



INSTITUTO
UNIVERSITÁRIO
DE LISBOA

Modelos preditivos das insolvências: O contributo da auditoria financeira externa nas PME portuguesas

João Miguel Oliveira Ilhéu

Mestrado em Contabilidade

Orientadores:

Professor Doutor Raul Manuel da Silva Laureano, Prof. Associado, ISCTE Business School, Departamento de Métodos Quantitativos para Gestão e Economia

Professora Doutora Márcia Rafaela Cadete dos Santos, Prof. Adjunta, Escola Superior de Ciências Empresariais – Instituto Politécnico de Setúbal, e Investigadora Associada do ISTAR - IUL - Centro de Investigação em Ciências da Informação, Tecnologias e Arquitetura (ISCTE - Instituto Universitário de Lisboa)

Novembro, 2021



**BUSINESS
SCHOOL**

Modelos preditivos das insolvências: O contributo da auditoria financeira externa nas PME portuguesas

João Miguel Oliveira Ilhéu

Mestrado em Contabilidade

Orientadores:

Professor Doutor Raul Manuel da Silva Laureano, Prof. Associado, ISCTE Business School,
Departamento de Métodos Quantitativos para Gestão e Economia

Professora Doutora Márcia Rafaela Cadete dos Santos, Prof. Adjunta, Escola Superior de
Ciências Empresariais - Instituto Politécnico de Setúbal, e Investigadora Associada do ISTAR -
IUL - Centro de Investigação em Ciências da Informação, Tecnologias e Arquitetura (ISCTE -
Instituto Universitário de Lisboa)

Novembro, 2021

Agradecimentos

Os anos de 2020 e 2021 foram, sem dúvida, anos desafiantes para todos nós em todos os aspectos com momentos bons e outros menos bons. Deste modo, quero agradecer a todos aqueles que contribuíram para que a realização deste projeto fosse possível.

Gostaria de agradecer ao professor Doutor Raul Laureano e à professora Doutora Márcia Santos por todo o conhecimento transmitido, ajuda, recomendações e disponibilidade ao longo deste ano.

Um agradecimento especial aos meus pais pelo percurso acadêmico que me proporcionaram e por todos os valores que me transmitiram. Sem eles não teria sido possível chegar aqui.

Por último, mas não menos importante, agradecer a toda a minha família, namorada e amigos pela motivação e palavras de incentivo.

O meu sincero obrigado a todos os que me apoiaram,
João.

Sumário

Em Portugal, desde a crise financeira, o número de PME que entram em insolvência é bastante elevado e preocupante pelos impactos que causam na economia e na sociedade. Embora já tenham sido desenvolvidos diversos modelos preditivos das insolvências cujos preditores são, essencialmente, a informação financeira, este tema ainda é crítico nos dias de hoje, pelo que é de enorme relevância continuar a investigar e a criar modelos com maior precisão que os anteriores. Deste modo, e como as demonstrações financeiras das empresas nem sempre transparecem a realidade económico-financeira das mesmas, o presente estudo avalia o impacto do trabalho em auditoria financeira externa e das características da auditora na predição das insolvências.

Para tal, recorre-se a técnicas de análise de dados mais avançadas, nomeadamente *text mining* e árvores de decisão com o algoritmo CART de forma a analisar as Certidões Legais de Contas/relatórios de auditoria, entre os anos de 2016 e 2020, de uma amostra de 2.040 empresas, 1.020 ativas e 1.020 insolventes.

Os resultados obtidos permitem identificar uma relação entre o trabalho efetuado em auditoria financeira externa, bem como as características da auditora e a insolvência das empresas, prevendo-se uma percentagem de exemplos corretamente classificados de 93%. O principal contributo empírico desta investigação é gerar melhor conhecimento sobre a inviabilidade das empresas através da atividade de auditoria financeira externa recorrendo-se a novas técnicas nunca antes utilizadas em modelos preditivos.

Palavras-chave: Insolvência, Auditoria Financeira Externa, *Text Mining*, Árvores de Decisão

JEL Classification System: M10, M41

Abstract

In Portugal, since the financial crisis, the number of SME's that have gone into insolvency is quite high and really worrying due the impacts they cause on the Portuguese economy and society. Although several insolvencies predictive models have already been developed whose predictors are, essentially, financial information, this matter is still critical today, so it is extremely important to continue to investigate and create models with greater precision than the previous ones. In this way, and as the company's financial statements do not always show their true economic and financial reality, this study assesses the impact of the work in external financial audit and the characteristics of the audit firm in predicting insolvency.

To this end, more advanced data analysis techniques are used, namely text mining and decision trees with the CART algorithm to analyses the audit report between the years 2016 and 2020 of a sample of 2.020 companies, 1.020 active and 1.020 insolvent.

The results obtained indicate the existence of a relationship between the work carried out by external financial audit as well as the characteristics of the audit firm and the company's insolvencies, achieving an accuracy of 93%. Therefore, the main empirical contribution of this investigation is to bring the best knowledge about the unfeasible of companies through the external audit activity, using new techniques have never used in predictive models.

Keywords: Insolvency, External Financial Audit, Text Mining, Decision Trees

JEL Classification System: M10, M41

Índice geral

Agradecimentos	I
Sumário	III
Abstract	V
Índice geral.....	VII
Lista de Acrónimos e Siglas	XIII
1. Introdução	1
1.1. Tema e a sua relevância	1
1.2. Problema e questão de investigação	2
1.3. Objetivos e contributos	3
1.4. Abordagem metodológica	4
1.5. Estrutura.....	5
2. Revisão da Literatura.....	7
2.1. Insolvências e a sua relevância para a economia.....	7
2.1.1. Conceito.....	7
2.1.2. Relevância.....	8
2.2. O relato financeiro como espelho da situação económica e financeira das empresas ..	9
2.3. Primeiras abordagens aos modelos preditivos das insolvências	10
2.4. Metodologias posteriores às adotadas nos primeiros modelos preditivos	12
2.4.1. <i>Text mining</i>	13
2.5. Auditoria em Portugal	13
2.5.1. Enquadramento e normativo vigente em Portugal.....	13
2.5.2. Princípios e responsabilidades gerais.....	14
2.5.3. A Certidão Legal das Contas / relatório de auditoria	14
2.5.4. Inclusão de uma opinião de auditoria	15
2.5.5. Ênfases.....	17
2.5.6. Incerteza material relacionada com a continuidade.....	17
2.6. O trabalho em AFE e características da auditora como preditor das insolvências	19
3. Hipóteses de investigação e metodologia.....	23

3.1.	Hipóteses de investigação e posicionamento da investigação	23
3.2.	População e amostra	23
3.2.1.	Processo de seleção da amostra	23
3.2.2.	Caracterização da amostra	25
3.3.	Dados	27
3.3.1.	Variável dependente	27
3.3.2.	Variáveis independentes	27
3.4.	Técnicas de análise de dados	29
3.4.1.	Text <i>mining</i> para identificar os tópicos nos comentários	31
3.4.2.	Modelos preditivos com árvores de decisão.....	33
4.	Resultados e discussão	35
4.1.	Identificação dos tópicos salientes presentes nos comentários.....	35
4.1.1.	Frequência absoluta dos termos nas subamostras das empresas ativas e insolventes 35	
4.1.2.	Tópicos nos comentários.....	36
4.2.	Caracterizar o trabalho da AFE e as características da auditora	39
4.3.	Caracterizar a relação entre as dimensões de variáveis de auditoria e o <i>status</i>	41
4.4.	Modelo preditivo das insolvências e perfis associados às empresas.....	44
4.4.1.	Modelo preditivo.....	44
4.4.2.	Perfis associados às empresas ativas e insolventes	48
5.	Conclusões.....	49
5.1.	Contributos.....	50
5.2.	Limitações e pistas para investigações futuras	50
Anexo	53
Referências bibliográficas	57

Índice de Figuras

Figura 1.1 - Impacto das insolvências registadas no primeiro semestre de 2020	3
Figura 2.2 - Estrutura da Certidão Legal das Contas / relatório de auditoria	15
Figura 2.3 - Tipos de opinião de auditoria	16
Figura 2.4 - Comunicações adicionais no relatório do auditor	17
Figura 2.5 - Implicações do pressuposto da continuidade no relatório do auditor	18
Figura 3.6 - Processo experimental do text mining	32
Figura 4.7 - Nuvem de palavras dos termos mencionados nos comentários	36
Figura 4.8 - Importância das variáveis do modelo preditivo D	47

Índice de tabelas

Tabela 2.1 - Capacidade Preditiva do modelo desenvolvido por Altman (1968).....	11
Tabela 2.2 - Modelos preditivos das insolvências/falência tendo como preditor a auditoria	19
Tabela 2.3 - Preditores relacionados com auditoria utilizados nos principais estudos.....	20
Tabela 3.4 - Processo de triagem dos dados.....	24
Tabela 3.5 - Distribuição das empresas da amostra por ano.....	24
Tabela 3.6 - Critério de Classificação das PME.....	25
Tabela 3.7 - Distribuição das empresas por dimensão	25
Tabela 3.8 - Distribuição das empresas por região (NUTS II).....	26
Tabela 3.9 - Distribuição da subamostra das empresas por secção CAE e dimensão	26
Tabela 3.10 - Fórmulas utilizadas na elaboração dos rácios.....	28
Tabela 3.11 - Descrição das variáveis independentes	29
Tabela 3.12 - Técnicas de análise de dados utilizadas	31
Tabela 3.13 - Parametrização dos modelos preditivos das insolvências	34
Tabela 3.14 - Exemplo de matriz confusão	34
Tabela 4.15 - Tópicos identificados nos comentários.....	37
Tabela 4.16 - Termos correlacionados com os tópicos.....	38
Tabela 4.17 - Medidas descritivas das variáveis qualitativas do trabalho em AFE e da auditora	40
Tabela 4.18 - Medidas descritivas da variável quantitativa da dimensão do comentário	41
Tabela 4.19 - Relação entre as variáveis qualitativas da auditoria e o <i>status</i>	43
Tabela 4.20 - Relação entre a variável quantitativa da dimensão dos comentários e o <i>status</i>	44
Tabela 4.21 - Resultados dos modelos preditivos das insolvências.....	45
Tabela 4.22 - TOP 3 das variáveis com maior importância nos modelos preditivos	46
Tabela 4.23 - Distribuição do <i>status</i> previsto para as categorias das variáveis mais importantes.	47

Lista de Acrónimos e Siglas

AFE - Auditoria Financeira Externa

CAE - Classificação Portuguesa de Atividades Económicas

CART - *Classification and Regression Trees*

CIRE - Código da Insolvência e da Recuperação de Empresas

CLC - Certidão Legal das Contas

CMVM - Comissão do Mercado de Valores Mobiliários

CSC - Código das Sociedades Comerciais

DF - Demonstrações financeiras

DGPJ - Direção-Geral da Política da Justiça

EOROC - Estatuto da Ordem dos Revisores Oficiais de Contas

IA - Inteligência artificial

IAASB - International Auditing and Assurance Standards Board

INE - Instituto Nacional de Estatística

ISA - *International Standard Auditing*

LDA - Sociedades por quotas de responsabilidade limitada

MEP - Método da Equivalência Patrimonial

NUTS - Nomenclatura das Unidades Territoriais para Fins Estatísticos

PECC - Percentagem de Exemplos Corretamente Classificados

PME - Pequenas e Médias Empresas

RAFE - Relatório de Auditoria Financeira Externa

RERE - Regime Extrajudicial de Recuperação de Empresas

ROC - Revisor Oficial de Contas

S.A. - Sociedades Anónimas

SROC - Sociedade de Revisores Oficiais de Contas

1. Introdução

A insolvência das empresas é um fenómeno a que os reguladores, gestores e *stakeholders* devem dar importância acrescida. Esta introdução aborda a relevância das insolvências no contexto das Pequenas e Médias Empresas (PME) nacionais e os impactos causados na economia, bem como determinadas características da Certidão Legal de Contas (CLC) / Relatório de Auditoria Financeira Externa (RAFE) e da auditora em si, como evidência de possíveis problemas que podem levar à insolvência das empresas. Para além disto, expõe o problema e os objetivos da investigação e apresenta a abordagem metodológica desenvolvida.

1.1. Tema e a sua relevância

A insolvência das empresas é um tema bastante abordado pelos impactos económicos que causam no tecido empresarial, nas pessoas e na economia dos países. De acordo com o Código da Insolvência e da Recuperação de Empresas (CIRE), as empresas entram em processo de insolvência quando se encontram impossibilitadas de cumprir com as suas obrigações já vencidas.

Portugal, desde a crise financeira vivida entre 2008 e 2014, aumentou o número de processos de insolvência face ao período anterior à crise. Apesar de virem a diminuir no pós crise, o número de empresas que não consegue liquidar as suas obrigações já vencidas ainda é elevado e preocupante. As PME nacionais são as mais afetadas e as que mais declaram insolvência comparativamente com as grandes empresas (INE, 2018).

As PME têm enorme relevância na economia da grande maioria dos países (Altman & Sabato, 2013). Em Portugal, este grupo de empresas representa, praticamente, todo o tecido empresarial e é responsável pelo emprego da grande maioria das pessoas no ativo (Pordata, 2018).

Assim, investigadores têm apostado em estudos que permitam prever as insolvências. Os primeiros avanços na criação de modelos preditivos foram concebidos tendo como preditores os indicadores financeiros das empresas (Beaver, 1996). Contudo, a gestão dos resultados e a manipulação dos mesmos praticada pelos gestores colocam em causa a fiabilidade e credibilidade das Demonstrações Financeiras (DF), nomeadamente nas PME (Constantino, 2020). Por este motivo, e também pela necessidade de se comparar indicadores financeiros com informação não financeira, é que a análise isolada dos rácios financeiros não garante a complexidade exigida e necessária para prever as insolvências (Cheng *et al.*, 2007).

Perante a dificuldade em prever as insolvências, a informação proveniente do trabalho realizado em Auditoria Financeira Externa (AFE), começou a ocupar um papel crucial nestes modelos tendo contribuído para a criação de modelos com boas capacidades preditivas (Muñoz-Izquierdo *et al.*, 2020).

De facto, o papel do auditor é formar uma opinião sobre se as DF estão preparadas, em todos os aspetos materiais, de acordo com o referencial de relato que lhe é aplicável. Para formar a opinião, o auditor deve concluir se obteve garantia razoável de que a informação financeira da empresa está isenta de distorções materiais devido a fraude ou erro (Internacional Standard on Auditing (ISA) 700) e que garante uma imagem verdadeira e apropriada da situação económico-financeira da empresa (IAASB, 2015). Deste modo, o trabalho em AFE permite colmatar a falta de transparência das DF, e por esse motivo, contribuir para a predição das insolvências. A CLC/RAFE é o resultado final do trabalho em AFE e onde constam todas as conclusões retiradas, sendo a única forma de comunicação dos auditores com os diferentes *stakeholders* e utilizadores da informação.

Apesar do trabalho em AFE já ser utilizado como preditor das insolvências e do potencial da CLC/RAFE na predição das insolvências, pouca é a informação deste documento utilizada na criação dos modelos, como concluem Muñoz-Izquierdo *et al.* (2019). Deste modo, estes autores introduziram nos modelos preditivos, pela primeira vez, a análise do conteúdo dos comentários emitidos pelo auditor, sejam reservas ou ênfases, de forma a perceber se o tema abordado nesses comentários permite diferenciar as empresas insolventes e ativas. De facto, a utilização de determinados termos podem ser um claro sinal de alerta que permita ajudar na prevenção das insolvências (Neves, 2012). Uma das desvantagens apontadas pelos autores é a metodologia utilizada, uma vez que a identificação dos temas mencionados nos comentários é feita de forma manual por peritos em contabilidade, o que torna o processo dispendioso, demorado e sujeito a diferentes interpretações.

