Machine Learning (2022) 111:4183–4215
https://doi.org/10.1007/s10994-022-06226-4

Machine learning in corporate credit rating assessment using the expanded audit report

Nora Muñoz-Izquierdo[1] · María Jesús Segovia-Vargas[2] · María-del-Mar Camacho-Miñano[3] · Yolanda Pérez-Pérez[3]

Received: 1 March 2021 / Revised: 14 July 2022 / Accepted: 15 July 2022 /
Published online: 30 September 2022

Abstract

We investigate whether key audit matter (KAM) paragraphs disclosed in extended audit reports—paragraphs in which the auditor highlights significant risks and critical judgments of the company—contribute to assess corporate credit ratings. This assessment is a complicated and expensive process to grade the reliability of a company, and it is relevant for many stakeholders, such as issuers, investors, and creditors. Although credit rating evaluations have attracted the interest of many researchers, previous studies have mainly focused only on financial ratios. We are the first to use KAMs for credit rating modelling purposes. Applying four machine learning techniques to answer this real-world problem—C4.5 decision tree, two different rule induction classifiers (PART algorithm and Rough Set) and the logistic regression methodology—, our evidence suggests that by simply identifying the KAM topics disclosed in the report, any decision-maker can assess credit scores with 74% accuracy using the rules provided by the PART algorithm. These rules specifically indicate that KAMs on both external (such as going concern) and internal (such as company debt) aspects may contribute to explaining a company's credit rating. The rule induction classifiers have similar predictive power. Interestingly, if we combine audit data with accounting ratios, the predictive power of our model increases to 84%, outperforming the accuracy in the existing literature.

Keywords Corporate credit rating · Machine learning techniques · Accounting ratios · Expanded audit report · Key audit matters

1 Introduction

For the past few decades, the analysis of corporate credit rating has attracted the attention of many researchers (Golbayani et al., 2020; Lee, 2007; Tsai & Chen, 2010; Zalata et al., 2020). Credit scoring models have been extensively used in the credit industry to assess the

Editors: João Gama, Alípio Jorge, Salvador García.

✉ Nora Muñoz-Izquierdo
nmunoz@cunef.edu

Extended author information available on the last page of the article

granting of credit, so these models are essential for financial institutions to decide whether or not to grant loans to their customers (Tsai & Chen, 2010). Corporate credit ratings are also important determinants of risk premiums and even the marketability of bonds (Huang et al., 2004). Thus, these ratings form the basis of important decisions and therefore they need to be as accurate as possible (Hájek & Michalak, 2013).

The assessment of a company's risk status involves an expensive and difficult process that generally takes months because it involves several experts studying a number of variables that reflect the company's underlying reliability (Hájek & Michalak, 2013). Rating agencies, such as Standard and Poor's (S&P) and Moody's, analyse various aspects of the companies -from strategic competitiveness to operational level details- to arrive at these ratings, investing a significant amount of time and effort (Huang et al., 2004). This paper might provide a solution that can reduce the financial costs and time involved in assessing credit ratings by applying machine learning techniques to audit data publicly available in audit reports, found on the first few pages of any company's annual report.[1]

The literature has demonstrated that machine learning techniques are accurate predictors of corporate credit ratings using financial and accounting ratios (Golbayani et al., 2020; Hájek, 2012; Huang et al., 2004; Pai et al., 2015). However, studies applying machine learning techniques to test corporate credit rating have not used the audit report for this purpose. The audit report is a document issued by independent auditors hired by companies that verifies if the company's financial statements give a true and fair view of the firm's financial position.

In an audit report, after identifying the name of the audited firm, the period covered and the financial reporting framework applied, the auditor issues an opinion, which can be unqualified (pass) or qualified (fail) and summarises the results of the audit process. An unqualified (pass) opinion indicates that the financial statements are free from material misstatements and that the financial statements faithfully represent the financial position of the company. However, a qualified (fail) opinion is issued if the auditor finds evidence of the existence of material misstatements or if the financial statements do not give a true and fair view of the firm's financial position. When an auditor believes that the company may not be able to remain in business in the foreseeable future, anticipating a critical viability concern, a qualified going concern opinion (GCO, hereinafter) is issued. Additionally, the report might also include emphasis of matter paragraphs, referring to any other matters to which the auditor wishes to draw attention by way of emphasis, without qualifying the opinion. Traditionally, the format of an audit report has been concise, giving the pass/fail opinion and, at times, some emphasis of matter paragraphs. However, there has been an international reform of the external audit regulations[2] and the format of the report has been

[1] The annual report is a mandatory document that companies must provide to shareholders on an annual basis that describes the company's operations, performance and financial position. Each annual report contains several sections, such as the financial statements and the audit report, among others. Financial statements include the standard mandatory elements like the balance sheet, profit and loss account and cash flow statement. The preparation of the company's financial statements is the responsibility of the directors of the firm. Published financial statements may be audited by an external and independent certified public accountant. In the case of publicly traded firms, an audit is required by law. When financial statements are audited, the directors include a report prepared by an independent accountant, called the audit report. The audit report is therefore a document prepared by the company's external auditor, addressed to the shareholders and subsequently attached by the company to its annual report.

[2] In 2016, the reform of the international audit regulations began with the updating and modification of some International Standards on Auditing (ISA). The reform took place due to complaints about the informativeness of audit reports. During and after the global financial crisis of 2008, audit report users, such as

expanded for listed firms, becoming more informative and including new sections. In this expanded format, in addition to the opinion and emphasis of matter paragraphs, auditors must include two new types of paragraphs that do not imply a fail opinion: (1) Key Audit Matter paragraphs (KAMs, hereinafter) and (2) Going concern uncertainty paragraphs (GCUPs, hereinafter). KAMs are disclosed in the report if auditors identify significant risks of material misstatements in specific areas of the company during their audit work or if they need to disclose areas in the financial statements that involved significant judgments by the directors (i.e., questionable accounting estimates made by the directors). GCUPs provide a statement on uncertainties relating to events that raise concerns about the company's viability over the following year.

This paper aims to examine how four different machine learning techniques can be used to assess corporate credit rating using the KAMs disclosed in the expanded audit report (model 1). In this model, we also test the predictive power of the GCUPs and emphasis of matter paragraphs, audit opinion and auditor size. The well-known techniques used are a decision tree, namely C4.5, two rule induction classifiers (the PART algorithm and the Rough Set) and the traditional (statistical) logistic regression analysis. Additionally, using the same techniques, we compare the predictive power of the disclosures in the expanded audit report (model 1) with the power of financial ratios alone (model 2) and with a combination of audit information and financial ratios (model 3). This is performed to complement our analysis and ensure the reliability of our evidence.

The main results of this paper suggest that the KAMs disclosed in expanded audit reports provide quite a robust predictive accuracy in anticipating corporate credit rating when applying different machine learning techniques. The predictive power of the main model (auditing variables model or model 1) tested stood at 74.14%, higher than that found in most of the previous studies in the field (see p. 3 of Golbayani et al., 2020 for accuracy percentages in the existing literature). Consequently, any user of annual reports, such as financial and credit institutions, investors, financial analysts and regulators, can benefit from our study. We provide a free of charge, quick and easy way to assess corporate credit ratings. Simply by looking at the first few pages of the annual report published on company websites, any interested decision-maker can read the KAM section of the audit report and roughly anticipate whether the firm's credit risk is high or low. Interestingly, when financial ratios are combined with audit data, the accuracy increases to 84.04%, which is higher than the percentages found in previous studies in the field of corporate credit rating.

Previous research has found that external audit quality (Zalata et al., 2020), audit effort (Ayres, 2015; Lim & Mali, 2020) and audit opinions (Feldmann & Read, 2013) issued in audit reports might have an effect on corporate credit rating. Our paper contributes to this research area as there are no previous studies of the accuracy of the expanded audit report and the KAM disclosures in predicting corporate credit rating applying data mining techniques. This study is also the first to test a combination of KAM disclosures and accounting ratios for the same purpose, responding to calls made in the existing literature for improvements by mixing the financial (ratios) and non-financial (audit report) variables to anticipate financial distress (Bellovary et al., 2007).

The remainder of the article is organised as follows. Section 2 reviews the relevant literature on credit rating assessment using financial and audit data and existing studies on

Footnote 2 (continued)

investors and creditors, were dissatisfied since the audit reports provided no warning about impending company bankruptcies (Geiger, Raghunandan and Riccardi 2014; Sikka 2009).

credit rating using machine learning techniques. Section 3 includes the data, dependent and independent variables and methodologies used in this study. The main model results from testing credit rating assessment using audit variables appear in Sect. 4 for the four machine learning techniques. To ensure the viability of the main findings, Sect. 4 also contains two additional models in which financial information is used for the same credit risk assessment purpose, but auditing and financial data are combined. Conclusions appear in Sect. 5.

2 Literature review

2.1 Credit rating assessment using financial and audit data

Previous studies have analysed the link between corporate credit rating assessment and the external audit profession using different approaches. In particular, research has focused on the connection between credit ratings and audit effort, as well as credit ratings and audit opinion.

Regarding credit ratings and audit effort, Lim and Mali (2020) suggest that companies with higher credit ratings demand greater audit effort, in terms of hours, than companies with lower credit ratings. The authors suggest that higher rated companies demand more audit effort to reduce information asymmetry and demonstrate the robustness of their financial reporting systems. They also show that firms audited by NonBig4 auditors demand additional audit effort with increasing credit ratings compared to Big4 customers (KPMG, PWC, Deloitte and EY).

Other authors investigate the relationship between audit effort and credit rating in the opposite direction, that is, the impact of audit effort on corporate credit ratings. Ayres (2015) concludes that higher accounting information risk levels, associated with disclosures on questionable fair value accounting estimates, negatively impact credit ratings. Elbannan (2008) confirms that companies disclosing internal control weaknesses are more likely to have lower credit ratings, speculative-grade rating, smaller size, lower profitability, lower cash flows from operating activities, net losses in the current and previous fiscal year, higher-income variability and higher leverage than firms with no such disclosures. This is consistent with the findings of Crabtree and Maher (2012) and Dedman and Kausar (2012) who suggest that external auditing brings benefits to private firms in terms of financial reporting quality and assurance and higher credit ratings. Zalata et al. (2020) show that credit rating is associated with external audit quality when companies are suspected of questionable managerial judgments. Interestingly, companies receive high ratings when they are audited by industry-specialised external auditors and credit rating agencies penalise suspect companies that pay high audit and non-audit fees. Additionally, low audit effort increases the extent to which managers can manipulate earnings (Caramanis & Lennox, 2008; Gandía & Huguet, 2020). In conclusion, it seems that the role of auditors in financial reporting supervision improves credit ratings.

Regarding credit rating and audit opinion, Feldmann and Read (2013) investigated this relationship in financially distressed companies. They concluded that the likelihood of an auditor issuing a GCO is associated with the credit rating issued by S&P prior to the audit report date. Moreover, their evidence suggests that after the issuance of a GCO, the rating tends to be downgraded. In companies with impending bankruptcy, Cha et al. (2016) reach the same conclusion using a credit rating issued by the National Information and Credit Evaluation Agency (NICE) in Korea. Strickett et al. (2021) examine this relationship from a bidirectional perspective. They test whether credit ratings inform auditors in the issuance of GCOs, and whether GCOs impact ratings. They show that the likelihood of issuing a GCO is related to the credit rating issued in the previous month by both S&P and Moody's. Additionally, their results show that in the month after an auditor issued a GC opinion, S&P reacted by downgrading its ratings 68% of the time while Moody's did this only 24% of the time. Funcke (2014) focuses specifically on the relationship between the credit rating and auditor reporting accuracy, showing that the existence of poor ratings and rating downgrades contain incremental information for GCOs. Finally, Chen et al. (2020) support the informativeness of audit opinions, both GCOs and other qualifications on violations of accounting standards or disclosure rules. They provide evidence in China that investors efficiently respond to qualified opinions through fast and unbiased stock price adjustments.

