal of Accounting Research*, 54(4), 1187-1230. https://doi.org/ 10.1111/1475-679X.12123

Moroney, R., Phang, S., & Xiao, X. (2021). When do investors value key audit matters? *European Accounting Review*, 30(1), 63-82. https://doi.org/ 10.1080/09638180.2020.1733040

Pelzer, J. R. E. (2016). *Understanding barriers to critical audit matter effectiveness: A qualitative and experimental approach*. [Doctoral dissertation, Florida State University]. FSU's Digital Repository. http://fsu.digital.flvc.org/ islandora/object/fsu%3A360417

Peterson, K., Schmardebeck, R., & Wilks, T. J. (2015). The earnings quality and information processing effects of accounting consistency. *The Accounting Review*, 90(6), 2483-2514. https://doi.org/10.2308/accr-51048

PricewaterhouseCoopers (PwC). (2015). *Delivering the value of the audit New insightful audit reports*. https://www.pwc.com/gx/en/audit-services/publications/ assets/pwc-auditing-report-new-insightful.pdf

Segal, M. (2017). ISA 701: Key audit matter – An exploration of the rationale and possible unintended consequences in a South African. *Journal of Economic and Financial Sciences,* 10(2), 376-391. https://doi.org/10.4102/jef.v10i2.22

Velte, P., & Issa, J. (2019). The impact of key audit matter (KAM) disclosure in audit reports on stakeholders' reactions: A literature review. *Problems and Perspectives in Management,* 17(3), 323-341. http://dx.doi.org/10.21511/ ppm.17(3).2019.26

Conclusions, Limitations, Contributions and Future Work

Summary and conclusions

This thesis was aimed to apply big data techniques and technology in trying to enhance and to assist auditors' procedures and documentation and, also, in analysing the results of the auditors' report. These objectives were considered relevant to provide evidence that certain technology could be used to support the assessment performed by auditors in relation to estimates made by management and, also, to corroborate that the use of technological resources could enhance audit quality and also provide meaningful information auditors' key audit matters (KAM) paragraphs

In relation to the decision tree model and the results of the investigation explained in Chapter I it was demonstrated that such white-box decision tree model technique constructed considering both historic quantitative information and also auditors' knowledge (qualitative information) was able to provide an assistance tool for auditors in their assessment of entity risk of failure and, consequently, such technique assisted in their analysis on whether to include a going concern paragraph in their auditor report as required by the applicable auditing standards. The automated tool was a decision tree (white-box easy to use) that, based on a set of financial indicators, provided additional evidence to auditors in their decision whether the auditors' report should include a going concern (the variable that the model is trying to predict) with a high degree of confidence. The innovation of this method is that it was aimed to prove a two-fold strategy to prepare an intuitive, white-box, and easy-to-use predictive model based upon simple decision tree questions considering both quantitative and qualitative indicators. Qualitative indicators considered the entity's financial figures based upon Z-Score supplemented with other quantitative indicators. The financial quantitative indicators also contributed important information about: (a) auditors' knowledge about the entity's risk considering their experience and expertise; and (b) the auditing firm's risk assessment of the company's industry being audited. This research was very useful in practice because it allows a more robust and, consequently, an enhanced audit evidence in relation to the auditors' assessment and it could be easily extended in other Firms or Cultures. Evidently, each firm should consider their data and elaborate their prediction model since data could vary because of portfolio of companies being audited and risk perception. This paper could also serve as an example for regulators on how machine learning could be applied for better quality audits.

In relation to the results obtained in Chapter II – Using Big Data Techniques to Analyse Key Audit Matters in the Auditors' Report, the conclusions obtained helped to understand the textual similarity of KAM disclosures by using KAM items in the 2017 and 2018 auditor's reports of Spanish companies. The Cosine Similarity (CS) was used and calculated to measure the KAM textual similarity. The paper provided empirical evidence that the average CS values in 2018, generally, are higher than in 2017, indicating the KAM disclosures are increasingly similar. Furthermore, it also observed that some accounting firms disclose relatively similar KAM items for specific KAM topics for clients within specific industries. The observation in textual similarity of KAM disclosures also provided helpful information for standard setters to understand whether the boilerplate issue empirically exists after the extended issuance of extended auditor's reports.