Assim, o tema do presente estudo é a predição das insolvências das empresas, tendo como preditores o trabalho da AFE e outras características relacionadas com a firma de auditoria. Em termos metodológicos, são utilizadas técnicas de análise de dados multivariada, nomeadamente árvores de decisão, e também técnicas alternativas à análise manual de conteúdo que assentam em Inteligência Artificial (IA), como o *text mining* na análise dos comentários emitidos pelos auditores na CLC/RAFE.

No que se refere aos estudos académicos publicados que visem desenvolver modelos preditivos de insolvência das empresas tendo como preditor a CLC/RAFE, até ao momento, os investigadores não aplicaram técnicas de *text mining* no tratamento de dados para construção desses modelos. Daí que, a aplicação desta técnica na presente investigação torne diferenciador o contributo científico ora produzido, na medida em que permite analisar os comentários dos auditores de uma grande amostra de empresas, cuja análise manual seria inviável.

1.2. Problema e questão de investigação

Nos últimos anos, o número de insolvências em Portugal é elevado face ao período anterior à crise financeira. Apesar de o número de empresas em insolvência vir a diminuir desde o período da crise, entre 2008 e 2014, a pandemia COVID-19 veio interromper esta tendência decrescente e, em 2020, o

número de empresas em insolvências voltou a aumentar ligeiramente no que diz respeito a dados do primeiro semestre de 2020 (Pordata, 2020). Estas insolvências representam uma perda do volume de negócios superior a 550 milhões de euros (figura 1.1). Outros impactos também estão associados ao aumento do número de insolvências, nomeadamente o desemprego e as dificuldades financeiras das famílias.

Figura 1.1 - Impacto das insolvências registadas no primeiro semestre de 2020



Fonte: Elaboração própria.

Para além dos impactos causados nas próprias empresas insolventes, estes efeitos emergem para os diversos *stakeholders*, tais como clientes, fornecedores, trabalhadores da empresa, investidores e outros credores, afetando a economia nacional. Os efeitos devastadores na economia dos países tornam as insolvências das PME um problema preocupante e alarmante que deve ser evitado.

Embora já tenham sido desenvolvidos alguns modelos preditivos das insolvências, este tema é crítico, pelo que é de grande relevância continuar a investigar e a criar modelos com maior precisão que os anteriores de forma a antecipar problemas de viabilidade da empresa a tempo de evitar as insolvências (Bauweraerts, 2016; Jardin, 2015).

Perante o problema mencionado, e no âmbito do tema da predição das insolvências, esta investigação tem como objetivo gerar conhecimento que permita apoiar profissionais e académicos a antecipar com maior precisão problemas de viabilidade das empresas através do trabalho realizado pela AFE e das características da auditora. Neste sentido, a presente investigação visa responder à seguinte questão de investigação: Como é que o trabalho da AFE permite prever as insolvências das empresas?

1.3. Objetivos e contributos

Tendo em conta a questão de investigação formulada anteriormente, o objetivo geral desta dissertação é identificar qual o contributo do trabalho da AFE bem como das características da auditora para a explicação das insolvências das empresas.

Para alcançar o objetivo geral proposto, são definidos quatro objetivos específicos:

- I. Identificar os tópicos salientes presentes nos comentários (reservas e ênfases) emitidos na CLC/RAFE;
- II. Caracterizar o trabalho da AFE e as características da auditora;
- III. Caracterizar a relação entre o trabalho em AFE e as características da auditora e o *status* da empresa;
- IV. Criar modelo preditivo das insolvências, identificando perfis associados às empresas insolventes e ativas.

Como resposta à questão de investigação através da concretização dos objetivos, esta investigação contribui para aumentar o conhecimento sobre modelos preditivos das insolvências e sobre o trabalho realizado pela AFE. Por outro lado, ao debruçar-se sobre o conteúdo dos comentários dos auditores na CLC/RAFE, este estudo vem colmatar a lacuna no conhecimento na disciplina de contabilidade e auditoria referente ao desenvolvimento de modelos preditivos de insolvências que apliquem técnicas de *text mining* para a criação destes modelos.

Adicionalmente, sendo identificados perfis de CLC/RAFE associados às insolvências, os gestores têm oportunidades de, atempadamente, delinear estratégias que possam evitar as insolvências. Por outro lado, os utilizadores da informação, tais como, investidores e credores passam a dar uma maior relevância ao trabalho realizado em AFE de forma a tomar melhores decisões de investimento e financiamento.

Relativamente ao contributo científico para a contabilidade, destaca-se a utilização com sucesso de técnicas mais avançadas de *text mining* na análise da CLC/RAFE, e também a relevância do trabalho realizado em AFE, que quando efetuado de forma responsável, ajuda na predição de problemas de inviabilidade das empresas.

1.4. Abordagem metodológica

Atendendo à questão de investigação e aos objetivos propostos, esta investigação adota um *mix* entre a metodologia qualitativa e quantitativa. Este estudo tem por base uma amostra de 2.040 empresas portuguesas, 1.020 com *status* insolvente e 1.020 com *status* ativa. Todos os dados são retirados da base de dados SABI, uma vez que esta base de dados disponibiliza toda a informação do trabalho em AFE e das características da auditora necessária para a concretização dos objetivos.

De forma a garantir a qualidade do estudo, recorre-se a diferentes técnicas já utilizadas e que evidenciaram bons resultados em outras investigações. Para além destas, também são utilizadas outras técnicas mais recentes de IA e pouco utilizadas na literatura. Inicialmente recorre-se a técnicas mais inovadoras ao nível do *text mining* para criação dos tópicos mencionados nos comentários emitidos pelo auditor. De seguida, utilizam-se técnicas de estatística uni e bivariada para analisar se as variáveis abordadas no estudo estão estatisticamente relacionadas com o *status* da empresa. Por fim,

e tendo em conta os objetivos propostos, recorre-se a árvores de classificação com o algoritmo *Classification and Regression Trees* (CART) para construção do modelo preditivo.

1.5. Estrutura

No que concerne à estrutura da dissertação e, de forma a compreender se o trabalho efetuado pela AFE é um bom preditor das insolvências, esta organiza-se em cinco capítulos.

No primeiro capítulo, este, são abordados o tema, a relevância do mesmo e a problemática associada. Para além disso, é apresentada a questão de investigação, os objetivos propostos e a abordagem metodológica a utilizar para a concretização desses mesmos objetivos.

Seguidamente, o segundo capítulo compreende o enquadramento teórico do tema. A revisão de literatura aborda a problemática das insolvências e os seus impactos, os modelos preditivos já desenvolvidos, bem como um enquadramento da AFE e a sua capacidade de prever as insolvências.

O terceiro capítulo, consubstancia-se na metodologia adotada, onde é demonstrado todo o procedimento de recolha e análise de dados, após terem sido definidas as hipóteses de investigação.

Os resultados e a sua discussão são apresentados no capítulo quatro.

Por fim, as conclusões a retirar do estudo, os contributos, as limitações e as pistas futuras para outras investigações estão presentes no capítulo cinco.

2. Revisão da Literatura

Este capítulo, em primeiro lugar, aborda a insolvências das PME e a sua relevância para a economia dos países, nomeadamente de Portugal, bem como a forma com que o relato financeiro retrata a realidade das empresas. Em segundo lugar, uma síntese dos preditores e das metodologias usadas nos modelos anteriormente concebidos, destacando a metodologia do *text mining* como uma potencial ferramenta de inteligência artificial na predição das insolvências. Em terceiro lugar, são apresentadas as responsabilidades do auditor, o conteúdo e a estrutura da CLC/RAFE, e também os efeitos destes mesmos relatórios nas empresas. Por último, são analisados modelos preditivos das insolvências de estudos anteriores que têm como preditores o trabalho realizado em AFE e outros atributos associados com as empresas auditadas e firmas de auditoria.

2.1. Insolvências e a sua relevância para a economia

2.1.1. Conceito

É considerado em situação de insolvência o devedor que se encontre impossibilitado de cumprir com as suas obrigações já vencidas, e também as pessoas coletivas e os patrimónios autónomos cujo seu passivo seja manifestamente superior ao ativo, avaliados segundo as normas contabilísticas aplicáveis, e encontra-se em situação económica difícil a empresa que enfrentar dificuldade séria para cumprir pontualmente as suas obrigações, designadamente por ter falta de liquidez ou por não conseguir obter crédito, de acordo com o CIRE.

Nos termos do artigo 1º do mesmo código, o processo de insolvência é um processo de execução universal, que tem como finalidade a satisfação dos credores, pela forma prevista, num plano de insolvência baseado na recuperação da empresa compreendida na massa insolvente ou, quando tal não se afigure possível, na liquidação do património do devedor insolvente e a repartição pelos credores.

A insolvência é um processo jurídico que se divide em duas vertentes, a jurídica e a económica. Do ponto de vista jurídico, a legislação tem como principal objetivo atenuar os efeitos gerados pela declaração da insolvência de um agente económico em toda a esfera económica, pois os efeitos não se refletem apenas na esfera jurídica do devedor. O processo de insolvência pode ocasionar impactos irreversíveis para os diversos *stakeholders*, tais como credores, investidores, fornecedores, clientes, colaboradores, tanto ao nível económico como psicológico. Com a declaração de insolvência e respetiva liquidação do património, os credores não conseguem reaver grande parte dos seus créditos, os investidores perdem os seus investimentos e os colaboradores ficam desempregados, dificultando o sustento das famílias (Muñoz-Izquierdo *et al.*, 2020).

A vertente económica envolve, entre outras, a noção de crédito. A perda de crédito de um agente económico provocada pela falta de confiança dos credores coloca em causa o equilíbrio no círculo das relações económicas, levando, por consequência, outros agentes económicos a passarem por dificuldades financeiras.

De acordo com o nº2 do Artigo 1º do CIRE, estando em situação económica difícil ou em situação de insolvência meramente iminente, isto é, havendo uma convicção de que a empresa, praticamente, se encontre impossibilitada de cumprir com as suas obrigações, esta pode requerer ao tribunal a instauração de Processo Especial de Revitalização (PER). Tendo em conta o Artigo 17º do CIRE, o PER tem carácter urgente e destina-se a permitir à empresa, que ainda seja suscetível de recuperação, estabelecer negociações com os respetivos credores de modo a estabelecer com estes acordos conducentes à sua revitalização. Por outras palavras, as empresas nacionais que se encontram em situação económica difícil podem recorrer ao PER com o intuito de recuperar o seu estado económico antes de serem declaradas insolventes.

2.1.2. Relevância

Os processos de insolvência, em Portugal, dispararam durante a crise financeira vivida entre 2008 e 2014, e desde então os números anuais de processos de insolvência têm sido bastante elevados face a outros países, de acordo com a Direção-Geral da Política da Justiça [DGPJ] (2020). Em Portugal, as empresas são muito dependentes do crédito bancário tanto para financiar o seu passivo corrente como os seus investimentos a longo prazo. Com as dificuldades vividas pelas instituições financeiras dos países europeus durante o período de recessão, as empresas viram os seus pedidos de crédito serem rejeitados, levando ao insucesso empresarial das mesmas. As microempresas continuam a representar a maioria dos casos de insolvência com uma quota de 65%, sendo o setor dos serviços o mais afetado (COSEC, 2020).

No primeiro semestre de 2020 o número de insolvências ascendia a 1313, mais 2% face ao período homólogo, o que representa um impacto de grande dimensão na economia do país (COSEC, 2020). Segundo a DGPJ (2020), a taxa de recuperação de créditos, isto é, a proporção do montante de créditos pagos face ao montante de créditos reconhecidos nos processos de insolvência foi de 13,7%, pelo que 86,3% dos créditos não foram pagos, tendo impactos irreversíveis para os *stakeholders*. Neste período, deram também entrada no tribunal (DGPJ, 2020). Segundo estes dados, a maioria das empresas que enfrenta dificuldades financeiras não consegue ser reestruturada.

Um estudo realizado em Portugal demonstrou que as principais causas que levam as empresas a passarem dificuldades são a redução da procura, falta de liquidez e dificuldade de obtenção de crédito (Assis, 2018). Já em termos gerais, Lukason & Hoffman (2014), mostram que as insolvências estão

associadas a problemas internos, tais como, incompetência da gestão, fraude e a problemas externos, nomeadamente, aumento da concorrência, crises económicas e falência de importantes fornecedores e clientes.

Outra investigação evidencia que, perdas acumuladas, investimentos de curto e longo prazo e processos judiciais são os problemas que estão mais associados ao insucesso empresarial (Muñoz-Izquierdo *et al.*, 2020).

As PME representam 99,9% do tecido empresarial português e a sua variação absoluta manteve-se nula e constante durante os últimos anos (Pordata, 2018). Segundo esta mesma fonte, em 2018 as PME foram responsáveis por 60% da riqueza (Valor Acrescentado Bruto) de Portugal, e geraram 56% do volume de negócios total das empresas nacionais. Estas empresas são responsáveis por, aproximadamente, dois terços de todos os colaboradores ativos nos países da OCDE de acordo com Altman *et al.* 2010, sendo que em 2018 empregavam, em Portugal, 3.230.077 indivíduos, o que correspondia a 78% das pessoas no ativo (Pordata, 2018). Por estes motivos, as PME têm um papel acrescido e importante na manutenção e criação de postos de trabalho.

2.2. O relato financeiro como espelho da situação económica e financeira das empresas

As DF fazem parte do processo de relato financeiro. Um conjunto de DF engloba, uma demonstração de resultados, uma demonstração das alterações da posição financeira (Balanço), uma demonstração de fluxos de caixa e uma parte de notas explicativas. O objetivos das DF é proporcionar informação relevante e fiável acerca do desempenho económico e financeiro da empresa a um vasto leque de utentes na tomada de decisão (IAASB, 2015).

De acordo com Altman *et al.* 2015, a confiabilidade e a viabilidade da informação financeira das PME é baixa devido à gestão dos resultados praticada pelas empresas. Para além disto, a maioria dos dados financeiros destas empresas não estão disponíveis de forma pública, tal como acontece com as grandes empresas cotadas (Altman *et al.* 2010).

Assim, é mais fácil identificar os sintomas das dificuldades financeiras nas grandes empresas mais cedo do que nas PME, por exemplo, através de indicadores relacionados com os mercados de capitais que as PME não têm porque não estão cotadas na bolsa de valores (Altman *et al.* 2015).

Apesar dos problemas mencionados relativos à informação financeira, os rácios/indicadores financeiros são utilizados na predição das insolvências nas PME, alcançando bons resultados (Sánchez *et al.*, 2013; Altman, 1968). Estes indicadores agrupam-se em diferentes categorias e traduzem relações entre contas do balanço, da demonstração de resultados, da demonstração dos fluxos de caixa ou de outras grandezas económico-financeiras e devem ser contruídos atendendo aos objetivos da análise (Neves, 2012).

2.3. Primeiras abordagens aos modelos preditivos das insolvências

Altman (1968) e Beaver (1966), são os pioneiros a criar modelos que pudessem antecipar as insolvências tendo como preditores conjuntos de rácios financeiros.