In this study, we extend the existing research on the link between audit and corporate credit ratings by examining the predictive power of KAMs on credit rating assessment. KAMs are a new feature of audits and are paragraphs included in the expanded audit report that disclose the company's risk of material misstatements or areas in the financial statements that involved significant management judgments, identified in the audit work. Theoretically, auditors ensure that the financial information reported is free from significant misstatements, because their role is to verify that the true and fair view of the company is shown in the financial statements (DeAngelo, 1981; Zalata et al., 2020). Therefore, the reliability of the information reported on companies is higher if the financial statements are audited. In particular, as the existing literature has identified, GCO disclosures, in which auditors warn investors about the future viability of the company, are associated with corporate credit ratings (Feldmann & Read, 2013). Given this evidence, we expect that other disclosures, such as KAMs, that also indicate auditor concerns about the significant risks and critical judgments of the company, should similarly contribute to predicting corporate credit ratings. When auditors mention risks found during the auditing process in these KAM paragraphs, this disclosure should affect different contract terms between the firm and its stakeholders, including credit rating agencies. We theorise that the KAMs disclosed will provide an accurate and robust assessment of the corporate credit rating. Thus, the first hypothesis is as follows:

> H1: KAMs disclosed in the expanded audit report are significant in assessing corporate credit ratings.

2.2 Credit rating assessment and machine learning techniques

In the existing literature, different approaches have been used to improve the accuracy of corporate credit rating assessments, such as applying different methodologies (Golbayani et al., 2020). Both traditional methods and more sophisticated machine learning techniques have been used for this purpose in the field of credit rating (Tsai & Chen, 2010).

The traditional methods used to predict corporate credit rating are based on statistical techniques. Previous studies testing the traditional models suggest that the most accurate types are the ordered logistic regression and ordered probit models (Hwang, 2013). These two models have outperformed the others, such as linear regression and multiple discriminant analysis (Hwang et al., 2008, 2010), as logit and probit models consider the order of rating classes (Hájek & Michalak, 2013). However, these traditional methods require specific assumptions to have theoretical validity. Thus, new machine learning techniques have also been used to predict corporate credit ratings. One of the advantages of these new techniques is that they do not require certain assumptions on the distribution of the data.

These techniques differ from the traditional models in that they allow models to learn from the data available (Huang et al., 2004).

In the extensive credit rating literature, several non-traditional techniques have been tested (Tsai & Chen, 2010). The most common data mining techniques are decision trees, artificial neural networks (West, 2000; Huang et al., 2004; Kim, 2005; Khashman, 2010; Pacelli & Azzollini, 2011; Addo et al., 2018; Caridad et al., 2019; Wallis, Kumar & Gepp, 2019), naïve Bayes classification and support vector machines (Hájek & Olej, 2011; Huang et al., 2004; Kim & Ahn, 2012; Lee, 2007; Pai et al., 2015; Wallis et al., 2019). The hybridisation procedure has also been an active research area, combining two or more of these sophisticated machine learning techniques (Tsai & Chen, 2010; Yeh, Li & Hsu, 2012) and applying multiple feature selection strategies (Pai et al., 2015; Tuv et al., 2009; Yu et al., 2008). These studies agree on the fact that the previously mentioned data mining techniques are superior to traditional statistical models because they provide more powerful and accurate credit rating assessments (Wallis et al., 2019).

Existing credit rating prediction studies have used financial ratios to develop credit rating models (Hájek & Michalak, 2013). The most commonly selected indicators for forecasting ratings are profitability, activity, liquidity, leverage and market ratios (Golbayani et al., 2020; Hájek, 2012; Hájek & Olej, 2011; Huang et al., 2004; Hwang et al., 2008; Kim, 2005; Pai et al., 2015). While considerable research has been devoted to credit rating prediction using financial data, little is known about how accurate audit information is as an indicator of corporate credit ratings. Considering that the auditing profession ensures the quality of annual accounts through the issuance of the audit report (Lennox, 1999), it is reasonable to expect that disclosures in such reports can similarly contribute to predicting credit ratings.

Past academic research in credit rating prediction using machine learning techniques has mainly been conducted on the US and Taiwan markets (Golbayani et al., 2020; Huang et al., 2004; Pai et al., 2015; Tsai & Chen, 2010). Fewer studies have explored other markets such as Korea (Kim & Ahn, 2012; Lee, 2007; Shin & Han, 2001), Japan (Yu et al., 2008), Australia (West, 2000), Germany (West, 2000; Kashman, 2010; Zhao et al., 2015), UK (Hájek, 2012; Yu et al., 2008) and Italy (Pacelli & Azzollini, 2011; Campanella, 2014; Moscatelli, Parpaliano, Narizzano & Viggiano, 2020). To our knowledge, this is the first study looking at the Spanish market when assessing corporate credit rating using machine learning techniques. The interest in examining this market is because companies in code law countries like Spain behave differently to companies in common law countries, such as the US and the UK, more frequently examined in these studies (La Porta et al., 2000).

In the related literature on bankruptcy and financial failure prediction applying machine learning techniques (Kumar & Ravi, 2007), audit data has been used for prediction purposes, demonstrating a predictive power like that of financial ratios when assessing firms' financial distress (McKee, 2003; Muñoz-Izquierdo et al., 2019a, 2019b). The efficiency of machine learning techniques has also been shown in explaining audit opinions (Sánchez-Serrano et al., 2020). Gaganis et al. (2007) find a high explanatory power of probabilistic neural networks to identify qualified audit opinions. These studies have tested audit data from traditional (pass/fail) format audit reports. However, no existing studies have focused on the predictive power of the new expanded audit report and its disclosed KAMs to anticipate credit ratings.

In summary, there are no existing studies on how machine learning techniques can assess corporate credit ratings using expanded audit reports and, more specifically, using the KAM paragraphs in which auditors mention possible risks identified in their audit work. We expect machine learning techniques to be accurate tools for credit rating assessment using KAMs, since the existing literature has already suggested their application and usefulness for rating prediction using financial ratios. Accordingly, the second hypothesis is proposed as:

> H2: The application of machine learning techniques to KAMs disclosed in expanded audit reports provides an accurate prediction of corporate credit rating

3 Methodology

3.1 Data and dependent variable

For our study, we used the entire population of 131 listed firms in the Spanish continuous trading market in 2017 with available financial data and expanded audit reports. The continuous trading market in Spain includes all stocks that trade simultaneously on the Madrid, Bilbao, Barcelona and Valencia stock exchanges.[3] We selected 2017 because it was the first year of implementation of the expanded audit reporting regulation after the passing of ISA 700,[4] adapting its implementation in Spain by resolution of the ICAC.[5]

We limit our data to one-year period given the recency of KAM disclosures and due to prior research that finds recurrent KAMs and textual similarity year over year. For instance, as per Kend and Nguyen (2020), around 70% of Australian listed firms have the same KAMs disclosed in both years 2017 and 2018. Consistently, using Thai listed companies, Suttipun (2022) show that, although the word count of KAMs reported fluctuate during the three-year period studied (2016–2018), the volume of matters reported is similar each year. Thus, due to the international adoption of the audit reporting regulation, we expect a similar number of KAMs disclosed in the Spanish market in 2018, so we believe that the first year of implementation of this new regulation best fits the investigation of the association between auditor disclosures and credit risk. Another reason supports our argument. Previous studies have showed support for audit firm and auditee industry effects when disclosing KAMs. For example, using Spanish listed companies during the period 2017–2018, the evidence found by Hsieh et al. (2021) indicate that auditors in the same audit firm tend to have recurring textual similarities under each KAM topic year over year, and these similarities increase for auditees that belong to the same industry. Therefore, in order to avoid this audit firm and auditee industry effects, we restrict our study to one fiscal year. Being this period the first implementation period of the new extended regulation, we avoid the influence of prior year auditors and the tendency to report industry-specific KAMs.

For our univariate analyses, we used the whole population of 131 financial and non-financial listed firms. However, for the application of the machine learning techniques, following standard practice, we removed banks and other financial institutions due to their different regulatory requirements and structural characteristics (Charitou et al., 2007). After excluding financial institutions, the final non-financial population consisted of 116 companies. The population construction is summarised in Table 1, Panel A.

[3] The 35 most liquid Spanish stocks that comprise the IBEX 35 (Spanish Exchange Index 35), which is the benchmark stock market index of the Bolsa de Madrid, are contained in our data.

[4] International Standard on Auditing (ISA) 700 (Revised), Forming an opinion and reporting on financial statements.

[5] ICAC (Spanish Accounting and Auditing Institute) resolution of December 23, 2016.

Table 1 Data and dependent variable construction

Panel A. Data construction	
Initial population: listed companies in the Spanish continuous trading market in 2017	143
(–) Companies with fiscal year ends prior to September (for which the expanded audit regulation does not apply in 2017)	(2)
(–) Companies under restructuring or being dissolved	(6)
(–) Companies with missing financial data	(4)
Final population: listed companies in the Spanish continuous trading market with available financial and audit data	131
(–) Banks and other financial institutions	(15)
Final non-financial population: non-financial listed companies in the Spanish continuous trading market	116

Panel B. Dependent variable construction: credit rating

StarMine credit rating	Assigned credit rating value (dummy variable: *CR*)	Number of firms in the non-financial population
AAA	0	1
AA+	0	4
AA	0	3
AA−	0	3
A+	0	5
A	0	10
A−	0	12
Subtotal of high credit rating firms		38
BBB+	1	13
BBB	1	23
BBB−	1	8
BB+	1	7
BB	1	7
BB−	1	4
B+	1	9
B	1	3
B−	1	1
CCC+	1	1
CCC	1	1
CCC−	1	1
Subtotal of low credit rating firms		78
Total firms in the non-financial population		116

Panel A reports the companies used in the study. Univariate analysis of the paper will be reported using all 131 observations, and machine learning techniques will be applied on the non-financial companies (116 observations). In Panel B, the two categories of the credit rating dependent variable are detailed. All A's credit ratings take the value of 0 (low credit risk), and the value of 1 is given to all B's and C's credit ratings (high credit risk)

Auditing and financial data were compiled for every firm from three different sources:

- Auditing variables: Expanded audit reports (included in annual reports) uploaded to the Spanish Stock Exchange's website[6]
- Financial and control variables: ORBIS database
- Credit ratings: StarMine Combined Credit Risk Model provided by Thomson Reuters

We manually collected auditing data from the expanded audit reports, downloading the annual reports from the firms' websites. From each expanded audit report, we manually extracted the audit firm name, type of audit opinion, existence of emphasis of matter paragraphs disclosed, GCUPs and KAMs. The raw KAM data was processed to categorise the content of these disclosures into a 5-item codification, explained in the following section. For the financial and control data, we took the balance sheet and income statement variables available from the ORBIS database.

For company credit rating, we used the StarMine Combined Credit Risk Model provided by Thomson Reuters. This corporate credit rating model generates a single estimate of the credit risk of public companies by combining three credit risk models: the StarMine Text Mining Credit Risk Model, the StarMine SmartRatios Credit Risk Model and the StarMine Structural Credit Risk Model. According to Thomson Reuters, this rating significantly outperforms the Altman Z-Score, accurately predicting 90.4% of default companies within a 12-month horizon in its bottom quintile of scored companies (Yan, Li & Bonne, 2014). The ratings can range from AAA (highest rating) to CCC- (lowest rating). For the purposes of our analysis, and to facilitate the discussion of our results, the multiple ratings are collapsed into two categories according to the schedule provided in Table 1, Panel B. Thus, we used a dummy credit risk assessment (*CR*) variable that takes the value 0 for all A credit ratings and 1 for all B and C credit ratings. When *CR* equals 0, the firm's credit rating is high, and it represents a low credit risk. If *CR* equals 1, the rating is poor, showing a high credit risk. Table 1, Panel B, also summarises the number of firms with high and low credit risks. Essentially, our final non-financial population for the credit rating analysis consists of 38 high credit risk companies and 78 low credit risk companies, which had available information in the three data sources noted above.