Original Contributions

The most important value of this thesis is that the proposed models were used in real practice, hence bringing the research to applied knowledge. Chapter I resulted in a model that was built, used, and compared with the auditors' report issued by a major audit Firm in Spain. It became evident that the use of big data techniques with data that auditors are able to gather (quantitative) and also auditors' knowledge can provide an easy-to-use model that can assist auditors in their documentation and also can assist in the achievement of a more quality audits. The results also provide empirical evidence about what regulators have mentioned in their study indicating that technology has much to offer the auditor in terms of efficiency and effectiveness as well as in supporting the assessment of the reasonableness of estimates made by management. Consequently, the results obtained provided an answer to the objectives that were set and, also, corroborated that the use of technological resources in real practice clearly could enhance audit quality.

Also, in relation to the Key Audit Matters (KAM), it became also evident that the new auditors' reports provide some very interesting data in their key audit matter (KAM). The use of big data techniques in real practice enabled the evaluation of the consistency of such KAMs following three combinations: (1) KAM topic, (2) KAM topic and auditor, and (3) KAM topic, auditor, and industry of the client being audited. The results provided indication of a surprisingly high consistency and provided empirical practical evidence about how auditors from the same accounting firm were using similar text under each KAM topic, and that such similarity increases for clients within the same industry. The results could be very helpful in any future discussion about the level of customization and customization guidance in the case that KAMs requirement are reviewed in the future.

Limitations

Regarding the Decision Tree Tool for Auditors' Going Concern Assessment in Spain, there were three limitations. First, this work has been performed on a set of Spanish data for one audit firm. As firm cultures vary among firms and countries, the final Decision Tree predictive model may not be entirely applicable for any other firm or country. Each firm should consider its own data and variables, especially subjective variables, and create its own Decision Tree. However, the procedure established in such study could be easily extrapolated and could benefit firms in setting up their predictive model based upon white-box machine learning technology. Second, the model has been built considering the available data for a short period (one year, 2019). Therefore, the model could have been different or even more accurate if more periods were to be included to build the model. Third, this Decision Tree model was not designed to predict real business failure, rather it was built to aid auditor's decision-making processes by formalizing auditors' past judgment history with qualitative and quantitative data. Finally, based upon the results obtained a model recalibration should be done every year as the model could change as quantitative data also change.

In relation to the results obtained in Chapter II – Using Big Data Techniques to Analyse Key Audit Matters in the Auditors' Report, there were natural limitations to this study. In the textual similarity of KAM Disclosures, it was limited to a specific time frame. Therefore, future researchers can include more KAM items in subsequent years to obtain more insights into the textual similarity of KAM disclosures with a longer time-series evolvement. Second, the impact of signing partner changes within the same accounting firm on the KAM disclosure similarity for the same company can also be investigated (increase granularity of the analysis). Finally, geography, as this same study could be replicated in other markets and maybe different results could be obtained.

Future Work

As indicated in Chapter I – Decision Tree Tool for Auditors' Going Concern Assessment in Spain, there were some clear reasons for some False Negative cases (Model indicating "NO" and auditors opt "YES") and also False Positive cases (Model indicating "YES" and auditors opt "NO"). Considering such results, future work could consider an additional investigation about the impact in additional new variables, such as mitigation plans (whether firms have external financing, or the coverage ratio of the liability) could be helpful to increase model efficiency. Also, as the model was constructed with historical financial information and qualitative score about a company or its industry and those may change through time, any future model should be continuously updated with new data each year. Finally, an interesting future research could be to add an additional layer to this work and compare real business failure with the Decision Tree Model results as well as the respective auditors' final conclusion to evaluate how many companies such ultimately entered into a bankrupt situation.

In relation to the results obtained in Chapter II – Using Big Data Techniques to Analyse Key Audit Matters in the Auditors' Report, considering the evidence found about the KAM similarity, if market participants consider that a more granular or customized KAM is necessary, then a more granular segmentation of KAM should be performed in the future. In addition, it would be worth any future work that would analyse the information a greater number of data (and maybe other geographies) and years to evaluate the trend in time. Also, given the results obtained and the level of highly similar KAM it might indicate that auditor's reports don't provide precise and detailed information that might be useful to support their decision-making, contrary to the spirit of expanded auditor's reports. Whether users of auditor's reports (financial statements) react to the textual similarity of KAM disclosures might also be an area for a next research topic that is worthy of investigation.

Finally, as it was seen throughout this thesis there are many areas where Big Data Techniques or Technological tools con assist auditors to document their work and analysis. Some other important areas that a similar investigation could be done is in relation to Decision Trees in the analysis of goodwill or other intangible assets impairment, which is an area that significant judgement (qualitative characteristics) must be applied considering a significant amount of data (quantitative data). Also, in relation to textual analysis techniques, another area of future work could be textual analysis of a company's management report of past press releases to identify paragraphs or situation that an auditor should evaluate in the riesk assessment phase of their work.

****