Beaver (1966), utiliza o método da análise univariada para prever o fracasso de 79 empresas americanas. Este é o primeiro estudo na predição do fracasso empresarial e funciona como um *benchmark* para investigações futuras. Para cada conjunto das DF foram escolhidos e comparados 30 rácios diferentes seguindo o critério da popularidade, nos 5 anos anteriores ao fracasso da empresa. O autor utiliza determinados critérios de forma a excluir a correlação entre os rácios, obtendo um conjunto final de 6 rácios financeiros:

- $Cash\ Flow\ to\ Total\ Debt = \frac{Cash\ Flow}{Total\ Debt}$.
- $Net\ Income\ to\ Total\ Assets = \frac{Net\ Income}{Total\ Assets}$.
- $Total\ Debt\ to\ Total\ Assets = \frac{Total\ Debt}{Total\ Assets}$.
- $Working\ Capital\ to\ Total\ Assets = \frac{Working\ Capital}{Total\ Assets}$.
- $Current\ Ratio = \frac{Current\ Assets}{Current\ Liabilities}$.
- *No-credit interval (defensive assets minus current liabilities to fund expenditures for operations)*

De acordo com este autor, o método estatístico utilizado apresenta uma complexidade bastante reduzida uma vez que examina a capacidade preditiva de cada rácio de forma individual, não estabelecendo relações entre os diferentes rácios. Outra das limitações apontadas pelo autor é o facto de a amostra ser constituída por empresas com ativos elevados, sendo que empresas maiores não têm tanta probabilidade de entrar em falência como as de pequena dimensão.

Tendo em conta os resultados obtidos, os primeiro dois rácios apresentam uma maior capacidade preditiva face aos restantes rácios, prevendo individualmente 87% do fracasso empresarial no ano anterior à falência. Dificuldades de obtenção de cash flow positivo nas PME podem estar relacionadas com a estrutura de funcionamento da empresa. Os rácios de funcionamento ajudam a explicar os impactos financeiros da gestão ao nível do ciclo de exploração, como por exemplo, o Prazo Médio de Pagamentos (PMP), o Prazo Médio de Recebimentos (PMR) e a duração média das existências em armazém. As PME tendem a estender o PMR para que possam captar mais clientes, contudo esta medida pode incapacitar as empresas de liquidar as suas dívidas de curto prazo (Altman *et al.*, 2010). Os rácios *Working Capital to Total Assets* e *Current Ratio* foram os piores rácios na predição das insolvências tendo em conta a sua capacidade preditiva e a deterioração à medida que se afastam do ano em que a empresa fracassou.

Pouco tempo depois, Altman (1968), utiliza o Modelo da Análise Discriminante Multivariada para prever a falência de 66 empresas industriais. Este modelo consiste na determinação de uma função, que pode ser usada para distinguir dois ou mais grupos através de uma combinação linear entre as diferentes variáveis. Esta combinação é a principal diferença entre a Análise univariada e multivariada. Por estes motivos, esta abordagem metodológica ainda é muito conceituada nos dias de hoje na predição das insolvências (Muñoz-Izquierdo *et al.*, 2020).

Os 22 rácios potencialmente relevantes em investigações anteriores recolhidos por Altman foram reduzidos de forma a evitar a correlação entre as variáveis. Foram então, selecionados cinco variáveis (rácios) que melhor discriminam os dois grupos *a priori* estabelecidos. Os rácios de rendibilidade, de atividade e de endividamento, são os que melhor discriminam os grupos, respetivamente, o que vai ao encontro do estudo de Peurseem & Chan (2014) que evidenciam que os rácios de rendibilidade são bons preditores das insolvências, uma vez que empresas em falência tendem a apresentar baixas rendibilidades face às empresas ativas. Em baixo estão identificados os 5 rácios utilizados no modelo:

- X1 (Rácio de Fundo de Maneio) = $\frac{\text{Fundo de maneio}}{\text{Ativo Total}}$;
- X2 (Rácio de Solvência) = $\frac{\text{Resultados Retidos}}{\text{Ativo Total}}$;
- X3 (Rácio de Rendibilidade) = $\frac{\text{EBIT}}{\text{Ativo Total}}$;
- X4 (Rácio de Endividamento) = $\frac{\text{Valor de Mercado}}{\text{Passivo Total}}$;
- X5 (Rácio de atividade) = $\frac{\text{Vendas}}{\text{Ativo Total}}$

O autor prova que o modelo da Análise Discriminante Multivariada é extremamente eficaz na predição das insolvências nos dois primeiros anos antes do fracasso empresarial, sendo que no primeiro ano apresenta uma Percentagem de Exemplos Corretamente Classificados (PECC) DE 95%. Entre o segundo e o terceiro ano, a capacidade preditiva do modelo diminui significativamente, pelo que o autor não consegue garantir a confiabilidade do modelo no 3º, 4º e 5º ano, como mostra o quadro abaixo.

Tabela 2.1- Capacidade Preditiva do modelo desenvolvido por Altman (1968)

Anos anteriores à falência	Dimensão da amostra	Capacidade Preditiva (%)
1	33	95%
2	32	72%
3	29	48%
4	28	29%
5	25	36%

Fonte: Adaptado de Altman (1968)

Apesar de ser um estudo antigo e um dos primeiros neste campo de investigação, é dos estudos com maior capacidade preditiva das insolvências tendo em conta o Método da Análise Discriminante Múltipla.

Contudo, de acordo com outros autores, a utilização de rácios financeiros sem o complementar da informação não financeira não produz informação credível para a análise da insolvência das empresas (Cheng *et al.*, 2007). A falta de confiabilidade da informação e a falta de informação nas PME levam à criação de novos modelos preditivos com diferentes preditores e metodologias.

Após as abordagens iniciais, alguns autores começaram a utilizar a informação não financeira nos seus modelos, nomeadamente a informação contida na CLC/RAFE proveniente do trabalho realizado pela AFE (Altman *et al.*, 2010, 2015) e a utilizar outras metodologias mais avançadas que permitam uma melhor predição das insolvências.

2.4. Metodologias posteriores às adotadas nos primeiros modelos preditivos

Após o aparecimento dos primeiros estudos que abordam os modelos preditivos, têm surgido novas metodologias com maior grau de complexidade de forma a aumentar a eficácia dos mesmos.

O modelo da Regressão Logística surge, pela primeira vez, numa investigação de Ohlson (1980). Uma das vantagens apontadas a este modelo é o facto de permitir estimar a probabilidade de ocorrência de um acontecimento. Esta metodologia é recomendada para situações em que a variável dependente é de natureza binária e é uma das metodologias mais abordadas nestes estudos dos modelos preditivos das insolvências. Por outro lado, a reduzida precisão na previsão das insolvências e a difícil interpretação são as principais desvantagens apontadas a este método (Constantino, 2020).

Mais recentemente têm vindo a ser utilizadas metodologias de IA como por exemplo, as árvores de decisão construídas através dos algoritmos CART e CHAID (*Chi-squared Automatic Interaction Detection*). Estes métodos são caracterizados pela fácil interpretação e obtenção de conhecimento, boa lidação com *outliers* e não respostas, bem como pela maior precisão na predição das insolvências (Abrantes, 2020).

Para além destas têm emergido novas abordagens metodológicas que acompanham toda a evolução digital que as empresas enfrentam nos dias de hoje, mas que ainda não são utilizadas neste tema da predição das insolvências, nomeadamente o *text mining* que tem um potencial enorme, na medida em que é um processo de análise assistida por computador e que pode ser aplicado em bases de dados massivas revelando-se vantajoso na componente de tempo e consumo de recursos na fase de extração e análise de dados (Zorio-Grima & Carmona, 2019).

2.4.1. Text mining

Os processos de *text mining* têm vindo a ter cada vez mais aplicabilidade na medida em que permitem transformar texto não estruturado em formato estruturado a fim de ser possível identificar padrões e conhecimento (Santos *et al.*, 2020). Segundo a *Internacional Business Machines* [IBM] (2020), cerca de 80% dos dados globais estão num formato não estruturado, pelo que estas técnicas podem ser ferramentas bastante valiosas para as organizações, permitindo uma melhor tomada de decisão.

Tendo em conta a IBM (2020), estas ferramentas têm impactado o modo como as indústrias trabalham, nomeadamente, na área da saúde, serviços ao cliente e na análise de risco. Na análise de risco, a utilização desta metodologia fornece *insights* relevantes sobre as tendências da indústria e dos mercados financeiros através da extração da informação presente nos relatórios anuais das empresas. Isto é, particularmente, muito valioso para os diferentes credores e investidores, nomeadamente para as instituições financeiras na concessão de crédito às empresas.

Apesar de ainda não ter sido utilizada no tema da predição das insolvências, este método já é abordado em estudos relacionados com o trabalho da AFE, nomeadamente, na análise da linguagem utilizada nas CLC/RAFE. De acordo com a literatura, a linguagem usada pelos auditores varia entre as diferentes firmas de auditoria e a jurisdição de cada país, sendo que os relatórios no Reino Unido têm tendência a ser mais longos e mais detalhados face aos países europeus (Zorio-Grima & Carmona, 2019).

2.5. Auditoria em Portugal

2.5.1. Enquadramento e normativo vigente em Portugal

Durante a crise financeira, o trabalho de auditoria foi muito criticado por não alertar os potenciais utilizadores da informação para possíveis insucessos empresariais que acabaram por suceder. Estes acontecimentos culminaram na perda de confiança dos utilizadores da informação no trabalho dos auditores (Freire, 2016).

Para restaurar a confiança, em 2016 foi implementada a reforma do enquadramento jurídico da atividade de revisão legal das contas em Portugal pela Lei n.º 140/2015, de 7 de setembro. Por um lado, as necessidades dos *stakeholders* em querer mais informações sobre os riscos das empresas bem como, por outro, o facto da CLC/RAFE não transparecer todo o conjunto de julgamentos e procedimentos desenvolvidos pelo auditor sobre áreas críticas do negócio que lhe permitiam emitir uma opinião, levaram à necessidade de criação de uma nova CLC/RAFE (Mazars, 2016).

Para além disto, este novo Estatuto da Ordem dos Revisores Oficiais de contas engloba a adoção das Normas Internacionais de Auditoria e também atribui à Comissão do Mercado de Valores

Imobiliários (CMVM) a supervisão pública de Revisores Oficiais de Contas (ROC) e de Sociedades de Revisores Oficiais de Contas (SROC).

A alteração na forma de comunicação e a prestação de informações por parte dos auditores são a mais importante revolução no setor da auditoria, através dos novos conteúdos definidos para a CLC/RAFE que a reforma prevê (Freire, 2016).

2.5.2. Princípios e responsabilidades gerais

As DF são um input bastante utilizado pelos investidores e restantes *stakeholders* (Chen *et al.*, 2016). De acordo com o *International Auditing and Assurance Standards Board* (IAASB), a finalidade de uma auditoria é aumentar o grau de confiança dos destinatários da informação financeira. O auditor deve expressar a sua opinião, através da CLC/RAFE, se as DF estão preparadas, em todos os aspetos materiais, de acordo com um referencial de relato que lhe é aplicável, de modo a garantir que dão uma imagem verdadeira e apropriada da realidade da empresa (Chen *et al.*, 2005; DeFond & Zhang, 2014; Sánchez *et al.*, 2013).

A reforma de auditoria veio enfatizar o papel que o auditor tem no processo de credibilização do relato financeiro e na qualidade da informação financeira das empresas. Segundo Lopes (2019), é da competência do Revisor Oficial de Contas:

- Identificar e avaliar os riscos de distorção material, executando procedimentos de auditoria adequados.
- Obter prova de auditoria suficiente e apropriada para a emissão da opinião e, assim, reduzir o risco de auditoria;
- Compreender o SCI;
- Analisar a adequação das políticas contabilísticas adotadas;
- Concluir sobre a apropriação do uso, pelo órgão de gestão, do pressuposto da continuidade;
- Avaliar a apresentação, o conteúdo e a estrutura das DF;
- Obter uma segurança razoável sobre se as DF estão como um todo estão isentas de distorções materiais devido a fraude ou erro e emitir a CLC/RAFE com uma opinião.

2.5.3. A Certidão Legal das Contas / relatório de auditoria

A CLC/RAFE são o resultado final do processo de auditoria. A CLC é o produto resultante da auditoria às contas que decorram de um dever legal ou estatutário, enquanto o RAFE provém do processo de auditoria que não resulta da lei, mas sim que foi pedido de forma voluntária, de acordo com o nº1 e 9 do Artº45 do Estatuto da Ordem dos Revisores Oficiais de Contas (EOROC). A CLC/RAFE, resultado do

trabalho realizado pela AFE, são a única forma direta de comunicação do auditor com os diferentes *stakeholders* e potenciais utilizadores da informação (DeFond & Zhang, 2014).

Na figura abaixo, de acordo com o nº2 do Art45º do EOROC, podemos verificar a estrutura de uma CLC/RAFE a partir da reforma da auditoria, tal como, as normas de referência às secções apresentadas.

Figura 2.2- Estrutura da Certidão Legal das Contas / relatório de auditoria

<u>Certidão Legal das Contas / Relatório de Auditoria</u>	ISA's
Identificação da entidade cujas contas foram objeto de revisão legal, especificando as contas (consolidadas ou individuais) e a data de relato.	
Descrição do âmbito da revisão legal das contas, que deve indicar, no mínimo as normas de auditoria segundo as quais foi realizada	
Opinião	ISA 700, 705
Bases para Opinião	ISA 700, 705
Ênfases*	ISA 706
Incerteza material relacionada com o pressuposto da continuidade*	ISA 570
Outras matérias*	ISA 706
Responsabilidades do órgão de gestão e fiscalização pelas Demonstrações financeiras	ISA 700
Responsabilidade do auditor pela Auditoria às Demonstrações Financeiras	ISA 700
Outras responsabilidades de relato (Sobre o relatório da gestão)	ISA 700

*Se necessário / Opcional

Fonte: Elaboração própria

2.5.4. Inclusão de uma opinião de auditoria

Segundo o nº2 do Artº45 do EOROC, o auditor deve incluir uma opinião de auditoria que pode ser modificada/qualificada ou não modificada/não qualificada.

De acordo com a ISA 700, estamos perante uma opinião não qualificada quando o auditor conclui que as DF estão, em todos os aspetos materiais, de acordo com o normativo contabilístico que lhe é aplicável. Neste caso, o auditor emite uma opinião sem reservas nem ênfases, ou com ênfases, em que estas ênfases não afetam a opinião.

Quando o auditor concluir que as DF apresentam reservas, reservas ou ênfases, existe a emissão de uma opinião adversa ou escusa de opinião, está-se na presença de uma opinião qualificada (Chen

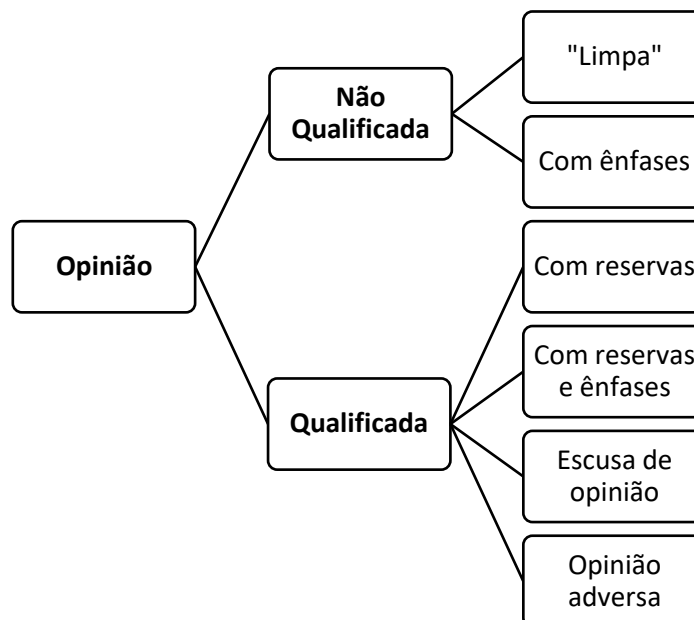
et al., 2005) (tabela 2.3). Neste caso, o parágrafo das “bases para opinião” passa a designar-se bases para opinião com reservas ou bases para opinião adversa, conforme o caso.