3.2 Independent and control variables

The independent variables in this study are summarised in Table 2.

The first nine variables are extracted from the expanded audit reports and represent the explanatory variables for our main analysis. They include audit size, audit opinion, type of paragraphs included and KAMs disclosed in the report.

AUSIZE presents the size of the auditor as a dummy variable, taking the value 1 if the company is audited by a Big 4 auditor, and 0 when the auditor is non-Big 4. The second independent variable is audit opinion (*AUOP*), which is 1 when the opinion is qualified (fail) and 0 if it is unqualified (pass). The third and fourth variables relate to other paragraphs in an audit report: emphasis of matter paragraphs (*EMP*) and the going concern uncertainty section (*GCUP*). Both are dummy variables and take the value 1 if the auditor issues them in the report, and 0 otherwise.

[6] http://cnmv.es/portal/home.aspx?lang=en.

Table 2 Definition of independent variables

Variable name	Explanation	Variable acronym
Audit firm size	Dummy variable with a value of 1 if the audit firm is a Big 4 (KPMG, PWC, Deloitte or EY), and 0 otherwise	*AUSIZE*
Audit opinion	Dummy variable with a value of 1 if the audit report is qualified (fail), and 0 if it is unqualified (pass)	*AUOP*
Type of paragraph		
Emphasis of matter	Dummy variable with a value of 1 if the audit report has an emphasis of matter paragraph, 0 otherwise	*EMP*
Going concern uncertainty	Dummy variable with a value of 1 if the audit report has a going concern uncertainty paragraph, 0 otherwise	*GCUP*
KAMs: Codification		
Going concern KAMs	Dummy variable with a value of 1 if the audit report has any going concern KAMs, 0 otherwise	*GCKAM*
Revenues KAMs	Dummy variable with a value of 1 if the audit report has any revenue KAMs, 0 otherwise	*REVKAM*
Others KAMs	Dummy variable with a value of 1 if the audit report has any KAMs related to other external issues, 0 otherwise	*OTHERKAM*
Assets KAMs	Dummy variable with a value of 1 if the audit report has any asset KAMs, 0 otherwise	*ASSETKAM*
Liabilities KAMs	Dummy variable with a value of 1 if the audit report has any liability KAMs, 0 otherwise	*LIABKAM*
Control variables	Dummy variable with a value of 1 if the company is experienced (21 years or more) according to Coad et al. (2013), 0 otherwise	*FIRMAGE*
	Dummy variable with a value of 1 if the company is large according to the European Commission definition, 0 otherwise	*FIRMSIZE*
	Categorical variable with a value of 0 if the industry is manufacturing, 1 for construction companies and 2 for commercial and services firms, according to the NACE codes	*INDUSTRY*
Financial variables	Continuous variable that presents the firm's liquidity calculated using the working capital (current assets minus current liabilities) to total assets ratio	*LIQUID*
	Continuous variable that presents the firm's cumulative profitability calculated using the retained earnings to total assets ratio	*CUMPROF*
	Continuous variable that presents the firm's profitability calculated using the earnings before interest and taxes (EBIT) to total assets ratio	*PROFITAB*
	Continuous variable that presents the firm's leverage calculated using the book value of equity to total liabilities ratio	*LEVERAGE*

This table explains independent variables for the main analysis (credit risk assessment using auditing variables) and for the additional analyses (credit risk assessment using only financial variables and a combination of auditing and financial variables). The construction of the three control variables of firms' age, size and industry is also reported

The division of the KAMs disclosed into five classifications is also shown in Table 2. KAM independent variables take the value of 0 if no KAMs on that category are disclosed, and 1 otherwise. The five KAM categories are: (1) KAMs related to going concern aspects (*GCKAM*), (2) KAMs explaining issues regarding performance and revenues recognised (*REVKAM*), (3) KAMs indicating aspects about information systems or the company being involved in business combinations (*OTHERKAM*),[7] (4) KAMs informing about valuation and recognition of company resources (*ASSETKAM*), and (5) KAMs disclosing aspects about company debts (*LIABKAM*).

We additionally include three control variables when applying our machine learning techniques (see Table 2). The first control dummy variable relates to the age of the firm (*FIRMAGE*) and takes the value 1 if the company is experienced (21 years or more) and 0 if the company is considered young or mature (up to 20 years). This taxonomy follows that proposed by Coad et al. (2013). The second control variable is the size of the firm measured in total assets. *FIRMSIZE* is a dummy variable that takes the value 1 if the company is large, and 0 when the company is medium. The categorisation of this variable has been made according to the size definition of the European Commission, considering EUR 43 million in assets as the cut-off point. The last control variable is the company's industry (*INDUSTRY*). For our analyses, industries have been split into three categories, according to the NACE[8] classification: manufacturing (value of 0), construction (value of 1) and commercial and services (value of 2).

For our additional analyses, we add four independent financial variables for the credit rating prediction, also reported in Table 2. The financial variables are the four accounting ratios that comprise Altman's Z''-Score model: working capital to total assets (*LIQUID*), retained earnings to total assets (*CUMPROF*), earnings before interest and taxes to total assets (*PROFITAB*), and book value of equity to total liabilities (*LEVERAGE*). The Z''-Score model is the latest version of the traditional model: the Z-Score. This updated version is used because our companies belong to different industries, whereas the original version of this model is only applied to manufacturing firms (Altman, 1983). Also, we use this measure of the firm's financial position because of its popularity and efficacy according to the literature (Altman, 1983; Altman et al., 2017; Balcaen & Ooghe, 2006; Bellovary et al., 2007). As per Altman et al. (2017), this model is applied worldwide as a main tool for analysing the financial position of businesses, both in academic research and practice.

LIQUID is a liquidity ratio that shows the value of net current assets over total assets, and it is expected to be low in financially troubled companies. *CUMPROF* expresses the cumulative profitability from prior periods as a proportion of total assets, whereas the other profitability ratio or return on assets ratio (*PROFITAB*) captures how productive a firm is in generating earnings during the current period. *LEVERAGE* captures if the value of the firm's equity is lower than total debt to external parties.

3.3 Machine learning techniques and their applications

To carry out our empirical analyses, we have developed four models applying different well-known machine learning techniques: a decision tree, two different rule induction

[7] We consider the category *OTHERKAM* as "other" as these aspects are not as controllable by the firm as the rest of the categories based on comments regarding data provided in the company's financial statements.

[8] The NACE codification is the Statistical Classification of Economic Activities in the European Union.

algorithms and a logistic regression model. The decision tree is C4.5, and the rule induction algorithms are the PART algorithm and the Rough Set. We apply these well-known techniques because the paper aims to test the predictive power of expanded audit reports in the field of credit rating. We do not seek to experiment with new classification methods or modify existing ones.

The combination of several machine learning techniques results in a comprehensive analysis of the problem from several points of view. In addition, this strategic mix of methodologies allows us to select the most relevant variables for further comparison. These algorithms have been chosen since their performance is superior to other techniques, according to the recent literature on classification problems, and they are explanatory methodologies. In fact, all these approaches have also been tested previously in assessing corporate credit rating (Golbayani et al., 2020; Tsai & Chen, 2010).

To evaluate the performance of the different models in corporate credit rating assessment, we use three performance metrics: accuracy (Acc), specificity (Sp) and sensitivity (Se). These measures can be obtained from the confusion matrix, in which the diagonal represents the correctly classified examples and the off-diagonal represents the classification errors.

Accuracy (Acc) is the rate of correctly classified observations. Specificity (Sp) and sensitivity (Se) identify the two possible error types in a binary classification problem: type I errors and type II errors. A type I error is associated with the specificity of the model (Type I error $=1-$ Sp). Sp is a statistical metric that identifies how well a binary classification model identifies negative cases. A higher *Sp* indicates a lower probability of type I misclassifications. In contrast, a type II error is related to the sensitivity (*Se*) of the model (Type II error $=1-Se$). *Se* shows the percentage of positive cases correctly classified in a binary model. A higher *Se* implies a lower probability of type II misclassifications.

In our study, a type I error indicates high credit rating firms incorrectly classified as risky (false positive) and a type II error shows low credit rating firms incorrectly classified as non-risky (false negative). We analyse these misclassifications due to their economic significance and to validate the evidence in our models. For instance, in the business failure line of research, type II misclassifications are considered more costly than type I errors (Gutiérrez et al., 2010; Hernandez Tinoco & Wilson, 2013), since failed companies about to go bankrupt are considered financially healthy. In line with this reasoning, for credit rating classification, a high rating given to risky firms (type II) seems to be more financially significant than other errors (type I).

In particular, for the calculation of these metrics with the C4.5 decision tree and the rule induction algorithms (PART and Rough Set), we apply the cross-validation procedure and, specifically, tenfold cross-validation. Our population is divided into k subsamples (10 in our study), so $k-1$ is used to estimate the model and the remaining ones are the evaluation subsamples. This process is repeated k times, so that each subsample is used once to evaluate the model and $k-1$ times to estimate it. The results are then averaged for the 10 different k-folds.

3.3.1 Decision tree (C4.5)

We use the C4.5 algorithm as the first machine learning technique. The C4.5 algorithm is a widely used decision tree developed by Quinlan (1993). The application of the decision tree contributes to a better interpretation of the results because it shows credit rating processes, that is, the ways or “paths” that predict credit scores (Díaz-Martínez et al., 2009).

After the decision tree is "pruned" (all branches for specific cases are eliminated because they do not assist in predicting results that can be generalised), the C4.5 provides a simple, accurate and robust tree. It shows (in the form of "branches" or rules) the best predictors of corporate credit rating.

Standard learning algorithms for decision trees generate a tree structure by splitting the training data into smaller and smaller subsets in a recursive top-down fashion. This splitting continues until all subsets are pure, or until their purity cannot be increased any further. A subset is pure if it contains instances of only one class. The aim is to achieve this with as few splits as possible so that the resulting decision tree is small and the number of instances in each subset is large (Díaz-Martínez et al., 2011). To this end, various split selection criteria have been designed and at each node, the learning algorithm selects the split that gives the best value for the splitting criterion.

Specifically, to carry out the partitions, C4.5 uses the information gain or gain ratio:

$$Gain\ ratio = \frac{I(X;Y_i)}{H(Y_i)}. \tag{1}$$

In the gain ratio, the numerator is the mutual information between X and Y, that is, the information provided by one of the variables about the other:

$$I(X;Y) = H(X) - H(X/Y) \tag{2}$$

Note that $H(X)$ is the entropy of X, defined in a similar way as the entropy of Y. $H(X|Y)$ is the conditional entropy of X given variable Y, and $H(Y)$ is the entropy of Y, both defined below:

$$H(X/Y) = \sum_{x,y} p(x,y) \log_2 \frac{1}{p(x/y)} \tag{3}$$

$$H(Y) = \sum_{x} p(y) \log_2 \frac{1}{p(y)} \tag{4}$$

Consequently, the gain ratio is based on the entropy of a random variable (a measure of the randomness or uncertainty of the variable) and the mutual information between different variables. Thus, the entropy indicates the uncertainty reduction of one of the variables produced when the value of the other (or others) is known (Reza, 1994; Ziemer & Tranter, 2002).

The C4.5 decision tree incorporates several additional features that turn it into a very powerful and flexible technique. For instance, this technique handles missing values and has a mechanism to prune the tree to avoid overfitting. This allows the C4.5 to provide more accurate results with a new dataset that has not been used to develop the model.

3.3.2 The PART algorithm

The PART algorithm, applied as the second machine learning technique, is a rule induction classifier developed by Frank and Witten (1998) that generates rules by incorporating a modified form of the C4.5 decision tree and eliminating some of the paths found in an initial decision tree structure. This algorithm builds partial decision trees instead of fully explored ones. Once the algorithm finds the partial tree, the tree-building stops and a rule is generated with the leaf that represents the greatest number of situations. In this study, the rules developed classify a firm into the high or low credit rating category based on the auditing data available, and the leaves contain the number of firms classified as one group or the other.

Since the PART algorithm is based on partial decision trees, its main advantage is its simplicity. However, its performance is like that of other machine learning algorithms. Consequently, this explanatory technique was chosen because of the simplicity of its rules, with no loss of accuracy. The results of the PART algorithm are easier to interpret than those of other classifiers as they are expressed in logical if/then statements (Díaz-Martínez et al., 2009). Additionally, the usefulness of this methodology has been corroborated in business studies (Camacho-Miñano et al., 2015), although it has not been applied before for credit rating prediction using auditing data.

3.4 The rough set

Rough Set theory (Pawlak, 1991) assumes that knowledge or information can be associated with every object of the universe considered. This information is expressed in the form of some attributes used for object description. In our study, the attributes of this decision rule model are represented by the independent variables (variables extracted from the expanded audit report), and the problem involves classifying high or low corporate credit ratings.

Knowledge is seen as the ability to classify objects, so it consists of a family of classification patterns of a domain of interest. If some objects are described by the same data, they are indiscernible. This indiscernibility relationship leads to the mathematical basis for the Rough Set theory: vague information causes indiscernibility and prevents the precise assignment of objects (firms) to a set (high or low credit rating). Intuitively, a rough set is a set of objects that cannot be exactly expressed by employing available knowledge or a set of attributes (Sanchis et al., 2007).

In addition, the Rough Set approach works by discovering dependencies between attributes in an information table and reducing the set of attributes by removing those that are not essential to characterise knowledge. A reduct is defined as the minimum subset of attributes that provides the same classification quality as the full set of attributes. A reduced information table may provide decision rules of the form 'if conditions then decisions'. These rules specify what decisions (actions) should be undertaken when some conditions are satisfied. The rules can be used to assign new objects to a decision class by matching the condition part of one rule to the description of the object (Pawlak & Skowron, 2007).

Even though this theory has been extended (Greco et al., 1998, 2001), we apply the classical approach. In this approach, attribute domains are not ordered (different values of the same attribute are equally preferable) and the predictive value of the attributes will only be factored into the model. We follow the classical approach because we are interested in the predictive power of the auditing variables (attributes) in assessing the classification problem of high versus low credit rating.

3.4.1 Logistic regression analysis

Logistic regression has become a widely used statistical technique and accepted method of analysing binary classification problems due to its flexibility. This traditional analysis has frequently been used in business, specifically determining corporate credit ratings.

Table 3 Summary statistics by industry (total financial and non-financial firms)

Data	Manufacturing	Construction	Commercial and services	Total
Number of observations	56	26	49	131
Financial data				
Age	54.8	51.2	45.7	50.6
Size	8,285	6,194	85,231	36,651
Current assets	1,968	2,826	2,695	2,385
Current liabilities	1,715	2,179	2,951	2,207
Total liabilities	4,972	4,454	6,919	5,476
Retained earnings	2,960	1,534	1,574	2,212
EBIT	427	177	595	426
Liquidity	0.09	0.17	0.06	0.10
Profitability	0.31	0.24	0.22	0.27
Leverage	0.07	0.02	0.09	0.06
Auditors				
Big 4 auditors	51	21	47	119
Non Big 4 auditors	5	5	2	12
Audit opinion data				
Unqualified opinions	55	24	49	128
Qualified opinions	1	2	0	3
Emphasis paragraphs	0	3	1	4
GC uncertainty paragraphs	3	5	4	12

In this table, descriptive statistics are provided by industry. The three industry categories are created based on the NACE codes. Age is expressed in years (averaged), size is the average of total assets, and all other financial variables (current assets, current liabilities, total liabilities, retained earnings, Earnings before interest and taxes or EBIT) are averaged and expressed in thousands of US dollars. Liquidity ratio is working capital (current assets minus current liabilities) to total assets. Profitability ratio is retained earnings to total assets. Leverage ratio is EBIT to total assets. All ratios are averaged. Regarding auditing data, this table discloses the companies' auditor (Big 4 versus non-Big 4 auditors) and opinion type (unqualified versus qualified), and presents the number of audit reports that include emphasis and GC uncertainty paragraphs

In this study, a logistic regression model is developed to test the predictive power of the expanded audit report in grouping credit ratings and identify which auditing variables significantly assist in achieving this classification. Like the previous techniques, credit rating is the dependent dichotomous variable in the regression and the independent variables consist of the auditing data extracted from the expanded audit report.

Generally, decision trees and rule induction classifiers outperform this traditional alternative in terms of accuracy because this parametric model may be hindered by missing variables and restrictive assumptions about the underlying distribution of the data. Therefore, we expect a slight decrease in the accuracy of our model using this methodology.

Table 4 KAMs by industry (total financial and non-financial firms)

KAMs data	Manufacturing	Construction	Commercial and services	Total
Going concern KAMs	4	0	2	6 (1.7%)
Business Combinations KAMs	13	2	13	28
Information systems KAMs	3	0	7	10
Others KAMs	16	2	20	38 (10.2%)
Fixed assets and investment property KAMs	18	16	9	43
Intangible assets KAMs	9	0	4	13
Goodwill KAMs	19	4	26	49
Inventory KAMs	6	5	2	13
Deferred tax assets & other tax KAMs	24	12	17	53
Financial investments KAMs	10	7	17	34
Assets KAMs	86	44	75	205 (55.2%)
Financial liabilities KAMs	9	2	4	15
Provisions and contingent liabilities KAMs	11	9	25	45
Liabilities KAMs	20	11	29	60 (16.2%)
Revenues KAMs	26	13	23	62 (16.7%)
Total KAMs	152	70	149	371

This table presents the number of KAMs disclosed in the 131 audit reports. Number of KAMs are provided by industry and by the 5-item KAM categories. The specific KAMs included in each category are also reported. 371 KAMs appear in the companies analised, and the most common category include asset valuation disclosures in the manufacturing industry

4 Results

4.1 Summary statistics

Summary statistics for our companies are provided in Tables 3 and 4. As seen, our data consists of 131 firms from different industries: manufacturing (42.0%), construction (19.8%) and commercial (38.2%). The average age is 51 years and the average size, measured by total assets, is USD 36 million.

On average, our firms show a healthy and profitable financial position as listed firms, with positive liquidity and a low level of indebtedness. Liquidity, measured by the working capital to total assets ratio, has a mean value of 10%. With positive working capital, firms do not seem to be dependent on external financing. Profitability is 27%, measured by the return on assets ratio. Therefore, our firms are generally profitable, as they get 27 dollars of profit for every 100 dollars of investment in assets. The companies also have low indebtedness, as the leverage ratio (book value of equity to total liabilities ratio) is 6%.

Regarding external audit information (Table 3), all companies were audited, since this is mandatory for listed companies in Spain. Most of them (91%) were audited by Big 4 audit firms (KMPG, PWC, Deloitte or EY), and only 9% by non-Big 4 auditors. All audit opinions were unqualified (clean), except in relation to three companies. Even though the opinions were clean, the univariate analysis seems to start signalling the usefulness of the expanded audit report as an appropriate tool to anticipate potential financial and credit corporate risks because there are four audit reports with emphasis of matter paragraphs and twelve with GCUPs. Without qualifying the audit opinion (the opinion is still clean), emphasis paragraphs indicate significant uncertainties, or any other matters disclosed in the notes to the financial statements. Likewise, GCUPs do not qualify the audit opinion but highlight that the auditors are questioning the company's future viability. This evidence is in line with the existing literature on qualified opinions, which supports the informativeness of audit opinions for investors, as stock prices adjust in response to qualified opinions on the going concern principle, violations of accounting standards and disclosure rules (Chen et al., 2020).

The distribution of KAMs by industry is also shown in Table 4. As with the emphasis paragraphs and GCUPs, the disclosure of a KAM section indicates a significant risk of material misstatement or a questionable management judgment, but the audit opinion is still clean. There are 371 KAMs, which suggests that each audit report discloses 2.83 KAMs on average. The most common KAM category relates to asset valuation (55.2%). In relation to the repeated observance of individual asset KAMs, the most frequent ones relate to taxes (53 observations) and goodwill (49 observations), as found in a recent study in Spain (Pérez Pérez, 2020). Revenue, liabilities and other KAM categories are less common, representing 16.7%, 16.2% and 10.2% of all KAMs, respectively. The least frequent KAM category is GC, accounting for 1.7% of all KAMs. Despite the small number of GC KAMs identified, it is important to mention that six reports highlight going concern risks, indicating that the company's ongoing future could be in danger. These going concern uncertainties once again show the usefulness of the audit report to investors (Chen et al., 2020).

4.2 Machine learning techniques: model 1 (auditing variables model) results

Moving on to the multivariate analyses, we use the non-financial population of 116 companies to apply machine learning techniques, eliminating the 15 financial institutions due to their special characteristics and regulatory requirements. Our main results (auditing variables model results) are presented in Table 5.

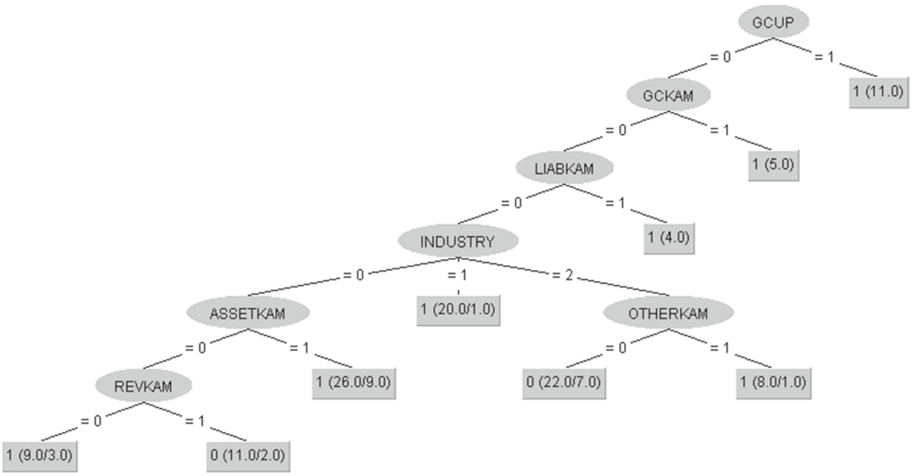
4.2.1 Decision tree (C4.5) results

The evidence using the C4.5 methodology is summarised in Table 5 Panel A, presented in a tree-shaped figure with 9 leaves. The results show that the main indicator for classifying credit rating scores is the going concern uncertainty paragraph (*GCUP*). When an audit report includes this paragraph (*GCUP* value is 1), the company has a low credit rating in 100% of cases (11 observations without any classification errors). Consequently, the company's risk is high. This result is in line with the existing literature on pass/fail audit reports (Gutierrez et al., 2020; Muñoz-Izquierdo et al., 2020; Camacho-Miñano, Muñoz-Izquierdo, Pincus & Wellmeyer, 2021) because the GCO is a common warning generally issued when a firm's financial viability is in doubt (Muñoz-Izquierdo et al., 2019a).