Quando a opinião é qualificada, o auditor deve identificar no parágrafo relativo às bases para a opinião, todas as circunstâncias que o levaram a não poder concluir que as demonstrações dão uma imagem verdadeira e apropriada da empresa auditada.

De acordo com a ISA 705, a decisão sobre qual o tipo de opinião depende da natureza da matéria que lhe dá origem, por outras palavras, se as DF estão efetivamente distorcidas (desacordo) ou no caso do auditor ser incapaz de obter prova de auditoria suficiente e apropriada (limitação de âmbito) concluir que as distorções por detetar, caso existam, sejam materiais.

A opinião do auditor é usada pelos utilizadores da informação para auxiliar nas suas decisões (Tahinakis & Samarinas, 2016), por exemplo na conceção de crédito às empresas por parte das instituições financeiras (Chen et al., 2016). Empresas com opiniões qualificadas têm maiores dificuldades, segundo estes autores, na obtenção de crédito. Por outro lado, outros autores concluem que a opinião do auditor tem efeitos insignificantes no mercado, isto é, os potenciais investidores e restantes *stakeholders* não são influenciados nas suas decisões pela opinião do auditor (Dodd et al., 1984).

Figura 2.3- Tipos de opinião de auditoria

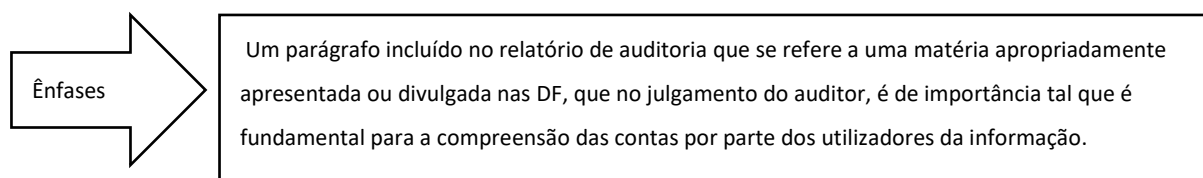


Fonte: Adaptado de Lopes (2019, pp 51).

2.5.5. Ênfases

A ISA 706, aborda as comunicações adicionais no relatório do auditor quando este considera necessário chamar à atenção dos potenciais utilizadores da informação para uma determinada matéria que, no seu julgamento, é fundamental para uma melhor compreensão das DF.

Figura 2.4 - Comunicações adicionais no relatório do auditor



Fonte: Elaboração própria

Importa referir que, de acordo com esta norma, se o auditor apresentar um parágrafo relativo a ênfases ou outras matérias, deve indicar de forma expressa que a opinião do auditor não é modificada a respeito da matéria enfatizada.

Alguns exemplos de circunstâncias que poderão determinar a inclusão de um parágrafo de ênfases no relatório de auditoria são, uma incerteza relacionada com o desfecho futuro de litígios ou ações regulatórias excepcionais, um acontecimento subsequente significativo entre o fecho das contas e a data do relatório de auditoria, uma catástrofe que tenha tido, ou continue a ter um efeito significativo na posição financeira da entidade (Lopes, 2019).

Os parágrafos de ênfases são importantes para a compreensão das DF das empresas e do trabalho realizado pela AFE, pelo que a sua análise é bastante relevante e não deve ser desconsiderada pelos utilizadores da informação pelo facto de não qualificarem a opinião do auditor (Muñoz-Izquierdo *et al.*, 2017).

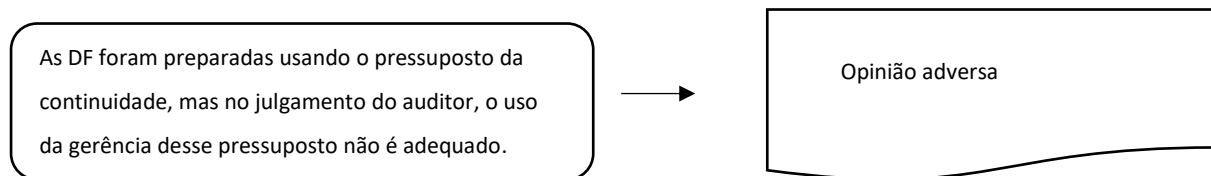
2.5.6. Incerteza material relacionada com a continuidade

Segundo a ISA 570, os ativos e passivos são registados nas DF no pressuposto de que a entidade será capaz de realizar os seus ativos e liquidar os seus passivos no âmbito normal da sua atividade. A responsabilidade do auditor é obter prova de auditoria suficiente e apropriada acerca da adequação do uso pela gerência do pressuposto de continuidade na preparação e apresentação das DF e, concluir, sobre a capacidade de a entidade prosseguir em continuidade.

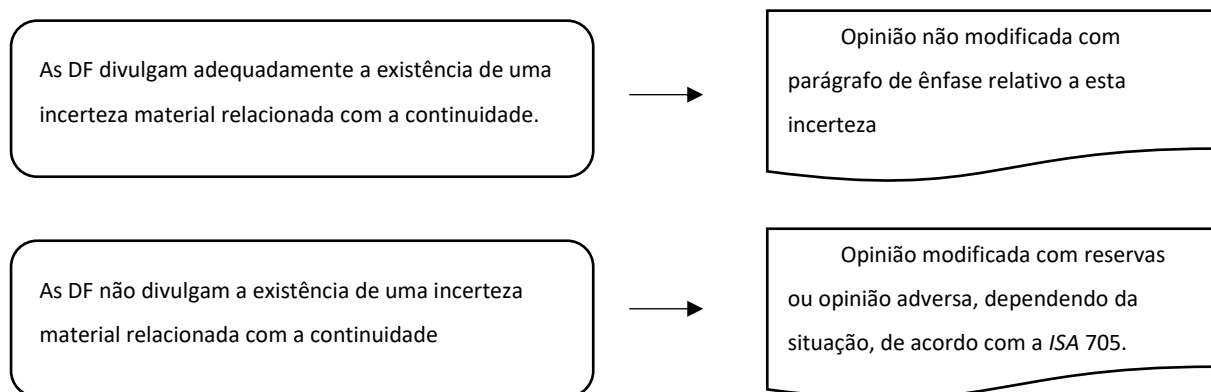
Esta ISA aborda as implicações que as conclusões do auditor têm no seu relatório, tal como ilustra a figura abaixo.

Figura 2.5 - Implicações do pressuposto da continuidade no relatório do auditor

O uso do pressuposto da continuidade não é apropriado:



O uso do pressuposto da continuidade é apropriado, mas existe uma incerteza material:



Fonte: Elaboração própria

Na primeira situação, segundo o parágrafo 21 da ISA 570, o auditor emite uma opinião adversa, independentemente das DF incluírem ou não uma divulgação sobre o uso não apropriado do pressuposto da continuidade pela gerência. A empresa poderá, neste caso, ser obrigada a preparar as suas DF tendo em conta a base de liquidação.

A decisão de emitir um parecer com uma violação do pressuposto de continuidade é sensível a pressões económicas tanto por parte do cliente, como a nível da reputação da empresa de auditoria (Carson *et al.*, 2013), sendo que este pressuposto está muito associado com os erros tipo I e tipo II (Berglund, 2020). Por um lado, se as auditoras emitirem um parecer em que colocam em causa a continuidade da empresa e o cliente acabar por não entrar em insolvência ou falência, a auditora incorre num erro tipo I. Se, por outro lado, o auditor emitir uma opinião “limpa” e a empresa, num período de 12 meses, parar as suas operações, o auditor incorre num erro tipo II.

Menon & Williams (2010) evidenciam que uma opinião qualificada relacionada com o pressuposto da continuidade pode incapacitar a empresa de continuar a operar pelos impactos causados ao nível do financiamento (DeFond & Zhang, 2014; Chen & Church, 1996).

2.6. O trabalho em AFE e características da auditora como preditor das insolvências

O objetivo da auditoria é garantir a credibilidade e fiabilidade das DF e emitir uma CLC/RAFE que expressa todo o trabalho desenvolvido. Deste modo, é razoável prever que a CLC/RAFE tenha informações relevantes para prever a insolvência das empresas (Muñoz-Izquierdo *et al.*, 2017).

A tabela 2.2 evidencia os principais estudos observados que abordam o trabalho efetuado pela AFE e outras características da auditora na predição das insolvências. Da análise da tabela, pode verificar-se que a regressão logística é a metodologia mais abordada pelos investigadores, sendo que apenas um dos estudos já utiliza técnicas de IA como as árvores de decisão.

Para além disto, todos os estudos, á exceção do estudo 3 que utiliza uma amostra de 50 empresas cotadas, utilizam como amostra as PME, uma vez que são empresas mais suscetíveis à insolvência. Note-se também, que os autores tendem a usar uma amostra mais reduzida na criação dos modelos.

Tabela 2.2 - Modelos preditivos das insolvências/falência tendo como preditor a auditoria

ID	Estudo	Amostra	Período de análise	Base de dados	Metodologia
1	Altman et al. (2010)	Reino Unido/PME privadas/5749148 totais empresas - 66833 com insucesso	2000-2007	-	Análise multivariada (Regressão Logística)
2	Sánchez et al. (2013)	Espanha/PME não financeiras/98 total da amostra com todas em dificuldades financeiras	1998-2008	SABI; Comercial Registry	Análise multivariada (Regressão Logística)
3	Van Peurseem & Chan (2014)	Nova Zelândia/Empresas cotadas/50 total da amostra - 25 falidas	2001-2010	NZX Company Research; New Zealand Companies Office websites	Análise univariada e multivariada
4	Muñoz-Izquierdo et al. (2019)	Espanha/PME/808 empresas - 404 com insucesso	2004-2014	BVD Data Base	Árvore de Decisão
5	Muñoz-Izquierdo et al. (2020)	Espanha/PME/808 empresas - 404 com insucesso	2004-2014	BVD Data Base	Análise univariada e multivariada (Regressão Logística)

Fonte: Elaboração Própria

As variáveis relacionadas com o trabalho efetuado pela AFE, nomeadamente o tipo de opinião, o número de comentários emitidos e o conteúdo dos comentários (reservas e ênfases) bem como outras

variáveis relacionadas com características da auditora, como por exemplo, a mudança e dimensão da mesma, são algumas das mais importantes utilizadas nos modelos preditivos das insolvências (tabela 2.3).

Da análise da tabela 2.3, podemos observar que o tipo de opinião é a variável que está presente em todas as investigações. De facto, o tipo de opinião aumenta a capacidade de predizer a insolvência das empresas e é uma variável diferenciadora entre as empresas ativas e insolventes, mostrando que a qualificação da opinião é um claro sinal de alerta das empresas que enfrentam dificuldades financeiras (Peurseem & Chan, 2014; Sánchez *et al.*, 2013). Adicionalmente, verificou-se que a qualificação da opinião é estatisticamente significativa para um período temporal de três anos (Altman *et al.*, 2015).

Contudo, relatórios com opiniões não qualificadas mas com parágrafos de ênfases, também estão positivamente relacionados com empresas que declaram insolvência ou falência (Muñoz-Izquierdo *et al.*, 2019). Esta investigação vai ao encontro do estudo de Muñoz-Izquierdo *et al.* (2017) que conclui que 87% das empresas que entraram em insolvência tinham comentários dos auditores, em alguns casos apenas ênfases, mas que também são relevantes para as análises.

Tabela 2.3 – Preditores de auditoria utilizados em cada um dos estudos

ID	Opinião de auditoria	Nº de comentários divulgados	Dimensão da auditora	Mudança da auditora	Conteúdo dos comentários
1	✓				
2	✓			✓	
3	✓		✓		
4	✓	✓			
5	✓	✓	✓		✓

Fonte: Elaboração própria

Ainda no que diz respeito ao trabalho realizado pela AFE importa referir que, a análise do conteúdo dos comentários emitidos pelo auditor é uma inovação no âmbito dos modelos preditivos das insolvências, tendo sido pela primeira vez incluída nestes modelos (Muñoz-Izquierdo *et al.*, 2020). De facto, estes autores evidenciam que o conteúdo dos comentários é um bom preditor das insolvências e que os auditores tendem em mencionar nesses mesmos comentários temas relacionados com gastos do período, investimentos e incertezas quanto à continuidade das operações. Importa referir que esta análise de comentários é feita manualmente por peritos de contabilidade.

Por outro lado, relativamente às características da firma responsável pela revisão das contas, existe uma tendência das empresas que entram em insolvência serem auditadas por firmas de

auditoria mais pequenas, enquanto as ativas por firmas de maior dimensão, as designadas por *Big 4* (Muñoz-Izquierdo *et al.*, 2020).

As *Big 4*, segundo Boone *et al.* (2010), conseguem manter um maior nível de independência face aos seus clientes, uma vez que o custo associado a erros é muito maior para as grandes empresas do setor (Geiger & Rama, 2016). Os auditores das *Big 4* ao detetarem distorções espelham-nas, muitas das vezes, nos seus relatórios de forma a alertar os utilizadores da informação e não cedem a pressões por parte da gestão, o que não acontece com tanta frequência nas empresas não *Big 4* (Berglund, 2020). Empresas com opiniões qualificadas resultantes de distorções detetadas por uma *Big 4* terão mais dificuldades na obtenção de financiamento (Schneider, 2018), pois os potenciais utilizadores da informação vêm as *Big 4* como as melhores empresas do setor na prestação de serviços de auditoria (Menon & Williams, 2010).

Outra das características abordadas na literatura é a mudança de auditora, que de acordo com alguns estudos é um bom preditor das insolvências, uma vez que as empresas tendem a mudar de auditor nos três anos anteriores à entrada em insolvência (Cheng *et al.*, 2007). Em concordância, Sánchez *et al.* (2013) também demonstra que as empresas insolventes tendem a mudar mais vezes de auditor, o que não acontece nas empresas ativas. Para estes autores, a mudança de auditor pode estar associada a desacordos entre gestores e auditores quanto à contabilização de transações. Contudo, e contrariamente ao evidenciado por estes autores, Muñoz-Izquierdo *et al.* (2017) conclui que não há evidência estatística que permita afirmar que a mudança do auditor está relacionada com a insolvência das empresas.

3. Hipóteses de investigação e metodologia

3.1. Hipóteses de investigação e posicionamento da investigação

Tendo em conta a questão de investigação formulada e a revisão de literatura formulam-se três hipóteses de investigação:

- H1: A dimensão das variáveis do trabalho realizado em auditoria é um bom preditor das insolvências, prevendo-se modelos com boa capacidade preditiva.
- H2: A dimensão das variáveis da auditora aumenta a capacidade preditiva dos modelos, quando adicionada no modelo com as variáveis da dimensão do trabalho em auditoria.
- H3: Os tópicos que caracterizam os comentários resultantes do trabalho em auditoria estão estatisticamente relacionados com o *status* da empresa, tendo estes uma importância significativa nos modelos preditivos.

A partir da década de 1970, a pesquisa positivista ou empírica passou a ser o paradigma dominante (Major, 2009). Este tipo de pesquisa permite analisar relações entre diferentes variáveis, expressas usualmente em hipóteses de investigação previamente formuladas tendo em conta a literatura existente (Major, 2017).