To identify high credit risk firms, the strongest branch of the tree is *GCUP = 0; GCKAM = 0; LIABKAM = 0; Industry = 1*, with 20 observations correctly classified and one error. This branch indicates that for a low credit rating assessment, the user must confirm the inexistence of GCUP and two types of KAMs: going concern KAMs (*GCKAM*) and liabilities KAMs (*LIABKAM*). Indeed, *GCKAM* seems to be a key factor in explaining the causes of failure (Altman et al., 2010). If there is a *GCKAM*, the corporate credit rating is low, indicating a higher credit risk. The existence of *LIABKAM* indicates concerns about

Table 5 Model 1 (auditing variables) results: C4.5, PART, Rough Set and logit regression

Panel A. Decision tree results: C4.5



Number of leaves: 9

Panel B. PART results: PART decision list

GCUP = 0 & *GCKAM* = 0 & *LIABKAM* = 0 & *INDUSTRY* = 1: 1 (20/1)

GCUP = 0 & *GCKAM* = 0 & *LIABKAM* = 0 & *OTHERKAM* = 1: 1 (21/6)

GCUP=0 & *GCKAM*=0 & *LIABKAM*=0 & *AUSIZE*=0 & *REVKAM*=1 & *FIRMAGE*=1 & *ASSETKAM*=0: 0 (15/4)

GCUP = 1: 1 (11)

GCKAM = 0 & *LIABKAM* = 0 & *INDUSTRY* = 2 & *REVKAM* = 0 & *FIRMSIZE* = 1 & *ASSETKAM* = 0: 1 (5/1)

GCKAM = 0 & *LIABKAM* = 0 & *INDUSTRY* = 0: 1 (23/8)

INDUSTRY = 2: 0 (14/3)

Number of rules: 7

Panel C. Rough Set results: Rough Set rules

# Rule	CLASS (*CR*)	*OTHERKAM*	*INDUSTRY*	*REVKAM*	*GCUP*	Strength
1	1	0	1			22
2	1		1	0		12
3	1				1	11

Number of total rules: 72

Panel D. Logit regression analysis results

Independent variables	Coefficients and statistical significance	Standard errors
Constant	−1.029	1.16
AUSIZE	1.841*	1.07
ENTKAM	1.192**	0.56
ACCKAM	0.283	0.47
FIRMSIZE	0.710	0.99
FIRMAGE	0.160	0.59

Table 5 (continued)

Panel D. Logit regression analysis results			
Independent variables	Coefficients and statistical significance	Standard errors	
INDUSTRY 1 (construction)	2.932*	1.28	
INDUSTRY 2 (commercial)	−0.202	0.48	
Observations: 103	Pseudo R-squared: 0.154	Wald chi2: 9.16	VIF of the model: 2.47

This table presents the results of model 1 (audit variables model) applying the four machine learning techniques. In the logit results, ***, **, * denote statistical significance at the 1%, 5%, and 10% levels, respectively. Decision tree results are shown in a tree-form graph

a company's current debt level, estimations of potential provisions or the need to record contingencies for potential debts (Muñoz Izquierdo et al., 2020). Together with this, the company's industry is also relevant in assessing credit rating because construction firms appear to be less risky when their reports do not disclose any of these paragraphs.

4.2.2 The PART algorithm results

The PART algorithm, which generates 7 rules to explain corporate credit ratings, is set out in Table 5 Panel B. The rules are like the branches obtained using the C4.5 decision tree, complementing the evidence from this initial technique (see Sect. 4.2.1). The evidence provided by the PART algorithm suggests that going concern uncertainty paragraphs (*GCUP*), going concern KAMs (*GCKAM*) and liability KAMs (*LIABKAM*) are accurate predictors of corporate credit rating. These disclosures in the audit report generally indicate high-risk firms.

4.2.3 The rough set results

The Rough Set technique obtains 72 rules. Some of these are only supported by one or two objects (audit variables), so only the strongest rules are set out in Table 5 Panel B.

The strongest rules explain low credit ratings (when *CR* class equals 1). This means that it is easy to identify patterns for high-risk firms. The first and second rules suggest that, for construction companies, the credit rating may be low even when the auditor does not disclose any risks in revenue recognition (*REVKAM*) and in external aspects less controllable by the firms (*OTHERKAM*). The third rule supports one of the results also found with the C4.5 and PART methodologies. It demonstrates that the disclosure of a GCUP is a clear sign of corporate viability problems.

In addition, one reduct in the Rough Set model considers two independent variables as redundant, so they are eliminated from the model. These variables are audit opinion (*AUOP*) and emphasis of matter paragraphs disclosed (*EMP*). This result corroborates the evidence provided by the C4.5 decision tree and PART algorithm, as these two variables were similarly eliminated in those other techniques.

4.2.4 Logistic regression analysis results

The results of the logit regression are summarised in Table 5 Panel 4. Due to the population size, we have aggregated the 5-category KAM codification into a 2-item codification to run a logistic regression: entity KAMs (*ENTKAM*) and accounting KAMs (*ACCKAM*). This classification has been used in prior papers such as Lennox, Schmidt and Thompson (2022), Sierra-García et al. (2019), and Camacho-Miñano et al. (2021). Entity KAMs (*ENTKAM*) relate to the risks of material misstatement that have a pervasive impact on the organisation and its financial statements, such as those related to a client's internal controls and GC. This category therefore includes GC and other KAMs. Accounting KAMs (*ACCKAM*) represent the risk of misstatements affecting a specific account in a client's financial statements, for example, the valuation of the firm's intangible assets and liabilities or the recognition of revenues. In our analyses, these two variables are dichotomous. *ENTKAM* equals 1 when the auditor mentions any of these risks, and 0 otherwise. Similarly, if accounting KAMs are disclosed, the variable *ACCKAM* takes the value 1, and if, according to the auditor, these risks do not exist, it takes the value 0.

In our main results, using logit, auditor size (*AUSIZE*), entity KAMs (*ENTKAM*) and the construction industry are significant in assessing corporate credit ratings. Thus, if the company has a GC discussion in its audit report, the variable *ENTKAM* equals 1 and, consequently, the firm shows low credit rating and high credit risk.

4.2.5 Accuracy comparison and discussion of the results

We have compared the performance of the four machine learning techniques using three metrics: accuracy (Acc), specificity (Sp) and sensitivity (Se). See Table 8 for the results for model 1.

Acc represents the rate of correctly classified firms. The PART algorithm achieves the highest classification accuracy of 74.14%, followed by the C4.5 decision tree (73.28%), the Rough Set (72.70%) and the logit model (69.90%). The lowest result being for the logit regression is consistent with the existing literature that suggests that other machine learning techniques provide better accuracy than traditional logistic regression (Crook et al., 2007; Huang et al., 2004; Ong et al., 2005).

Sp and Se measures help to identify type I and II errors. Se is higher than Sp for all methods, reaching 84.61% in the PART algorithm. These results indicate that there are fewer type II errors than type I errors. These findings demonstrate the usefulness of our model 1 since type II misclassifications can be the costliest for decision makers when assessing credit ratings (false negative or assigning high rating to a risky firm).

In conclusion, our evidence from model 1 supports our first hypothesis (H1), suggesting that KAMs disclosed in the expanded audit report are significant in assessing corporate credit ratings. In particular, the best predictors appear to be KAM disclosures related to GC (*GCKAM*) and concerns about company debts (*LIABKAM*).

Our evidence demonstrates the reliability and significant prediction power of the disclosure in the audit report related to going concern uncertainties. Previous research supports the usefulness of GC opinions for assessing firms' financial positions (Altman & McGough, 1974; Altman et al., 2010). However, we extend this evidence, finding that it is not necessary to qualify (fail) the opinion with a GC qualification to arrive at a low credit rating. When the auditor issues an unqualified (pass) expanded report including a GCUP or a GCKAM, these in themselves are signals of a low credit rating.

When the auditor highlights issues relating to company debts through KAMs (*LIAB-KAM*), these may represent credit risk signals. This finding complements previous studies about other disclosures in traditional pass/fail reports. Muñoz-Izquierdo et al. (2020) suggest that the number of disclosures included in traditional reports, as well as remarks on assets and the profit generated, are the best corporate distress predictors. Our paper adds to these findings, suggesting that a company's external financing may similarly represent a warning for investors when deciding whether to invest in a business.

With regards to our second hypothesis (H2), our evidence supports this, suggesting that the application of machine learning techniques to KAMs disclosed in the expanded audit report (model 1) results in up to 74% of companies being correctly classified. Comparing this with existing research, this percentage is an improvement on the rates obtained by most of the previous studies on corporate credit rating (Golbayani et al., 2020). Not only do we achieve slightly higher accuracy than in the existing literature, but also we do this using easier, zero-cost information. The decision-maker does not need to have a strong accounting background because KAM topics are easy to understand and to find, as they appear on the first pages of every company's annual report. KAMS are usually shown in a table and/or with headings. Instead of analysing financial data to assess corporate credit rating, as seen in previous studies, any decision-maker may assess credit ratings with similar accuracy by simply identifying the KAM topics disclosed in the report.

4.3 Machine learning techniques: additional analyses

4.3.1 Model 2 (financial variables model) results

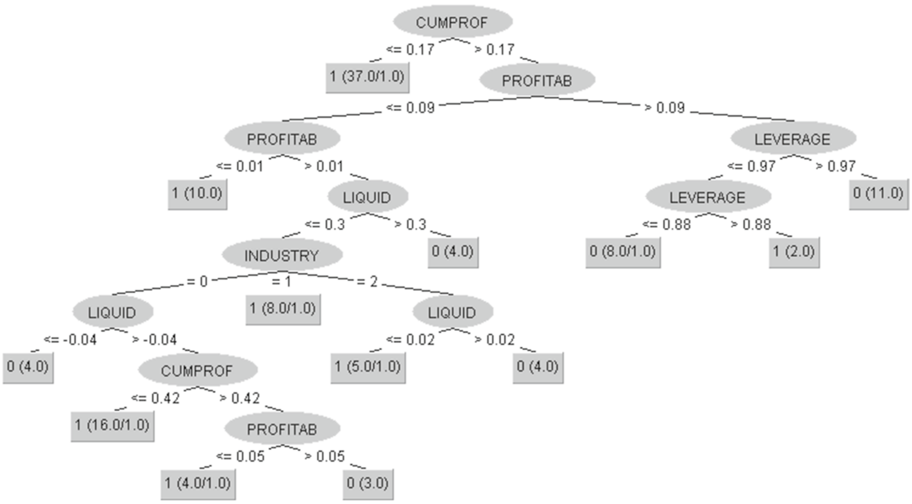
Model 2, the model only using financial variables, tests the predictive value of financial variables in explaining corporate credit rating. The findings using the four machine learning techniques appear in Table 6.

In our study, when financial data are used to predict corporate credit ratings, we categorize each company's financial position into the four accounting ratios of the Altman's Z''-Score model, as mentioned earlier, due to its relevance, frequent use and popularity in the existing research (Altman et al., 2017). As stated in the methodology (Sect. 3), the Z''-Score model includes two profitability measures (retained earnings to total assets and return on assets), one liquidity ratio (working capital to total assets) and a leverage measure (total equity to total liabilities).

In the financial variables only prediction model using the C4.5 methodology, our decision tree contains 13 leaves and suggests mainly that low profitability is a sign of a low credit rating. Two significant leaves corroborate this result: 37 observations with 1 error, using each company's cumulative profitability (retained earnings by total assets) and 10 correctly classified observations using the current profitability ratio (return on assets). These results can be explained with similar evidence found in the related field of corporate bankruptcy prediction. According to previous research, return on assets appears to be the most powerful predictor, as it continually outperforms other measures in assessing the failure risk (Altman et al., 2017). Therefore, if a negative correlation between profitability measures and financial distress is found, a decline in profitability might also signal a low corporate credit rating.