Assim, tendo em consideração os objetivos propostos, pode afirmar-se que este estudo se insere num posicionamento positivista e segue um *mix* entre a metodologia qualitativa e quantitativa, uma vez que nos modelos preditivos são utilizadas variáveis qualitativas, resultantes da análise do conteúdo das CLC/RAFE, e quantitativas.

Este estudo consubstancia um estudo exploratório, uma vez que ao abordar o tema da predição das insolvências tendo como preditores as dimensões das variáveis do trabalho realizado em AFE e as dimensões das variáveis da auditora em si, contribuiu para obter conhecimento deste tema pouco abordado na literatura.

3.2. População e amostra

A informação incluída neste trabalho é retirada da base de dados SABI, que contém toda a informação necessária para a realização da investigação. A partir desta base de dados é possível retirar informação de sociedades anónimas (S.A.) e das sociedades por quotas de responsabilidade limitada (LDA), sujeitas a revisão oficial de contas, quer por obrigatoriedade ao abrigo do artigo 262º do Código das Sociedades Comerciais, quer por voluntariedade.

3.2.1. Processo de seleção da amostra

A amostra é constituída por empresas portuguesas, obtida através de um processo com várias fases (tabela 3.4). Numa primeira instância, é recolhida uma amostra de 25.172 empresas, 21.461 empresas

com *status* ativa e 3.711 com *status* insolvente, entre os anos de 2016 a 2020. Posteriormente, retira-se à amostra inicial todas as empresas cuja forma jurídica é distinta de S.A. e LDA, bem como todas as empresas pertencentes às secções K e O da Classificação Portuguesa de Atividades Económicas (CAE), restringindo-se a amostra a 20.551 empresas, 3.338 com *status* insolvente e 17.213 ativas. Tendo em conta o objetivo do estudo exclui-se, também, todas as empresas que não são sujeitas a revisão oficial de contas e que têm dados omissos numa das variáveis utilizadas na investigação de forma a garantir que a empresa ainda é auditada e ainda se encontra em atividade no ano de análise, restringindo-se a amostra a 2.040 empresas, 1.020 em insolvência e 1.020 ativas. Importa referir que, nesta última fase, a seleção aleatória final das empresas ativas é baseada tendo em conta o número e perfil das empresas insolventes selecionadas, de modo que a amostra final seja composta pelo mesmo número de empresas ativas e insolventes e cujos perfis (dimensão e setor de atividade) sejam idênticos.

Tabela 3.4 - Processo de triagem dos dados

Descrição	Nº de empresas insolventes
Base de dados SABI	25.172
Após eliminação de empresas com formas jurídicas distintas de S.A. e LDA e pertencentes às secções K e O da CAE	20.551
Após eliminação de empresas que não são sujeitas a revisão oficial de contas ou que têm dados omissos e redução aleatória das empresas ativas por perfil	2.040

Fonte: Elaboração própria

Assim, este estudo aborda uma amostra final de 2.040 empresas, 1.020 com *status* insolvente e 1.020 com *status* ativa, correspondente ao período de 2016 a 2020.

A tabela 3.5 distribui o total das 2.040 empresas pelos diferentes anos de análise. Assim, pode verificar-se que são analisadas 297 empresas que entraram em insolvência no ano de 2020, bem como 297 empresas ativas em 2020, totalizando 594 empresas analisadas em 2020. Relativamente a 2019 são analisadas 246 empresas que entraram em insolvência neste mesmo ano e, também, outras 246 empresas ativas em 2019. Note-se ainda que o número de empresas tende a aumentar ao longo do período de análise.

Tabela 3.5 - Distribuição das empresas da amostra por ano

	2016	2017	2018	2019	2020	Total
Insolventes	125	177	175	246	297	1020
Ativas	125	177	175	246	297	1020
Total	250	354	350	492	594	2040

Fonte: Elaboração própria

3.2.2. Caracterização da amostra

Primeiramente, a amostra das empresas é classificada como micro, pequena ou média empresa de acordo com o Decreto-Lei n.º 372/2007, de 6 de novembro, que estipula que a categoria das PME é constituída por empresas cujo volume de negócio anual não exceda 50 milhões de euros ou cujo total do balanço anual não exceda 43 milhões de euros e empreguem menos de 250 pessoas. Evidencia, também, que uma microempresa é definida por empregar menos de 10 pessoas, sendo que o seu volume de negócios anual ou o balanço total não excede 2 milhões de euros. Por outro lado, uma pequena empresa é caracterizada por empregar menos de 50 colaboradores e cujo seu volume de negócios ou balanço total não ultrapassa os 10 milhões de euros (tabela 3.6).

Tabela 3.6 - Critério de Classificação das PME

Dimensão	Nº de colaboradores	Volume de Negócios (VN) ou Balanço Total (BT)
Micro	< 10	<= 2 milhões de euros de VN
Pequena	< 50	<= 10 milhões de euros de BT
Média	< 250	<= 50 milhões de euros de VN ou 43 milhões de euros de BT

Fonte: Adaptado do IAPMEI, I.P. – Agência para a Competitividade e Inovação

A tabela 3.7, classifica a amostra das 2.040 empresas em micro, pequenas e médias empresas, tendo em conta os critérios de classificação presentes na tabela 3.6. As empresas que não se enquadram nestas condições são classificadas como não PME. De acordo com a tabela 3.7, podemos concluir que os grupos com maior representatividade são as pequenas e microempresas tendo, respetivamente, 962 (47%) e 888 (43%) empresas. O grupo das não PME regista apenas 12 empresas (1%).

Tabela 3.7 – Distribuição das empresas por dimensão

	Micro	Pequena	Média	Não PME	Total
Empresas	888	962	178	12	2040
%	43%	47%	9%	1%	100%

Fonte: Elaboração própria

Caracterizando as empresas por regiões (NUTS II), verifica-se que a região do Norte regista no total 731 empresas (36%), tal como evidenciado na tabela 3.8. A região de Lisboa e Vale do Tejo regista 876 empresas (43%). Com ainda uma percentagem considerável, o centro de Portugal regista 269 empresas (13%).

Tabela 3.8 - Distribuição das empresas por região (NUTS II)

NUTS II	Nº de observações	%
Norte	731	36%
Lisboa e Vale do Tejo	876	43%
Centro Portugal	269	13%
Alentejo	60	3%
Algarve	39	2%
Região Autónoma da Madeira	37	2%
Região Autónoma dos Açores	28	1%
Total	2040	100%

Fonte: Elaboração Própria

Por fim, as empresas da amostra são classificadas por setor de atividade, tendo em conta a CAE. Para garantir a representatividade dos setores opta-se por agrupar alguns setores, tais como os setores das secções C à F e da H à S, sendo estes representativos dos setores da Indústria e Serviços, respetivamente, tal como acontece em investigações idênticas (Constantino, 2020). Importa referir que, nesta classificação opta-se por dividir a amostra total em subamostras, considerando uma subamostra de 1.020 empresas insolventes e outra subamostra de 1.020 empresas ativas. Com esta divisão pretende-se evidenciar que ambos os grupos de empresas, insolventes e ativas, têm perfis idênticos no que respeita o setor de atividade e à dimensão.

Da análise da tabela 3.9, pode observar-se que o setor da Indústria é o mais representado com 56,6% das empresas, seguindo-se o setor dos serviços com 21,9%. O setor do comércio por grosso e a retalho; reparação de veículos automóveis e motociclos regista um total de 200 empresas (19,6%). Com menor representatividade, destaca-se o setor da indústria extrativa e o setor da agricultura, produção animal, caça, floresta e pesca.

Tabela 3.9 - Distribuição da subamostra das empresas por secção CAE e dimensão

Secções CAE	Descrição	Micro	Pequena	Média	Não PME	Total	%
A	Agricultura, produção animal, caça, floresta e pesca	12	4	0	0	16	1,6
B	Indústrias extrativas	2	2	0	0	4	0,4
C à F	Indústria	227	287	60	3	577	56,6
G	Comércio por grosso e a retalho; reparação de veículos automóveis e motociclos	86	101	12	1	200	19,6
H à S	Serviços	117	87	17	2	223	21,9
Total		444	481	89	6	1020	100,0

Fonte: Elaboração própria

3.3. Dados

Recordando, esta investigação tem como principal objetivo analisar o contributo do trabalho efetuado em AFE e das características da auditora na predição das insolvências. Neste sentido, para a concretização dos objetivos foram recolhidos, para além dos dados demográficos, dados relacionados com a AFE e com a situação financeira das empresas. Para a recolha destes dados recorreu-se à base de dados SABI. A SABI representa a base de dados mais completa de análise financeira de empresas portuguesas e espanholas e, contém, informação atual e histórica sobre mais de dois milhões e meio de empresas.

3.3.1. Variável dependente

A variável dependente é uma variável dicotómica, de forma a distinguir o *status* das empresas:

- Empresas com o *status* “Insolvência/ Trâmites de composição”
- Empresas com o *status* “Ativa”

Este tipo de variável é comum na maioria dos estudos abordados na revisão de literatura em que o objetivo é a elaboração de modelos preditivos das insolvências (Muñoz-Izquierdo *et al.*, 2019, 2020).

3.3.2. Variáveis independentes

Relativamente às variáveis independentes utilizadas neste estudo, estas dividem-se em três dimensões: trabalho realizado em AFE, características da auditora e económico-financeira. No que diz respeito à dimensão do trabalho em AFE considera-se a variável Tipo_opinião, que classifica as empresas que têm a CLC/RAFE com uma opinião não qualificada e as empresas que têm a CLC/RAFE com uma opinião qualificada, Mud_tipo_opinião que diferencia as empresas cuja opinião não é alterada entre o ano N-1 e N, das empresas em que a opinião muda de não qualificada para qualificada e, também, das empresas em que a opinião varia de qualificada para não qualificada.

Ainda na primeira dimensão, consideram-se as variáveis Comentário, Dim_comentário e Tópico. Com a variável Comentário pretende-se analisar se a presença de comentários, seja reserva, ênfase ou ambas, é um bom preditor das insolvências. A variável diferencia as empresas que têm comentário na CLC/RAFE, das empresas que não têm qualquer comentário, isto é, quando o auditor emite uma opinião sem qualquer comentário (comentário limpo/opinião limpa). Importa referir que, para a classificação desta variável, não importa o número de comentários emitidos ou o tipo de comentário (reserva ou ênfase). A variável Dim_comentário evidencia o número de caracteres do comentário de modo a perceber-se se a dimensão dos comentários é um bom preditor das insolvências. Por último a variável Tópico, que resulta do trabalho realizado em *text mining*, onde se pretende analisar se os tópicos mencionados nos comentários são um bom preditor das insolvências. Esta variável distingue

as empresas que têm os comentários associados ao tópico 1 capital e continuidade, das empresas que têm os comentários associados ao tópico 2 gastos e ativo, das empresas cujo comentário aborda o tópico 3 investimento e imparidade e, também, das empresas em que o comentário está relacionado com o tópico 4 resultados e caixa. Importa referir que os comentários de uma empresa só podem estar relacionados a um tópico, o mais evidente.

Já em relação à segunda dimensão de variáveis, considera-se: Dim_auditora que diferencia as empresas que são auditadas por uma firma de auditoria que pertence ao grupo das *Big 4* das empresas que são auditadas por firmas não *Big 4* e, também, a variável “Mudança_auditora” que diferencia as empresas que não mudam de auditoria entre N-1 e N, das empresas que mudam para uma não *Big 4*, das empresas que mudam para *Big 4*, das que mudam mas continuam a ser auditadas por uma não *Big 4* e, por fim, das empresas das empresas que mudam mas continuam a ser auditadas por uma *Big 4*. Note-se que pertencem ao grupo das *Big 4*, as firmas PricewaterhouseCoopers (PwC), KPMG, Ernest & Young (EY) e Deloitte.

Por fim, a terceira dimensão de variáveis diz respeito aos dados da situação financeira das empresas, nomeadamente, EBITDA, solvabilidade, endividamento, ativo (AT) e vendas (V), que também são consideradas em investigações anteriores e que têm um boa capacidade preditiva das insolvências (Altman, 1968; Beaver, 1966). A tabela 3.10 ilustra as fórmulas utilizadas na conceção dos rácios financeiros.

Relativamente às variáveis demográficas são consideradas a forma jurídica, a NUTSII e a CAE, tal como usual neste tipo de estudos (Abrantes, 2020).

Tabela 3.10 - Fórmulas utilizadas na elaboração dos rácios

Rácios	Fórmulas utilizadas
Solvabilidade	$\frac{\textit{Capital próprio}}{\textit{Passivo}}$
Endividamento	$\frac{\textit{Passivo}}{\textit{Ativo}}$

Fonte: Elaboração própria.

A tabela 3.11, evidencia as variáveis independentes, a sua escala, a unidade de medida e para que objetivos cada variável é importante. Note-se que são, também, recolhidos dados dos anos de 2015 e 2014 para analisar as empresas em insolvência e ativas de 2016 e 2017.

Tabela 3.11 - Descrição das variáveis independentes

Variáveis	Descrição	Escala ⁽²⁾	Unidade de medida	Objetivos
Características das duas dimensões de auditoria				
Tipo de opinião ⁽¹⁾	Tipo de opinião	QN	n.a.	II, III e IV
Mud_tipo_opinião ⁽¹⁾	Mudança do tipo de opinião	QN	n.a.	II, III e IV
Comentário ⁽¹⁾	Presença de comentário	QN	n.a.	II, III e IV
Dim_comentário ⁽¹⁾	Dimensão do comentário	QD	unidade	II, III e IV
Tópico ⁽¹⁾⁽³⁾	Tópicos salientes	QN	n.a.	I, II, III e IV
Dim_auditora ⁽¹⁾	Dimensão da auditora	QN	n.a.	II, III e IV
Mud_auditora ⁽¹⁾	Mudança de auditora	QN	n.a.	II, III e IV
Características económico-financeiras				
V ⁽¹⁾	Vendas	QC	x 1000 euros	II, III
EBITDA ⁽¹⁾	EBITDA	QC	x 1000 euros	II, III
Solvabilidade ⁽¹⁾	Solvabilidade	QC	x 1000 euros	II, III
Endividamento ⁽¹⁾	Endividamento	QC	x 1000 euros	II, III
AT ⁽¹⁾	Ativo	QC	x 1000 euros	II, III
Características demográficas				
Forma_legal	Forma jurídica	QN	n.a.	IV
NUTSII	Região	QN	n.a.	IV
Dimensão	Dimensão da empresa	QO	n.a.	IV
CAE	Setor de atividade	QN	n.a.	IV

Nota: (1) Variáveis calculadas para os anos N e N-1. (2) QN: Qualitativa nominal; QO: Qualitativa ordinal; QD: Quantitativa discreta; QC: Quantitativa contínua. (3) Variável criada com *text mining* conforme descrito no ponto 3.4.1.

Fonte: Elaboração própria.

3.4. Técnicas de análise de dados

A qualidade da informação retirada de uma base de dados tem um papel crucial no desenvolvimento de um estudo. Deste modo, verificou-se que a informação exportada da base de dados SABI continha não respostas e *outliers*. Consideram-se *outliers* todos os valores que se afastam, tanto para cima como para baixo, em mais do que três desvios padrões da média. De forma a não impactarem os resultados, estes *outliers* são substituídos pela média mais ou menos três desvios padrões (Jacinto, 2019). Tendo em conta as não respostas, estas são consideradas como dados omissos, uma vez que é impossível conhecer essa informação.

Na análise da informação financeira verificam-se algumas limitações no âmbito da qualidade da informação pelo que se descarta a utilização da terceira dimensão de variáveis nos modelos preditivos.

Contudo, a análise de cada uma destas variáveis e a sua relação com o *status* da empresa é realizada, sendo apresentados os resultados nos anexos A e B.

Importa referir que, para determinadas variáveis, o número de não respostas é bastante elevado, nomeadamente, na variável Dim_comentário e Tópico devido à falta de comentários divulgados pela base de dados.

Para além disso, é importante salientar que o número de dados válidos para a variável Tópico não coincide com o número de dados válidos da variável Dim_comentário. Esta situação deve-se ao facto de, na identificação dos tópicos presentes nos comentários emitidos pelos auditores, através da ferramenta de *text mining*, terem sido utilizados o mesmo número de empresas ativas e insolventes com comentários, daí nesta análise não se considerar algumas empresas ativas com comentário. Se fosse utilizado um maior número de empresas ativas do que empresas insolventes, provavelmente, não iria ser possível identificar tópicos associados às insolvências.

Durante a preparação dos dados verifica-se que existe muita falta de informação relativa ao ano da insolvência das empresas. Devido a essa falta de informação, o estudo aborda o ano anterior à insolvência como o ano de referência. Por exemplo, uma empresa que entrou em insolvência em 2020 têm como ano de referência o ano de 2019, sendo o ano de 2019 igual ao ano N e o ano de 2018 igual ao ano N-1. Decidiu-se analisar os dois anos anteriores à insolvência porque a informação da auditoria é estatisticamente significativa na diferenciação das empresas ativas e insolventes num período temporal de até três anos, tendo em conta o ano da insolvência (Altman *et al.*, 2015).

No que diz respeito aos dados, estes são analisados recorrendo a diferentes técnicas de análise de dados. Primeiramente recorreu-se a técnicas de *text mining* para a criação dos tópicos salientes presentes nos comentários dos auditores. Importa referir que se recorre a esta técnica devido à extensão e à subjetividade na categorização dos comentários, em contraste com estudos anteriores onde é aplicada a análise manual do conteúdo dos comentários por parte de peritos em contabilidade (Muñoz-Izquierdo *et al.*, 2019). Depois, recorre-se a técnicas de estatística descritiva uni e bivariada para testar a qualidade dos dados e avaliar a relação entre as variáveis (Laureano, 2020). Por último, recorre-se a técnicas de estatística multivariada, nomeadamente as árvores de decisão com o algoritmo CART para a criação dos modelos preditivos do *status* da empresa (ativa e insolvente).

A tabela 3.12 sintetiza as principais técnicas utilizadas no estudo, relacionando-as com os objetivos específicos.

Tabela 3.12 - Técnicas de análise de dados utilizadas

Técnicas	Objetivos
Técnicas de <i>text mining</i>	I
Técnicas de estatística descritiva univariada	Preparação dos dados / II
Análise de <i>outliers</i>	Preparação dos dados
Técnicas de estatística descritiva bivariada	III
Árvores de decisão para classificação - CART	IV

Fonte: Elaboração própria

3.4.1. Text mining para identificar os tópicos nos comentários

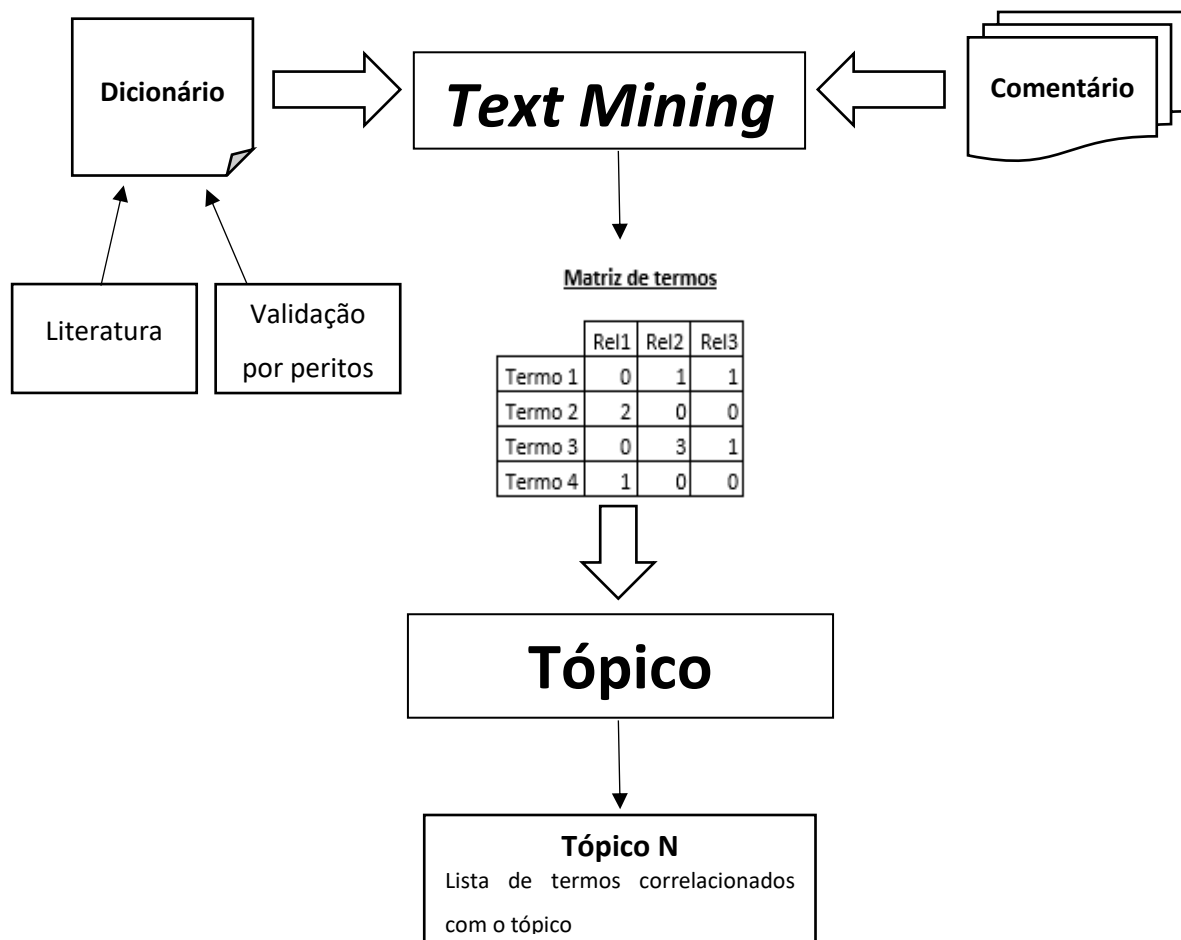
Neste estudo, recorre-se a técnicas de *text mining* para analisar, de forma automática, o conteúdo dos comentários (ênfases e reservas) emitidos na CLC/ RAFE. O objetivo da aplicação desta técnica de análise é identificar tópicos de forma a perceber se existe uma tendência dos termos mencionados, que permita diferenciar as empresas insolventes das ativas. A opção por esta técnica de análise de conteúdo, assistida por computador, torna-se evidente face ao volume de comentários a analisar.

O procedimento da análise desses comentários está descrito na figura 3.6. São necessários dois *inputs*, um dicionário de termos relevantes que possam ser mencionados pelo auditor (anexo C) e os comentários.

O dicionário, anexo C, é criado tendo em conta, para além de outros aspetos, investigações anteriores (Muñoz-Izquierdo *et al.*, 2017). Importa referir que o dicionário é revisto por um especialista, em particular pelo professor doutor Nuno Magro, docente do departamento de contabilidade do Iscte, que sugeriu que se acrescentassem termos como Artigo 35º do Código das Sociedades Comerciais, ativos biológicos, *goodwill*, ativo, passivo, Método da Equivalência Patrimonial (MEP) e Regime Extrajudicial de Recuperação de Empresas (RERE). O dicionário incorpora termos similares, tais como, formas no plural e sinónimos, como por exemplo, “cliente”, “clientes” e “devedores”, de forma a garantir a representatividade dos termos com significado semelhante.

Relativamente aos comentários, importa referir que, para a concretização do primeiro objetivo específico, alguns comentários das empresas ativas são ignorados de forma a ter o mesmo número de comentários em ambas as subamostras, insolventes e ativas. Assim, são analisados 162 comentários em N (81 de empresas ativas e 81 de empresas insolventes) e 168 em N-1 (84 de empresas ativas e 84 de empresas insolventes). Esta seleção das empresas ativas com comentários é feita de forma aleatória tendo em conta o ano, de forma a analisar o mesmo número de comentários relativos ao ano 2020, 2019 e assim sucessivamente, nas empresas ativas e insolventes.

Figura 3.6 - Processo experimental do *text mining*



Fonte: Adaptado de Santos *et al.* (2020, pp 7)

Após todo este processo da elaboração e validação do dicionário, os *inputs* são processados no *software Orange* de forma a obter uma matriz de termos, que indica o número de vezes que os termos aparecem nos comentários. Através desta matriz é possível analisar a frequência absoluta dos termos na amostra.

Posteriormente, a aplicação da técnica de *topic modeling* permite gerar diferentes combinações de termos associados a tópicos, aos quais estão associados diferentes probabilidades marginais aos tópicos. Após ser testado diferentes números de tópicos, conclui-se que quatro tópicos é o cenário em que as probabilidades marginais relativas de cada tópico são mais significativas (iguais ou superiores a 0,1). Opções com um maior número de tópicos faria com que as probabilidades marginais fossem muito reduzidas em determinados tópicos.

Importa referir que para a obtenção da matriz de termos são desenvolvidos dois modelos, um para as empresas insolventes e outro para as ativas, de forma a obter a frequência absoluta de palavras para estes dois subgrupos e, posteriormente, para a criação de duas nuvens de palavras distintas de modo a analisar estes dois subgrupos separadamente. Já no que diz respeito à identificação dos tópicos

é utilizado apenas um modelo para o total da amostra de forma a ser possível criar a variável Tópico que possa diferenciar as empresas (ativas e insolventes) de acordo com os tópicos associados aos comentários.

3.4.2. Modelos preditivos com árvores de decisão

As árvores de decisão são uma técnica estatística multivariada que podem ser contruídas através de diferentes algoritmos, nomeadamente, o CART e o *Chi-squared Automatic Interaction Detection* (CHAID), que são os mais abordados na literatura (Pestana & Gageiro, 2014). As árvores de decisão são classificadas em árvores de regressão quando a variável de resposta (variável dependente) é quantitativa e em árvores de classificação quando a variável é qualitativa, que é o caso deste estudo (insolvente ou ativa).

Nesta investigação são concebidas árvores de classificação através do algoritmo de CART, uma vez que não requerem a verificação de pressupostos e lidam bem com todos os tipos de variáveis, *outliers* e não respostas (Constantino, 2020). O algoritmo CART cria árvores de decisões binárias, ou seja, divide os dados em dois nós tendo em conta a homogeneidade dos dados e caracteriza-se pela sua fácil interpretação e boa capacidade preditiva (Pestana & Gageiro, 2014).

Para analisar a validação dos modelos preditivos opta-se pela utilização do método de validação *holdout* que divide a amostra em duas partições, a de treino (30%) e a de teste (70%). Para além deste, o método de validação cruzada *k-fold* também é utilizado de forma a complementar o método anterior. Este segundo método tem um funcionamento muito semelhante ao método *holdout* mas com rotação das *k* partições. Esta rotação permite que os resultados obtidos para o treino e para o teste sejam o reflexo da totalidade do conjunto de dados utilizados, o que permite aproveitar toda a amostra do estudo. Define-se $k=10$, uma vez que esta repartição é útil para amostras de pequenas dimensões (Cabete & Cardoso, 2006) e, também, porque depois de várias tentativas é a repartição que permite obter melhores resultados. Os melhores modelos são aqueles que apresentam valores mais elevados na métrica PECC no método da validação cruzada.

De forma a melhorar a capacidade preditiva dos modelos são definidos parâmetro que permitem contrariar o sobre e subajustamento dos modelos, nomeadamente, o número máximo de níveis da árvore (profundidade), a fixação do número mínimo de casos por nó pai bem como, o número mínimo de casos por nó filho. Após várias experiências, a tabela 3.13 evidencia a parametrização que permite atingir melhores resultados em cada modelo, bem como as diferentes dimensões de variáveis utilizadas nesses modelos e os anos a que respeitam.

Tabela 3.13 - Parametrização dos modelos preditivos das insolvências

Parâmetros	Modelos					
	A	B	C	D	E	F
Profundidade máxima	5	5	5	5	5	5
Mínimo de casos por nós pai	2	2	2	5	6	4
Mínimo de casos por nós filho	1	1	1	2	2	2
Ganho mínimo na pureza	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001
Validação cruzada - partições	10	10	10	10	10	10
Treino (70%) / Teste (30%)	Sim	Sim	Sim	Sim	Sim	Sim
Dimensão das variáveis	AFE	AFE Auditora	AFE	AFE Auditora	AFE Auditora	AFE Auditora Demográficas
Anos	N1	N1	N	N	N N1	N N1

De forma a avaliar a qualidade dos modelos recorre-se a diferentes métricas, tais como, à PECC, à sensibilidade, isto é, a percentagem das empresas insolventes que o modelo acertou e estão de facto em insolvência e, também, à especificidade que representa a percentagem de casos corretamente classificados de empresas ativas. Todas estas técnicas mencionadas anteriormente são baseadas na matriz confusão (tabela 3.14).

Tabela 3.14 – Exemplo de matriz confusão

Status observado	Status previsto	
	Insolvente	Ativa
Insolvente	Verdadeiro positivo	Falso negativo
Ativa	Falso positivo	Verdadeiro negativo

Nota: O alvo é prever as insolvências

4. Resultados e discussão

Este capítulo apresenta os resultados da análise de dados e uma discussão sobre os mesmos, à luz da metodologia aplicada nesta investigação para cada um dos objetivos formulados.

4.1. Identificação dos tópicos salientes presentes nos comentários

4.1.1. Frequência absoluta dos termos nas subamostras das empresas ativas e insolventes

Para melhor compreensão dos resultados, numa primeira fase, o estudo analisa as duas subamostras separadamente, de forma a identificar os termos nos comentários dos auditores com referência às empresas ativas e insolventes. O anexo C e a figura 4.7, evidenciam a frequência absoluta dos termos em ambas as subamostras. Pode verificar-se que o termo continuidade é o mais mencionado pelo auditor nas empresas insolventes (117 vezes), enquanto nas empresas ativas só é mencionado 32 vezes. Estudos anteriores evidenciam que o auditor tende a emitir um comentário com uma incerteza material relacionada com a continuidade das operações nas empresas insolventes (DeFond & Zhang, 2014). Por outro lado, o termo gasto é o mais mencionado nas empresas ativas (91 vezes), enquanto nas empresas insolventes só é citado pelo auditor 39 vezes.

A discrepância dos termos pode ser melhor visualizada e interpretada através da nuvem de palavras (figura 4.7) onde o tamanho das palavras é proporcional ao número de vezes que aparecem nos comentários. Da sua análise identificam-se outros termos que tendem a diferenciar as empresas ativas das insolventes, nomeadamente, os termos acionistas, art35, financiamento, insolvência e plano especial de revitalização que estão presentes com alguma frequência nos comentários dos auditores com referências às empresas insolventes comparativamente com as empresas ativas. Termos como, por exemplo, ativos biológicos e contas individuais não aparecem nem nos comentários das empresas ativas, nem das empresas insolventes.