Table 6 Model 2 (financial variables) results: C4.5, PART, Rough Set and logit regression

Panel A. Decision tree results: C 4.5



Number of leaves: 13

Panel B. PART results: PART decision list

CUMPROF <= 0.17: 1 (37/1)

PROFITAB > 0.09 & *FIRMAGE* = 1 & *LEVERAGE* > 0.97: 0 (10)

PROFITAB <= 0.01: 1 (10)

INDUSTRY = 1: 1 (8/1)

INDUSTRY = 2 & *LIQUID* > 0.01: 0 (8)

LIQUID > − 0.04 & *LIQUID* <= 0.3 & *INDUSTRY* = 0 & *FIRMAGE* = 1 & *CUMPROF* > 0.22 & *CUMPROF* <= 0.42: 1 (13)

INDUSTRY = 0 & *FIRMAGE* = 1 & *PROFITAB* > 0.05: 0 (10/1)

LIQUID <= − 0.04: 0 (7/1)

INDUSTRY = 2: 1 (5/1)

FIRMAGE = 0: 1 (3)

CUMPROF > 0.48: 1 (3)

CUMPROF > 0.48: 0 (2)

Number of rules: 12

Panel C. Rough Set results: Rough Set rules

# Rule	CLASS (*CR*)	*LEVERAGE*	*PROFITAB*	Strength
1	1	0.2		12
2	1		0.1	10
Number of total rules: 38				

Panel D. Logit regression analysis results

Independent variables	Coefficients and statistical significance	Standard errors
Constant	2.594**	1.02

Table 6 (continued)

Panel D. Logit regression analysis results

Independent variables	Coefficients and statistical significance	Standard errors	
LIQUID	1.095	1.84	
CUMPROF	−8.092***	2.94	
PROFITAB	−3.560	2.22	
LEVERAGE	0.547	0.69	
FIRMAGE	0.453	0.65	
FIRMSIZE	0.047	0.74	
INDUSTRY 1 (construction)	2.443*	1.25	
INDUSTRY 2 (commercial)	−0.667	0.60	
Observations: 116	Pseudo R-squared: 0.323	Wald chi2: 27.07	VIF of the model: 3.03

This table presents the results of model 2 (financial variables model) applying the four machine learning techniques. In the logit results, ***, **, * denote statistical significance at the 1%, 5%, and 10% levels, respectively

Similarly, the C4.5 decision tree also builds a significant branch using profitability measures for the high credit rating classification. 11 observations (and no misclassifications) suggest that high profitability and a leverage ratio close to 1 (specifically 0.97) are accurate predictors of a high credit rating. As the leverage ratio measures the relationship between internal and external financing, the value of this measure indicates that companies with a low credit risk do not depend on funding from external parties. In contrast, shrinkage in leverage could warn the decision-maker about possible financial difficulties.

The remaining methodologies validate and complement these findings. PART rules are very similar to the branches of the C4.5 decision tree, the profitability and leverage measures being the ones with highest predictive power. In the Rough Set, 38 rules are generated but the strongest ones also suggest that high credit risk appears when leverage and profitability are low (12 observations and 10 observations, respectively). In the logit model, the most statistically significant variable in assessing corporate credit ratings is cumulative profitability ($p = 0.006$). As explained earlier, this finding is associated with the ratios identified as significant in assessing corporate financial distress. For example, in the review of research by Bellovary et al. (2007), they find that the cumulative profitability measure of the Z''-Score model is one of the top ten ratios used in the bankruptcy prediction literature.

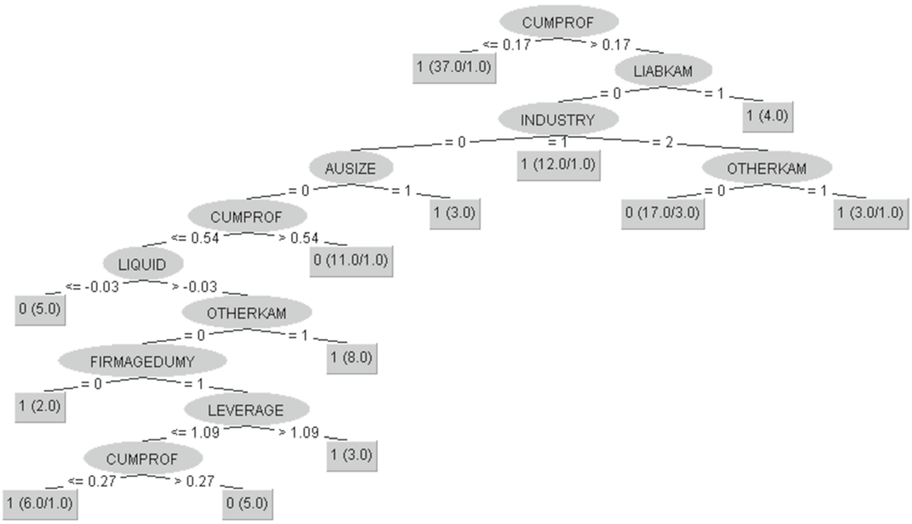
4.3.2 Model 3 (combined auditing and financial variables model) results

In models 1 and 2, the predictive power of auditing and financial data on corporate credit ratings was examined separately. In model 3, we test the incremental predictive power of combining both data sources. This model supplements the benchmark model (auditing variables model) by adding the financial information variables, examining their effect on classification accuracy.

In the related research area of predicting corporate financial distress, Balcaen and Ooghe (2006) highlight the importance of supplementing the commonly used accounting ratios with non-financial data. Specifically, in our combined model 3, we test the predictive

Table 7 Model 3 (combined) results: C4.5, PART, Rough Set and logit regression

Panel A. Decision tree results: C4.5



Number of leaves: 13

Panel B. PART results: PART decision list

CUMPROF <= 0.17: 1 (37/1)

LIABKAM = 0 AND *INDUSTRY* = 1: 1 (12/1)

LIABKAM = 0 AND *PROFITAB* > 0.09 AND *REVKAM* = 1: 0 (13)

LIABKAM = 0 AND *FIRMSIZE* = 1 AND *AUSIZE* = 0 AND *PROFITAB* > 0.01 AND *CUMPROF* > 0.54: 0 (7)

LIABKAM = 0 AND *FIRMSIZE* = 1 AND *AUSIZE* = 0 AND *LIQUID* > −0.04 AND *OTHERKAM* = 1: 1 (10/1)

LIABKAM = 0 AND *OTHERKAM* = 0 AND *FIRMSIZE* = 1 AND *AUSIZE* = 0 AND

LEVERAGE > 1.01: 1 (6)

LIABKAM = 0 AND *LEVERAGE* <= 0.41: 1 (8/1)

LIABKAM = 0 AND *AUSIZE* = 0 AND *ASSETKAM* = 0: 0 (7)

LIABKAM = 1: 1 (4)

INDUSTRY = 0 AND *AUSIZE* = 0 AND *CUMPROF* <= 0.3: 1 (3/1)

AUSIZE = 0: 0 (5)

AUSIZE = 0: 1 (4/1)

Number of rules: 12

Panel C. Rough Set results: Rough Set rules

# Rule	CLASS (*CR*)	*OTHERKAM*	*INDUSTRY*	*LEVERAGE*	*REVKAM*	LIQUID	PROFITAB	GCUP	Strength
1	1	0	1						22
2	1			0.2					12
3	1		1		0				12
4	1	0				0.1	0		11
5	1							1	11
6	1						0.1		10

Table 7 (continued)

Panel C. Rough Set results: Rough Set rules

# Rule	CLASS (*CR*)	*OTHERKAM*	*INDUSTRY*	*LEVERAGE*	*REVKAM*	LIQUID	PROFITAB	GCUP	Strength
Number of total rules: 266									

Panel D. Logit regression analysis results

Independent variables	Coefficients and statistical significance	Standard errors	
Constant	1.049	1.90	
AUSIZE	3.514***	1.26	
REVKAM	– 1.092*	0.57	
OTHERKAM	0.754	0.77	
ASSETKAM	– 0.047	0.69	
LIQUID	2.719	2.66	
CUMPROF	– 8.703**	3.96	
PROFITAB	– 18.236**	8.35	
LEVERAGE	0.355	0.89	
FIRMAGE	0.275	0.82	
FIRMSIZE	2.797**	1.28	
INDUSTRY 1 (construction)	2.779	1.79	
INDUSTRY 2 (commercial)	– 0.517	0.75	
Observations: 94	Pseudo R-squared: 0.423	Wald chi2: 34.08	VIF of the model: 3.30

This table presents the results of model 3 (auditing and financial variables combined model) applying the four machine learning techniques. In the logit results, ***, **, * denote statistical significance at the 1%, 5%, and 10% levels, respectively

power of the four accounting ratios of Altman's Z''-Score model and the KAMs disclosed in the expanded audit report. Results are set out in Table 7.

Our results show that the combined use of auditing and financial factors leads to a more accurate prediction of corporate credit rating than the individual use of each of these factors. Accounting ratios do not include all the significant information for predicting distress and non-financial variables, more precisely audit variables, are likely to offset this deficiency.

Two significant branches are built by the C4.5 decision tree. One is to identify firms with low ratings and the other is to identify those with high ratings. If the retained earnings over assets ratio is low, this indicates low credit ratings (37 observations and 1 misclassification). In contrast, when assessing the high ratings of commercial and service companies in particular, the audit report is important because a combination of high retained earnings over assets and the absence of KAMs in the report provides the most significant path (17 correctly classified observations with 3 misclassifications). Specifically, for high ratings, KAMs disclosing liabilities (*LIABKAM*) and general (*OTHERKAM*) risks (such as

Table 8 Performance metrics comparison by model and techniques

	Performance metrics								
	Model 1			Model 2			Model 3		
Techniques	Acc (%)	Sp (%)	Se (%)	Acc (%)	Sp (%)	Se (%)	Acc (%)	Sp (%)	Se (%)
Decision tree C4.5	73.28	55.26	82.05	75.00	55.26	84.61	80.17	71.05	84.16
PART Algorithm	74.14	52.63	84.61	74.14	60.53	80.77	75.86	65.79	80.76
Rough Set	72.70	68.42	82.35	73.40	75.00	72.60	75.50	61.11	82.43
Logit regression	69.90	60.53	75.38	81.03	60.53	91.03	84.04	71.05	92.86

In this table, we show three performance metrics of the 3 models comparing the machine learning techniques applied. Acc states for accuracy (correctly classified companies); Sp is the specificity and Se the sensitivity of the models applying every technique. Note that in model 1 of the logit regression analysis, we aggregate the 5-item KAM codification into a 2-item (entity and accounting KAMs) classification, which slightly improves the accuracy of the results

issues regarding information systems or business combinations) should be absent. Similar evidence can be found using the PART algorithm, which results in 12 rules, the strongest one being the same as that found in the decision tree (low cumulative profitability to assess risky firms) for construction companies.

The Rough Set in model 3 generated a total of 266 rules and the most significant ones classify low credit rating firms. Adding to the evidence of the previous methodologies discussed, the Rough Set suggests that in construction companies, it is important to analyse both the KAMs and the accounting ratios, as the rating may be low even without warnings in the audit report ($OTHERKAM=0$; $INDUSTRY=1$ and $REVKAM=0$; $INDUSTRY=1$). This finding confirms the conjecture that accuracy in predicting corporate credit ratings increases when auditing data are combined with financial data.

In summary, the novelty of our study is that it extends the corporate credit rating assessment literature to include KAM disclosures. Additionally, the predictive power of our model improves significantly when combining both auditing data and financial variables (see Sect. 4.3.3 for details).

4.3.3 Performance metrics comparison and discussion of the results

Three performance metrics were compared for the three estimated models to test the predictive power of the machine learning techniques applied. Table 8 summarises the results. Model 1's results have been explained in Sect. 4.2.5. The performance of models 2 (only financial variables) and 3 (combined auditing and financial variables) follow this.

The efficacy of the Z''-Score model has been tested in the related literature on bankruptcy prediction. Our logit results for model 2 improve on the ones obtained by Altman et al. (2017). They found that the model performs reasonably well for most countries and in their Spanish sample, the classification performance was 73.40%. The accuracy of the logit regression in our model 2 is 81.03%. Although the other techniques provide slightly lower accuracy, it is still at a reasonable level. The rate of correctly classified cases is 75.00% with the C4.5 decision tree, 74.14% with the PART algorithm and 73.40% with the Rough set technique.