No contexto da pandemia COVID-19, é possível verificar que os auditores já passaram a mencionar o termo COVID-19 nos seus comentários referentes a 2019, aparecendo 24 vezes nas empresas ativas e 12 vezes nas empresas insolventes. Um dos comentários, aleatoriamente selecionado, emitido por um auditor relativamente a este tema é “os recentes desenvolvimentos resultantes da pandemia Covid-19 (Coronavírus) têm um impacto significativo na saúde das pessoas e na sociedade como um todo, aumentando a incerteza sobre o desempenho operacional e financeiro das organizações. A nossa opinião não é modificada em relação a esta matéria.”

Figura 4.7 - Nuvem de palavras dos termos mencionados nos comentários



Fonte: Elaboração própria.

4.1.2. Tópicos nos comentários

A tabela 4.15 identifica os tópicos criados que estão presentes nos comentários provenientes do trabalho em AFE. Os resultados mostram que o tópico 1 capital e continuidade é o que melhor caracteriza os comentários, uma vez que é o tópico com uma maior probabilidade marginal (0,370), seguindo-se o tópico 3 imparidade e investimento, o tópico 2 gastos e ativo e o tópico 4 resultados e caixa, com probabilidades marginais de 0,294, 0,236 e 0,100, respetivamente.

Tabela 4.15 – Tópicos identificados nos comentários

Tópicos	Probabilidade marginal dos tópicos
Tópico 1 capital e continuidade	0,370
Tópico 2 gastos e ativo	0,236
Tópico 3 imparidade e investimento	0,294
Tópico 4 resultados e caixa	0,100
Total	1

Fonte: Elaboração própria

Os nomes atribuídos aos tópicos estão relacionados com os termos mais correlacionados com cada um dos tópicos de forma a conseguir-se perceber, através do nome do tópico, que temas são referidos pelo auditor nos comentários emitidos na CLC/RAFE (tabela 4.16).

Dos resultados obtidos, os termos mais correlacionados com o tópico que melhor caracteriza a amostra (tópico 1 capital e continuidade) são, passivo, continuidade, capital, financiamento, limitação de âmbito, acionistas e Artigo 35º do Código das Sociedades Comerciais (CSC). Por outro lado, os termos que mais se relacionam com o tópico 2 gastos e ativo são: gasto, inventário, ativo, depreciação e resultado líquido. Relativamente ao tópico 3 investimento e imparidade, o termo com correlação mais significativa é mensuração, seguindo-se outros termos como, imparidade, investimento, devedor, depreciação e credor. Por último, o tópico 4 resultados e caixa, com menor probabilidade marginal tem como termos mais correlacionados: resultado transitado, vida útil e caixa. Note-se que uma empresa que tem um determinado tópico associado ao comentário, o comentário em causa pode não incluir todos os termos correlacionados com esse mesmo tópico.

A contextualização dos termos de cada tópico no comentário do auditor implica olhar para alguns exemplos exemplificativos. A título de exemplo, o auditor nos comentários caracterizados pelo tópico 1 capital e continuidade refere que “(...) verifica-se o incumprimento das disposições legais dos artigos 35º do Código das Sociedades Comerciais, pelo facto de estar perdido mais de metade do capital social. No entanto, conforme divulgado na nota 2 do anexo e no relatório de gestão, as demonstrações financeiras foram preparadas com base na continuidade das operações, a qual se encontra dependente do apoio dos seus acionistas, dos financiamentos da banca e outras entidades, do sucesso das operações e atividades futuras”.

Por outro lado, nos comentários associados com o tópico 4 resultados e caixa, o auditor refere que “tendo por referência as vidas úteis dos ativos imobiliários e dos ativos intangíveis definidas ao nível das demonstrações financeiras e considerando-as mais aderentes ao período de vida útil estimado dos referidos ativos do que aquelas que a Entidade utiliza na preparação das suas demonstrações financeiras, o Ativo, os Resultados Transitados e o Resultado Líquido a 31 de dezembro de 2018 encontram-se, respetivamente, subavaliados (...)”.

Tabela 4.16 - Termos correlacionados com os tópicos

Tópico 1 capital e continuidade	β	Tópico 2 gastos e ativo	β
Passivo	0,19	Gasto	0,34
Continuidade	0,18	Inventário	0,15
Capital	0,16	Ativo	0,10
Financiamento	0,10	Depreciação	0,09
Limitação de âmbito	0,09	Res. líquido	0,06
Acionistas	0,09		
Art35	0,05		
Tópico 3 imparidade e investimento	β	Tópico 4 resultados e caixa	
Mensuração	0,26	Res. transitado	0,32
Imparidade	0,19	Vida útil	0,30
Investimento	0,16	Caixa	0,24
Devedor	0,13		
Depreciação	0,08		
Credor	0,07		

Nota: A correlação entre o tópico e o termo é representada por β

Fonte: Elaboração própria.

Relativamente ao tópico 3 imparidade e investimento, o auditor refere que “(...) no decurso do nosso trabalho não obtivemos confirmação externa de saldo de clientes com antiguidade superior a 24 meses, nem nos foi possível realizar procedimentos alternativos que nos permitissem validar o montante deste ativo para podermos concluir sobre a recuperabilidade destas dívidas a receber de clientes. Assim, existem insuficiências de perdas por imparidade que poderão ascender ao valor de 2.666.173 euros, pelo que o ativo e os capitais próprios se encontram sobrevalorizados por este montante. Não nos foi disponibilizada informação suficiente que nos permita validar a quantia de 100.000 euros registada em outros investimentos financeiros, pelo que o ativo e os capitais próprios se encontram sobrevalorizados nesta quantia.”.

Por último, e no que diz respeito ao tópico 2 gastos e ativo, o auditor menciona nos comentário que “considerando que pela natureza dos inventários da empresa, designadamente sucatas de ferro e outros metais, os custos de movimentação são significativos, não procedendo a Empresa à respetiva pesagem à data do Balanço, encontramos-nos impossibilitados de realizar o procedimento de auditoria que consiste na assistência às contagens físicas dos elementos constitutivos desta rúbrica. Esta circunstância não nos permite a obtenção de evidência suficiente que nos permita concluir acerca da razoabilidade substantiva dos valores apresentados em Inventários (4.200.668 euros; 4.216.551 euros, no exercício transato) e dos seus eventuais efeitos no Custo das Mercadorias Vendidas e Matérias Consumidas”

4.2. Caracterizar o trabalho da AFE e as características da auditora

Tendo em conta a análise univariada, no que respeita às variáveis qualitativas das duas dimensões de variáveis (tabela 4.17), pode analisar-se que a variável Tipo_opinião segue uma distribuição mais ou menos equitativa pela qualificação e não qualificação da opinião, apesar de haver mais CLC/RAFE com uma opinião qualificada, tanto em N, como em N-1 (56,4% e 54,5%, respetivamente). Contrariamente, relativamente à Mud_tipo_opinião, verifica-se que a não mudança do tipo de opinião entre os anos analisados é predominante face à mudança de opinião, sendo que 93,7% das empresas têm uma CLC/RAFE com o mesmo tipo de opinião em N e N-1.

A maioria das empresas tende a ter comentários nas suas CLC/RAFE, tanto em N, como em N-1 (60,9% e 56,3%, respetivamente). Pelo contrário, 39,1% das empresas no ano N não tem nenhuma reserva ou ênfase emitida pelo auditor, pelo que se pode concluir que as CLC/RAFE têm uma opinião limpa.

Tendo em conta, ainda, a dimensão do trabalho em AFE, pode analisar-se que, relativamente à variável Tópico, tanto em N, como N-1, o tópico 1 capital e continuidade predomina face aos restantes (38,9% em N e 35,1% em N-1). Para além disso, importa referir que o número de empresas com comentários associados ao tópico 3 imparidade e investimento mantém-se constante de N-1 para N. No que diz respeito ao tópico 2 gastos e ativo, apesar da sua presença significativa nos comentários do ano N-1 (28%), em N esta percentagem diminui consideravelmente (19,1%).

Por outro lado, no que diz respeito à dimensão das características da auditora responsável pela AFE às contas das empresas, 54,8% são empresas que pertencem ao grupo das não *Big 4* e 45,2% são empresas *Big 4* (ano N). O ano N-1 segue uma distribuição semelhante à do ano N. Relativamente à mudança de auditora, a tabela 17 evidencia que 92,4% das empresas não muda de auditora da transição do ano N-1 para o ano N.

Através da tabela 4.18, no que diz respeito à variável Dim_comentário pode analisar-se que, em média, o número de caracteres dos comentário emitidos pelo auditor é maior no ano N-1 face ao ano N, sendo que em N-1 é de 1.112 caracteres e em N é de 1.037. Note-se que o desvio padrão é bem elevado comparativamente com a média, pelo que não se consegue tirar grandes conclusões da análise desta tabela. No entanto, em termos medianos, também o número de caracteres em N (747) é inferior ao número de caracteres do ano anterior (878).

Tabela 4.17 - Medidas descritivas das variáveis qualitativas do trabalho em AFE e da auditora

Variável	Descrição	Nº	%
Tipo_opinião_N	Opinião não qualificada	890	43,6
	Opinião qualificada	1150	56,4
	Total	2040	100,0
Tipo_opinião_N1	Opinião não qualificada	928	45,5
	Opinião qualificada	1112	54,5
	Total	2040	100,0
Mud_tipo_opinião	Opinião não muda	1912	93,7
	Muda de não qualificada para qualificada	83	4,1
	Muda de qualificada para não qualificada	45	2,2
	Total	2040	100,0
Comentário_N	Há comentário	1242	60,9
	Não há comentário	798	39,1
	Total	2040	100,0
Comentário_N1	Há comentário	1149	56,3
	Não há comentário	891	43,7
	Total	2040	100,0
Tópicos_N	Tópico 1 capital e continuidade	63	38,9
	Tópico 2 gastos e ativo	31	19,1
	Tópico 3 imparidade e investimento	47	29,0
	Tópico 4 resultados e caixa	21	13,0
	Total	162	100,0
Tópicos_N1	Tópico 1 capital e continuidade	59	35,1
	Tópico 2 gastos e ativo	47	28,0
	Tópico 3 imparidade e investimento	50	29,8
	Tópico 4 resultados e caixa	12	7,1
	Total	168	100,0
Dim_auditora_N	Auditada por não <i>Big 4</i>	630	54,8
	Auditada por <i>Big 4</i>	520	45,2
	Total	1150	100,0
Dim_auditora_N1	Auditada por não <i>Big 4</i>	664	56,9
	Auditada por <i>Big 4</i>	503	43,1
	Total	1167	100,0
Mud_auditora	Não muda	1056	92,4
	Muda para não <i>Big 4</i>	3	0,3
	Muda para <i>Big 4</i>	16	1,4
	Muda mas continua não <i>Big 4</i>	43	3,8
	Muda mas continua <i>Big 4</i>	25	2,2
	Total	1143	100,0

Tabela 4.18 - Medidas descritivas da variável quantitativa da dimensão do comentário

Variável	Nº	Média	Desvio padrão	Mínimo	Mediana	Máximo
Dimensão_comentário_N	306	1037	789	94	747	2683
Dimensão_comentário_N1	253	1112	860	61	878	3047

4.3. Caracterizar a relação entre as dimensões de variáveis de auditoria e o *status*

De forma a caracterizar a relação entre as variáveis independentes nominais com o *status* da empresa recorre-se ao teste de independência Qui-quadrado (χ^2) e à medida de associação do teste do V de Cramer (VC), uma vez que, são adequados para este tipo de variáveis (tabela 4.19).

Dos resultados obtidos, e para a dimensão do trabalho em AFE, pode concluir-se que as variáveis Tipo_opinião e Comentário, tanto em N como em N-1 são as que mais se relacionam (relação forte) com o *status* da empresa, pois são aquelas cujo valor do teste do V de Cramer mais se aproxima da unidade.

No que diz respeito à variável Tipo_opinião_N1, 82,5% das empresas ativas têm uma opinião não qualificada o que se verifica também no ano N. No caso das empresas insolventes podemos analisar que em N-1 91,5% das empresas têm uma opinião qualificada e que no ano N há uma subida do número de CLC/RAFE com opiniões qualificadas de, aproximadamente, 4 pontos percentuais, sendo que em N quase todas as empresas insolventes (95,3%) apresenta uma opinião qualificada, o que corrobora o estudo de Altman (2010).

Relativamente à variável Comentário_N1, podemos concluir que 93,6% das empresas insolventes têm comentários emitidos pelo auditor, e à medida que se aproxima o ano da insolvência, essa percentagem de empresas aumenta ligeiramente (96,8%). Por outro lado, apesar de 25% das empresas ativas também terem comentários emitidos pelos auditores, 75% das empresas não têm, isto para o ano N.

Ainda dentro da dimensão do trabalho em AFE podemos concluir que a variável Tópico_N e Tópico_N1 têm uma relação fraca a moderada com o *status* da empresa (VC = 0,221 e VC = 0,357, respetivamente). Da análise da tabela pode concluir-se que o tópico 1 capital e continuidade destaca-se nos comentários das empresas insolventes, estando mencionado em cerca de 50% das empresas, tanto em N, como N-1. Pode também concluir-se que, nestas mesmas empresas, o tópico 3 investimento e imparidade tende a ser mencionado à medida que estas se aproximam da insolvência, pois em N-1 está presente em 23,8% dos comentários das empresas e em N está em 27,2% dos comentários. Por outro lado, e contrariamente às empresas insolventes, o tópico 3 investimento e imparidade diminui a sua presença no ano N (30,9%) comparativamente com o ano N-1 (35,7%).

No que diz respeito às empresas ativas, não há um tópico que se destaque significativamente dos outros, sendo que tanto em N, como em N-1, o tópico 3 Investimento e imparidade é o que mais está

presente nos comentários. Note-se que, há um aumento de cerca de 10 pontos percentuais do ano N-1 para o ano N no que diz respeito ao tópico de capital e continuidade. Estas conclusões podem estar relacionadas com a situação pandémica que o país e o mundo ultrapassa e com o facto de os auditores sentirem a necessidade de referir este tópico por desconhecerem os impactos que poderão ter na continuidade das empresas, conforme exemplo anteriormente descrito.

Tendo em conta a dimensão das características da auditora, pode concluir que, no que concerne o ano N e N-1, as empresas insolventes são auditadas maioritariamente por firmas não *Big 4*, sendo o seu peso de 95,4% e 96,1%, respetivamente, o que vai ao encontro de outra investigação que conclui o mesmo resultado (Peurseem & Chan, 2014). Por outro lado, nas empresas ativas existe uma distribuição mais ou menos equitativa entre empresas auditadas por firmas *Big 4* ou não *Big 4*, sendo que 50,4% são auditadas pelas grandes firmas de auditoria e 49,6% pelas mais pequenas (ano N).

Para a variável *Mud_auditora* não existe evidência estatística para se concluir quanto há existência de relação entre esta variável e o status das empresas ($\chi^2 = 9,3$; $p = 0,069$), considerando-se um nível de significância de 0,05.

Para a variável quantitativa da dimensão dos comentários, recorre-se ao teste t para duas amostras independentes e à medida de associação Eta. Pode concluir que, tanto em N, como em N-1, existe uma relação forte entre estas variáveis e o *status* da empresa, uma vez que o valor da medida de associação Eta está próximo da unidade (tabela 4.20). Pode também analisar-se que, em média, tanto em N como em N-1, a média de caracteres nos comentários emitidos é superior nas empresas ativas face às empresas insolventes, apesar desta diferença não ser significativa.