The pattern in the combined model of auditing and financial data (model 3) is similar in terms of accuracy. The most powerful predictor is the logit regression (84.04%). The second, third and fourth most accurate techniques are, in order, the C4.5 decision tree (80.17%), the PART algorithm (75.86%) and the Rough Set (75.50%).

As with model 1, in terms of type I and type II errors, type II misclassifications are less common than type I errors in most cases, meaning that there are more non-risky companies rated as risky than there are risky companies classified as non-risky. This verifies the usefulness of both models 2 and 3 for users of annual reports. Data from analysing the topics in KAMs and the accounting ratios in corporate financial statements can be used to accurately suggest whether to invest in that business.

5 Conclusion

In this paper, we apply well-known machine learning techniques, namely the C4.5 decision tree, PART algorithm rule classifier, Rough set methodology and logistic regression, to a real-world problem: the assessment of corporate credit ratings. The novelty of this study lies in it being the first to introduce data from the expanded audit report to the field of credit rating. We hypothesise that the information included in expanded audit reports, and in particular the KAMs disclosed, add explanatory power when predicting corporate credit ratings and that machine learning algorithms or data mining techniques are appropriate tools to test this prediction.

Using the listed firms on the Spanish trading market in 2017, the first year in which it was mandatory to implement the expanded audit report, our evidence supports the hypotheses proposed. The results are consistent across the four different techniques used and suggest that the KAMs disclosed in expanded audit reports have up to 74.14% accuracy when explaining corporate credit ratings. The highest predictive percentage is obtained with the PART algorithm, and this is slightly higher than the figures published in the previous literature (Golbayani et al., 2020). This evidence contributes to the credit rating literature. Previous authors have stated that even a slight improvement in credit rating accuracy might reduce large credit risks and translate into significant future savings (Tsai & Chen, 2010).

In addition to the predictive power of the models tested in this study, our evidence indicates that KAM disclosures that mention both internal and external aspects of the company contribute to explaining credit ratings. This evidence supports others in the related literature on business failure who have proposed integrating exogenous and endogenous factors to offer a complete explanation of the potential for failure (Amankwah-Amoah, 2016). In particular, KAMs disclosing aspects on corporate debt indicate high possibilities of low credit ratings. Similarly, KAMs regarding GC are warnings from the auditors about the viability of the company in the foreseeable future. A combination of asset KAMs and KAMs about external aspects are signals of low credit ratings. For commercial companies, the auditor comments that explain their corporate credit rating most are KAMs related to exogenous factors and KAMs about company performance or revenues generated by the firm. This evidence is consistent with the previous research on the impact of corporate governance on credit ratings carried out by Ashbaugh-Skaife et al. (2006). They found that credit ratings are positively related to accrual quality and timeliness of earnings. Our main credit rating assessment results using auditing data alone are complemented by our additional analysis combining KAM variables and accounting ratios. Combining both auditing data and accounting variables, the predictive power rises to 84.04%, the logit regression being the best tool to apply.

Some implications can be drawn from these results. First, this paper responds to calls in the literature for improvements to credit rating prediction (Golbayani et al., 2020). Second, our evidence implies that credit rating can be anticipated not only by using financial and accounting ratios, which is the source of information most commonly used (Hájek, 2012; Huang et al., 2004; Pai et al., 2015), but also by reading the disclosures published in the expanded audit report, easily accessible in publicly available corporate annual reports. Our models achieve a high explanatory power when predicting corporate credit ratings relying on audit information alone, which is an innovative contribution. Any reader of an annual report can easily find the auditor's disclosures in the first pages of the report and understand the topic to which they relate, without any specific accounting knowledge or background in finance. Third, this paper extends the previous work on credit rating prediction by ascertaining whether corporate aspects in general, and which types of these specifically, significantly assist in predicting ratings. Examples include debts, GC and some specific external factors mentioned by the auditor in the KAMs disclosed. Fourth, to complement our main results, when combining the KAMs disclosed with the accounting ratios the predictive power of our models notably improved to a level higher than in the previous literature (Golbayani et al., 2020). Finally, our evidence could represent a timely and important contribution to the current international audit reform. The reform began in 2016 and in 2020 is still being implemented in countries such as the US. The aim is to improve the transparency and informational value of the audit report to inform about relevant aspects of the audit procedure. Our study has found evidence that indicates that the expanded audit report is indeed a valuable tool for anticipating a company's credit rating.

This study is not free of its limitations. Due to our population size, some KAMs may not appear to be significant in the logistic regression model and others are not included in the other three machine learning techniques (decision tree and rule induction classifiers) when explaining credit ratings. However, we analyse 116 companies, which is the total population of non-financial Spanish listed firms required to issue the expanded audit report in its first year of implementation. We acknowledge that the number of firms studied is small, but this is due to the dimensions of the market chosen (not the sample selection). Other recent studies on expanded auditor disclosures have also focused on small markets, such as Spain (Hsieh et al., 2021), Thailand (Suttipun, 2022), Jordan (Abdullatif & Al-Rahahleh, 2020) or Brazil (Ferreira & Morais, 2019), and small samples in the UK market (Sierra-García et al., 2019). Hence, small markets and reduced samples are relevant and important to address, so we believe that whether KAM paragraphs contribute to assess corporate credit ratings using the Spanish market is an important empirical question to assess. As a future line of research, we recommend increasing the number of observations to obtain more robust results. For instance, a multinational study about the expanded audit reporting regulation could shed interesting results in credit rating assessment, including and comparing auditor disclosures from several countries.

Authors' contributions Conceptualisation, All; Data curation, NMI and YPP; Formal analysis, NMI and MJSV; Investigation, NMI and MJSV; Methodology, software and validation, MJSV, NMI and MMCM; Project administration and supervision, NMI; Writing—original draft, NMI and YPP; Writing—review & editing, All.

Funding This paper is part of an ongoing project of the Banco de Santander and UCM research call (reference number PR87/19-22586). Also, we gratefully acknowledge the financial support provided by the

Spanish Ministry of Science and Innovation (reference number PID 2020-115700RB-I00). Finally, we thank the editor and the anonymous reviewers for helpful comments and suggestions.

Availability of data and materials The data that support the findings of this study are available from the authors upon reasonable request.

Code availability Not applicable.

Declarations

Conflict of interest The authors declare no conflicts of interest.

Consent to participate Not applicable.

Consent for publication Not applicable.

Ethics approval Not applicable.

References

Abdullatif, M., & Al-Rahahleh, A. S. (2020). Applying a new audit regulation: Reporting key audit matters in Jordan. *International Journal of Auditing, 24*(2), 268–291.

Addo, P. M., Guegan, D., & Hassani, B. (2018). Credit risk analysis using machine and deep learning models. *Risks, 6*(38), 1–20.

Altman, E. I. (1983). *Corporate financial distress. A complete guide to predicting, avoiding, and dealing with bankruptcy*. Wiley.

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 and Accounting, 27*, 131–171.

Altman, E. I., & McGough, T. P. (1974). Evaluation of a company as a going concern. *Journal of Accountancy, 138*, 50–57.

Altman, E. I., Sabato, G., & Wilson, N. (2010). The value of non-financial information in small and medium-sized enterprise risk management. *Journal of Credit Risk, 2*(6), 95–127.

Amankwah-Amoah, J. (2016). An integrative process model of organisational failure. *Journal of Business Research, 69*(9), 3388–3397.

Ashbaugh-Skaife, H., Collins, D. W., & LaFond, R. (2006). The effects of corporate governance on firms' credit ratings. *Journal of Accounting and Economics, 42*(1–2), 203–243.

Ayres, D. (2015). *Accounting Information Risk and Credit Ratings*. PhD diss., University of Tennessee. https://trace.tennessee.edu/utk_graddiss/3321

Balcaen, S., & Ooghe, H. (2006). 35 years of studies on business failure: An overview of the classic statistical methodologies and their related problems. *The British Accounting Review, 38*, 63–93.

Bellovary, J. L., Giacomino, D. E., & Akers, M. D. (2007). A review of bankruptcy prediction studies: 1930 to present. *Journal of Financial Education, 33*, 1–42.

Camacho-Miñano, M. M., Muñoz-Izquierdo, N., Pincus, M., & Wellmeyer, P. (2021). Are key audit matter disclosures useful in assessing the financial distress level of a firm? *SSRN:* https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3744282

Camacho-Miñano, M. M., Segovia-Vargas, M. J., & Pascual-Ezama, D. (2015). Which characteristics predict the survival of insolvent firms? An SME reorganisation prediction model. *Journal of Small Business Management, 53*(2), 340–354.

Campanella, F. (2014). Assess the rating of SMEs by using classification and regression trees (CART) with qualitative variables. *Review of Economics & Finance, 4*(3), 16–32.

Caramanis, C., & Lennox, C. (2008). Audit effort and earnings management. *Journal of Accounting and Economics, 45*(1), 116–138.

Caridad, D., Hančlová, H., Bousselmi, H. W., & López del Río, L. C. (2019). Corporate rating forecasting using artificial intelligence statistical techniques. *Investment Management & Financial Innovations, 16*(2), 295–312.

Cha, M., Hwang, K., & Yeo, Y. (2016). Relationship between audit opinion and credit rating: Evidence from Korea. *Journal of Applied Business Research, 32*(2), 621–634.

Charitou, A., Lambertides, N., & Trigeorgis, L. (2007). Earnings behavior of financially distressed firms: The role of institutional ownership. *Abacus, 43*, 271–296.

Chen, S., Hu, B., Wu, D., & Zhao, Z. (2020). When auditors say 'no', does the market listen? *European Accounting Review, 29*(2), 263–305.

Coad, A., Segarra, A., & Teruel, M. (2013). Like milk or wine: Does firm performance improve with age? *Structural Change and Economic Dynamics, 24*, 173–189.

Crabtree, A., & Maher, J. (2012). Credit ratings, cost of debt, and internal control disclosures: A comparison of SOX 302 and SOX 404. *Journal of Applied Business Research, 28*, 885–902.

Crook, J. N., Edelman, D. B., & Thomas, L. C. (2007). Recent developments in consumer credit risk assessment. *European Journal of Operational Research, 183*(3), 1447–1465.

DeAngelo, L. E. (1981). Auditor size and audit quality. *Journal of Accounting and Economics, 3*(3), 183–199.

Dedman, E., & Kausar, A. (2012). The impact of voluntary audit on credit ratings: Evidence from UK private firms. *Accounting and Business Research, 42*(4), 397–418.

Díaz-Martínez, Z., Sánchez-Arellano, A., & Segovia-Vargas, M. J. (2009). Analysis of financial instability by means of decision trees and lists. In R. O. Bailly (Ed.), *Emerging topics in macroeconomics* (pp. 303–327). Editorial Nova Publishers.

Díaz-Martínez, Z., Sánchez-Arellano, A., & Segovia-Vargas, M. J. (2011). Prediction of financial crises by means of rough sets and decision trees. *Innovar, 21*(39), 83–100.

Elbannan, M. (2008). Quality of internal control over financial reporting, corporate governance and credit ratings. *International Journal of Disclosure and Governance, 6*, 127–149.

Feldmann, D., & Read, W. J. (2013). Going-concern audit opinions for bankrupt companies—Impact of credit rating. *Managerial Auditing Journal, 28*(4), 345–363.

Ferreira, C., & Morais, A. I. (2019). Analysis of the relationship between company characteristics and key audit matters disclosed. *Revista Contabilidade & Finanças, 31*, 262–274.

Frank, E., & Witten, I. H. (1998). *Generating accurate rule sets without global optimization*. Working paper 98/2. University of Waikato, Hamilton, New Zealand.

Funcke, N. (2014). *Credit rating changes and auditor reporting accuracy*. Working paper. Erasmus University Rotterdam.

Gaganis, C., Pasiouras, F., & Doumpos, M. (2007). Probabilistic neural networks for the identification of qualified audit opinions. *Expert Systems with Applications, 32*(1), 114–124.

Gandía, J. L., & Huguet, D. (2020). Audit fees and earnings management: Differences based on the type of audit. *Economic Research-Ekonomska Istraživanja, 31*(1), 2628–2650.

Geiger, M. A., Raghunandan, K., & Riccardi, W. (2014). The global financial crisis: US bankruptcies and going-concern audit opinions. *Accounting Horizons, 28*(1), 59–75.

Golbayani, P., Florescu, I., & Chatterjee, R. (2020). A comparative study of forecasting corporate credit ratings using neural networks, support vector machines, and decision trees. *The North American Journal of Economics and Finance, 54*, 101251.

Greco, S., Matarazzo, B., & Slowinski, R. (1998). A new rough set approach to evaluation of bankruptcy risk. In C. Zopounidis (Ed.), *New operational tools in the management of financial risks* (pp. 121–136). Kluwer Academic Publishers.

Greco, S., Matarazzo, B., & Slowinski, R. (2001). Rough sets theory for multicriteria decision analysis. *European Journal of Operational Research, 129*(1), 1–47.

Gutierrez, E., Krupa, J., Minutti-Meza, M., & Vulcheva, M. (2020). Do going concern opinions provide incremental information to predict corporate defaults? *Review of Accounting Studies, 25*(4), 1344–1381.

Gutiérrez, P. A., Segovia-Vargas, M. J., Salcedo-Sanz, S., Hervás-Martínez, C., Sanchís, A., Portilla-Figueras, J., & Fernández-Navarro, F. (2010). Hybridizing logistic regression with product unit and RBF networks for accurate detection and prediction of banking crises. *Omega, 38*(5), 333–344.

Hájek, P. (2012). Credit rating analysis using adaptive fuzzy rule-based systems: An industry-specific approach. *Central European Journal of Operations Research, 20*(3), 421–434.

Hájek, P., & Michalak, K. (2013). Feature selection in corporate credit rating prediction. *Knowledge-Based Systems, 51*, 72–84.

Hájek, P., & Olej, V. (2011). Credit rating modelling by kernel-based approaches with supervised and semi-supervised learning. *Neural Computing and Applications, 20*(6), 761–773.

Hernandez Tinoco, M. H., & Wilson, N. (2013). Financial distress and bankruptcy prediction among listed companies using accounting, market and macroeconomic variables. *International Review of Financial Analysis, 30*, 394–419.

Hsieh, S. F., Beretta-Custodio, C., & Vasarhelyi, M. A. (2021). The Textual Similarity of KAM Disclosures for Spanish Companies. *International Journal of Digital Accounting Research, 21*, 183–202.

Huang, Z., Chen, H., Hsu, C. J., Chen, W. H., & Wu, S. (2004). Credit rating analysis with support vector machines and neural networks: A market comparative study. *Decision Support Systems, 37*(4), 543–558.

Hwang, R. C. (2013). Forecasting credit ratings with the varying-coefficient model. *Quantitative Finance, 13*(12), 1947–1965.

Hwang, R. C., Cheng, K. F., & Lee, C. F. (2008). On multiple-class prediction of issuer credit ratings. *Applied Stochastic Models in Business and Industry, 25*(5), 535–550.

Hwang, R. C., Chung, H., & Chu, C. K. (2010). Predicting issuer credit ratings using a semiparametric method. *Journal of Empirical Finance, 17*(1), 120–137.

Kend, M., & Nguyen, L. A. (2020). Investigating recent audit reform in the Australian context: An analysis of the KAM disclosures in audit reports 2017–2018. *International Journal of Auditing, 24*(3), 412–430.

Khashman, A. (2010). Neural networks for credit risk evaluation: Investigation of different neural models and learning schemes. *Expert Systems with Applications, 37*(9), 6233–6239.

Kim, K. S. (2005). Predicting bond ratings using publicly available information. *Expert Systems with Applications, 29*(1), 75–81.

Kim, K. J., & Ahn, H. (2012). A corporate credit rating model using multi-class support vector machines with an ordinal pairwise partitioning approach. *Computers & Operations Research, 39*(8), 1800–1811.

Kumar, P. R., & Ravi, V. (2007). Bankruptcy prediction in banks and firms via statistical and intelligent techniques–A review. *European Journal of Operational Research, 180*(1), 1–28.

La Porta, R., Lopez-de-Silanes, F., Shleifer, A., & Vishny, R. (2000). Investor protection and corporate governance. *Journal of Financial Economics, 58*(1–2), 3–27.

Lee, Y. C. (2007). Application of support vector machines to corporate credit rating prediction. *Expert Systems with Applications, 33*(1), 67–74.

Lennox, C. S., Schmidt, J. J., & Thompson, A. M. (2022). Why are expanded audit reports not informative to investors? Evidence from the United Kingdom. *Review of Accounting Studies*. https://doi.org/10.1007/s11142-021-09650-4.

Lennox, C. S. (1999). The accuracy and incremental information content of audit reports in predicting bankruptcy. *Journal of Business Finance & Accounting, 26*(5–6), 757–778.

Lim, H. J., & Mali, D. (2020). Do credit ratings influence the demand / supply of audit effort? *Journal of Applied Accounting Research, 22*(1), 72–92.

McKee, T. E. (2003). Rough sets bankruptcy prediction models versus auditor signaling rates. *Journal of Forecasting, 22*, 569–586.

Moscatelli, M., Parlapiano, F., Narizzano, S., & Viggiano, G. (2020). Corporate default forecasting with machine learning. *Expert Systems with Applications, 161*, 113567.

Muñoz-Izquierdo, N., Camacho-Miñano, M. M., Segovia-Vargas, M. J., & Pascual-Ezama, D. (2019a). Explaining the causes of business failure using audit report disclosures. *Journal of Business Research, 98*, 403–414.

Muñoz-Izquierdo, N., Camacho-Miñano, M. M., Segovia-Vargas, M. J., & Pascual-Ezama, D. (2019b). Is the external audit report useful for bankruptcy prediction? Evidence using artificial intelligence. *International Journal of Financial Studies, 7*(20), 1–23.

Muñoz-Izquierdo, N., Laitinen, E. K., Camacho-Miñano, M. 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.

Ong, C. S., Huang, J. J., & Tzeng, G. H. (2005). Building credit scoring models using genetic programming. *Expert Systems with Applications, 29*(1), 41–47.

Pacelli, V., & Azzollini, M. (2011). An artificial neural network approach for credit risk management. *Journal of Intelligent Learning Systems and Applications, 3*, 103–112.

Pai, P. F., Tan, Y. S., & Hsu, M. F. (2015). Credit rating analysis by the decision-tree support vector machine with ensemble strategies. *International Journal of Fuzzy Systems, 17*(4), 521–530.

Pawlak, Z. (1991). *Rough sets theoretical aspects of reasoning about data*. Kluwer Academic Publishers.

Pawlak, Z., & Skowron, A. (2007). Rudiments of rough sets. *Information Sciences, 177*(1), 3–27.

Perez Pérez, Y. (2020). *The risk on financial information*. PhD thesis dissertation. Complutense University of Madrid, Spain.

Quinlan, J. R. (1993). *C4.5: Programs for machine learning*. Morgan Kaufmann.

Reza, F. M. (1994). *An introduction to Information Theory*. Dover Publications.

Sánchez-Serrano, J. R., Alaminos, D., García-Lagos, F., & Callejón-Gil, A. M. (2020). Predicting audit opinion in consolidated financial statements with artificial neural networks. *Mathematics, 8*(8), 1288.

Sanchis, A., Segovia, M. J., Gil, J. A., Heras, A., & Vilar, J. L. (2007). Rough sets and the role of the monetary policy in financial stability (macroeconomic problem) and the prediction of insolvency in insurance sector (microeconomic problem). *European Journal of Operational Research, 181*(3), 1554–1573.

Shin, K. S., & Han, I. (2001). A case-based approach using inductive indexing for corporate bond rating. *Decision Support Systems, 32*(1), 41–52.

Sierra-García, L., Gambetta, N., García-Benau, M. A., & Orta-Pérez, M. (2019). Understanding the determinants of the magnitude of entity-level risk and account-level risk key audit matters: The case of the United Kingdom. *The British Accounting Review, 51*, 227–240.

Sikka, P. (2009). Financial crisis and the silence of the auditors. *Accounting, Organizations and Society, 34*(6–7), 868–873.

Strickett, M., Hay, D. C., & Lau, D. (2021). The going-concern opinion and the adverse credit rating: An analysis of their relationship. *Accounting Research Journal, 35*(4), 470–489.

Suttipun, M. (2022). External auditor and KAMs reporting in alternative capital market of Thailand. *Meditari Accountancy Research, 30*(1), 74–93.

Tsai, C. F., & Chen, M. L. (2010). Credit rating by hybrid machine learning techniques. *Applied Soft Computing, 10*(2), 374–380.

Tuv, E., Borisov, A., Runger, G., & Torkkola, K. (2009). Feature selection with ensembles, artificial variables, and redundancy elimination. *The Journal of Machine Learning Research, 10*, 1341–1366.

Wallis, M., Kumar, K., & Gepp, A. (2019). Credit rating forecasting using machine learning techniques. In P. Blayney, H. Huo, A. Ryzhov, K. D. Strang, A. Stranieri, L. Sun, M. Zaharia, P. Zhang (eds.)*Managerial perspectives on intelligent big data analytics* (pp. 180–198). IGI Global.

West, D. (2000). Neural network credit scoring models. *Computers and Operations Research, 27*(11–12), 1131–1152.

Yan, X., Li, Y., & Bonne, G. (2014). *Starmine combined credit risk model: Overview and global performance*. Thomson Reuters.

Yeh, C. C., Lin, F., & Hsu, C. Y. (2012). A hybrid KMV model, random forests and rough set theory approach for credit rating. *Knowledge-Based Systems, 33*, 166–172.

Yu, L., Wang, S., & Lai, K. K. (2008). Credit risk assessment with a multistage neural network ensemble learning approach. *Expert Systems with Applications, 34*(2), 1434–1444.

Zalata, A. M., Elzahar, H., & McLaughlin, C. (2020). External audit quality and firms' credit score. *Cogent Business & Management, 7*(1), 1–16.

Zhao, Z., Xu, S., Kang, B. H., Kabir, M. M. J., Liu, Y., & Wasinger, R. (2015). Investigation and improvement of multi-layer perceptron neural networks for credit scoring. *Expert Systems with Applications, 42*(7), 3508–3516.

Ziemer, R. E., & Tranter, W. H. (2002). *Principles of communications: Systems, modulation, and noise*. Wiley.

Authors and Affiliations

Nora Muñoz-Izquierdo[1] · María Jesús Segovia-Vargas[2] · María-del-Mar Camacho-Miñano[3] · Yolanda Pérez-Pérez[3]

María Jesús Segovia-Vargas
mjsegovia@ccee.ucm.es

María-del-Mar Camacho-Miñano
marcamacho@ccee.ucm.es

Yolanda Pérez-Pérez
yperez@kpmg.es

[1] Finance and Accounting Department, CUNEF University, Pirineos 55, 28040 Madrid, Spain

[2] Financial and Actuarial Economics and Statistics Department, Facultad de Ciencias Económicas y Empresariales, Complutense University of Madrid, Campus de Somosaguas, 28223 Pozuelo de Alarcón, Madrid, Spain

[3] Accounting and Finance Department, Facultad de Ciencias Económicas y Empresariales, Complutense University of Madrid, Campus de Somosaguas, 28223 Pozuelo de Alarcón, Madrid, Spain