Tabela 4.19 - Relação entre as variáveis qualitativas da auditoria e o *status*

Variáveis	Descrição	Status				Test Qui-quadrado (χ^2) V de Cramer (VC)
		Ativa		Insolvência		
		Empresas	%	Empresas	%	
Tipo_opinião_N	Opinião não qualificada	842	82,5	48	4,7	$\chi^2 = 1256,6$; $p = 0,001$ VC = 0,785; $p = 0,001$
	Opinião qualificada	178	17,5	972	95,3	
	Total	1020	100,0	1020	100,0	
Tipo_opinião_N1	Opinião não qualificada	841	82,5	87	8,5	$\chi^2 = 1123,9$; $p = 0,001$ VC = 0,742; $p = 0,001$
	Opinião qualificada	179	17,5	933	91,5	
	Total	1020	100,0	1020	100,0	
Mud_tipo_opinião	Opinião não muda	941	92,3	971	95,2	$\chi^2 = 28,0$; $p = 0,001$ VC = 0,117; $p = 0,001$
	Muda de não qual. para qual.	39	3,8	44	4,3	
	Muda de qual. para não qual.	40	3,9	5	0,5	
	Total	1020	100,0	1020	100,0	
Comentário_N	Há comentário	255	25,0	987	96,8	$\chi^2 = 1102,9$; $p = 0,001$ VC = 0,735; $p = 0,001$
	Não há comentário	765	75,0	33	3,2	
	Total	1020	100,0	1020	100,0	
Comentário_N1	Há comentário	194	19,0	955	93,6	$\chi^2 = 1154,0$; $p = 0,001$ VC = 0,752; $p = 0,001$
	Não há comentário	826	81,0	65	6,4	
	Total	1020	100,0	1020	100,0	
Tópicos_N	Tópico 1 capital e continuidade	24	29,6	39	48,1	$\chi^2 = 7,9$; $p = 0,048$ VC = 0,221; $p = 0,048$
	Tópico 2 gastos e ativo	17	21,0	14	17,3	
	Tópico 3 imparidade e investimento	25	30,9	22	27,2	
	Tópico 4 resultados e caixa	15	18,5	6	7,4	
	Total	81	100,0	81	100,0	
Tópicos_N1	Tópico 1 capital e continuidade	17	20,2	42	50,0	$\chi^2 = 21,5$; $p = 0,001$ VC = 0,357; $p = 0,001$
	Tópico 2 gastos e ativo	26	31,0	21	25,0	
	Tópico 3 imparidade e investimento	30	35,7	20	23,8	
	Tópico 4 resultados e caixa	11	13,1	1	1,2	
	Total	84	100,0	84	100,0	
Dim_auditora_N	Auditada por não Big 4	505	49,6	125	95,4	$\chi^2 = 98,6$; $p = 0,001$ VC = 0,293; $p = 0,001$
	Auditada por Big4	514	50,4	6	4,6	
	Total	1019	100,0	131	100,0	
Dim_auditora_N1	Auditada por não Big4	517	51,0	147	96,1	$\chi^2 = 110,2$; $p = 0,001$ VC = 0,307; $p = 0,001$
	Auditada por Big4	497	49,0	6	3,9	
	Total	1014	100,0	153	100,0	
Mud_auditora	Não muda	934	92,1	122	94,6	$\chi^2 = 9,3$; $p = 0,069$ VC = 0,090; $p = 0,069$
	Muda para não Big4	2	0,2	1	0,8	
	Muda para Big4	15	1,5	1	0,8	
	Muda mas continua não Big4	43	4,2	0	0,0	
	Muda mas continua Big4	20	2,0	5	3,9	
	Total	1014	100,0	129	100,0	

Tabela 4.20 - Relação entre a variável quantitativa da dimensão dos comentários e o *status*

Variável	Status	Média	Desvio Padrão	Teste t Eta / Eta ²
Dim_comentário_N	Ativa	1066,13	792,65	t(304) = 1,1; $p = 0,284$
	Insolvência	956,30	779,20	Eta = 0,919; Eta ² = 0,896
Dim_comentário_N1	Ativa	1147,80	883,61	t(251) = 0,9; $p = 0,355$
	Insolvência	1041,40	810,51	Eta = 0,960; Eta ² = 0,939

4.4. Modelo preditivo das insolvências e perfis associados às empresas

4.4.1. Modelo preditivo

Os modelos preditivos estimados são apresentados na tabela 4.21. Primeiramente importa analisar que todos os modelos apresentam bons resultados. Apesar dos modelos que apenas contemplam a dimensão do trabalho em auditoria nos anos N e N-1 serem aqueles com resultados mais baixos face aos restantes (modelo A e C, respetivamente), é importante ter em conta que têm uma PECC de validação cruzada de aproximadamente 90%. Estas conclusões permitem validar a primeira hipótese de investigação, pois pode concluir-se que o trabalho efetuado pela AFE é um bom preditor das insolvências.

Os resultados obtidos evidenciam que o modelo D é o que têm maior capacidade preditiva, uma vez que apresenta o melhor resultado de PECC da validação cruzada, e também boas métricas na especificidade e sensibilidade (PECC = 92,6%; especificidade = 89,4%; sensibilidade: 93,6%).

Tabela 4.21 - Resultados dos modelos preditivos das insolvências

	Modelos					
Métricas	A	B	C	D	E	F
Complexidade						
Profundidade	5	5	5	5	5	5
Nº total de nós	29	33	47	31	43	47
Nº total de nós terminais	15	17	24	16	22	22
Qualidade						
Treino						
PECC	91,1%	93,4%	91,5%	93,3%	93,5%	93,4%
Especificidade	90,5%	95,3%	86,9%	91,7%	91,8%	93,7%
Sensibilidade	91,7%	91,5%	96,1%	94,9%	95,2%	93,1%
Teste						
PECC	90,9%	93,3%	90,9%	91,5%	92,1%	92,4%
Especificidade	89,9%	93,9%	89,4%	89,4%	89,8%	92,2%
Sensibilidade	91,8%	92,6%	92,4%	93,6%	94,4%	92,6%
Validação cruzada - PECC	89,9%	92,3%	89,1%	92,6%	92,0%	90,8%

De forma a concluir quanto às variáveis com maior capacidade preditiva de cada modelo estimado, a tabela 4.22 apresenta o TOP 3 das variáveis com maior importância para cada modelo. Os resultados evidenciam que as variáveis que surgem com maior frequência são as variáveis da dimensão do trabalho realizado em AFE, nomeadamente, Tipo_opinião_N, Tipo_opinião_N1, Comentário_N e Comentário_N1. Importa referir que, quando a dimensão das variáveis da auditora é incluída nos modelos, tanto em N como em N1, a variável da Dim_auditora aparece no TOP 3 das variáveis mais importantes, substituindo a variável da Dim_comentário, como se pode observar nos modelos B e D.

De facto, tendo em conta o ano N, quando as características da auditora se juntam ao trabalho efetuado pela AFE, a capacidade dos modelos preditivos aumenta, embora ligeiramente, de 89,9% (modelo A) para 92,3% (modelo B). Já no que respeita o ano N-1 pode observar-se as mesmas conclusões, sendo que a capacidade preditiva do modelo aumenta de 89,1% (modelo C) para 92,6% (modelo D). Estes resultados, permitem validar quer, tanto a primeira, como a segunda hipótese de investigação

Tabela 4.22 - TOP 3 das variáveis com maior importância nos modelos preditivos

Modelos	TOP 3 das variáveis mais importantes		
A	Tipo_opinião_N1	Comentário_N1	Dim_comentário_N1
B	Comentário_N1	Tipo_opinião_N1	Dim_auditora_N1
C	Tipo_opinião_N	Comentário_N	Dim_comentário_N
D	Tipo_opinião_N	Comentário_N	Dim_auditora_N
E	Tipo_opinião_N	Comentário_N1	Tipo_opinião_N1
F	Tipo_opinião_N	Comentário_N1	Tipo_opinião_N1

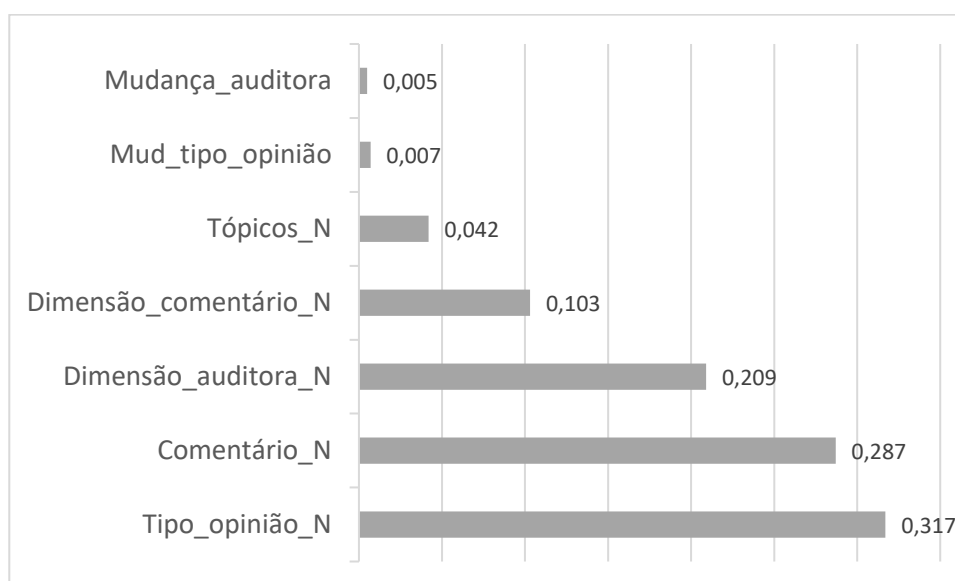
Sendo o modelo D o que apresenta melhor resultado como se evidencia anteriormente, a figura 4.8 apresenta o nível de importância de todas as variáveis utilizadas neste modelo. Os resultados mostram que tanto as variáveis do trabalho realizado em AFE, como as variáveis da auditora, são bons preditores das insolvências. Na primeira dimensão destaca-se o Tipo_opinião que é a variável com maior capacidade preditiva e que melhor explica a insolvência das empresas e, também, o Comentário_N. Na segunda dimensão, a variável Dim_auditora também têm um papel importante na previsão da insolvências das empresas. Estas conclusões corroboram estudos anteriores que evidenciam que tanto a opinião do auditor como a dimensão da auditora são duas variáveis importantes na predição das insolvências (Muñoz-Izquierdo *et al.*, 2020; Peursem & Chan, 2014).

Contudo, não se pode concluir que as empresas não *Big 4* estão associadas às empresas insolventes pelo facto de estas não emitirem qualquer comentário na CLC/RAFE, pois dos resultados obtidos a grande maioria das empresas em insolvências têm uma qualificação da opinião.

Importa referir que, a variável Tópico_N, criada através do *text mining*, têm uma importância mais reduzida face às outras variáveis. Deste modo, e apesar da primeira parte da terceira hipótese de investigação ser validada, uma vez que a variável Tópico está estatisticamente relacionada com o *status* da empresa para um nível de significância de 0,05, não é possível validar a segunda parte desta hipótese de investigação, pois esta variável não tem uma importância significativa na predição das insolvências comparativamente com as restantes variáveis.

Importa referir também, que a mudança de auditora entre o ano N-1 e N não é um bom preditor das insolvências, sendo mesmo considerada a variável com menor importância do modelo D. Estas conclusões não estão de acordo com o estudo de Cheng *et al.* (2007) que indica que a mudança de auditor permite diferenciar as empresas insolventes das ativas, pois as primeiras tendem a mudar de auditor quando se aproximam do ano da insolvência.

Figura 4.8 – Importância das variáveis do modelo preditivo D



Com o intuito de se perceber com maior detalhe a importância das variáveis, a tabela 4.23 diferencia o *status* previsto do modelo, segundo as diferentes categorias das três variáveis consideradas mais importantes no modelo D. Os resultados obtidos evidenciam que, para a dimensão das variáveis do trabalho em AFE, o modelo prevê que para as empresas com opinião não qualificada, 98% sejam ativas e 2% insolventes. Por outro lado, prevê-se que para as empresas com opinião qualificada 10% sejam ativas e 90% insolventes. No que diz respeito à variável comentário, o modelo preditivo prevê que para as empresas que não têm comentário 100% sejam ativas e 0% insolventes. Já para as empresas que têm comentário, o modelo prevê que 15% sejam ativas e 85% insolventes.

No que concerne à variável da dimensão da auditora, o modelo tende a prever que as empresas auditadas por uma firma *Big 4* se mantenham ativas (100%), enquanto para as empresas auditadas por uma não *Big 4* o modelo já prevê algumas insolventes (28%). Note-se que a falta de informação do tipo de auditora nas empresas insolventes pode ter influenciado e predição do mesmo.

Tabela 4.23 - Distribuição do *status* previsto para as categorias das variáveis mais importantes.

Variáveis	Descrição	Status previsto			
		Ativa		Insolvente	
		Empresas	%	Empresas	%
Tipo_opinião_N	Opinião não qualificada	874	98,0	16	2,0
	Opinião qualificada	112	10,0	1038	90,0
Dim_auditora_N	Auditada por não <i>Big 4</i>	452	72,0	178	28,0
	Auditada por <i>Big 4</i>	520	100,0	0	0,0
Comentário_N	Há comentário	188	15,0	1054	85,0
	Não há comentário	798	100,0	0	0,0

4.4.2. Perfis associados às empresas ativas e insolventes

De forma a identificar perfis de CLC/ RAFE associados às empresas ativas e insolventes, recorre-se à análise dos nós terminais com elevada confiança e suporte relevante da árvore de classificação proveniente do modelo D, melhor modelo (Anexo F).

Posto isto, o perfil de CLC/ RAFE relacionado com as empresas ativas pode ser obtido pela análise da regra de decisão:

- Se (Tipo_opinião_N = não qualificada) e (Dim_auditora_N = Big 4) e (Dim_comentário_N \leq 812,5) então *status* = ativa (nó 19; suporte = 416 e confiança de 0,971)

Este é o perfil associado às empresas ativas que melhora conjuga o suporte e a confiança. Já para o perfil de CLC/RAFE relacionado com as empresas insolventes tem-se, por exemplo:

- Se (Tipo_opinião_N = qualificada) e Dim_auditora_N = não Big 4), e (Tópicos_N = tópico 3 ou Tópicos_N = tópico 4 ou Tópicos_N = 2), e (Dim_comentário_N \leq 1356 e Dim_comentário_N $>$ 283), então *status* = insolvente (nó 22; suporte = 982 e confiança = 0,936).

Tal como nas empresas ativas, este é o perfil mais associado às insolvências tendo em conta o suporte e a confiança do nó.

Dos resultados obtidos pode concluir-se que a dimensão da auditora e o tipo de opinião na CLC/ RAFE são dos atributos que melhor diferencia as empresas insolventes e ativas. Embora na análise descritiva univariada, o valor médio e mediano do número de caracteres ser superior nos comentários das empresas ativas comparativamente com as empresa insolventes, note-se que empresas com relatórios mais extensos estão mais propensas à insolvência, o que vai de encontro com Neves (2012). Aqui pode observar-se uma das vantagens da utilização do método das árvores de decisão, em que a conjugação da variável da dimensão dos comentários com outras variáveis permite diferenciar as empresas ativas e insolventes.

Os tópicos 3 investimento e imparidade e 2 gastos e ativo também estão associadas às insolvências quando conjugadas com as restantes variáveis. De facto, Muñoz-Izquierdo *et al.* (2019) conclui que as empresas com comentários associados a investimentos e gastos estão mais associadas às insolvências.

Globalmente, e tendo em conta outras investigações que estudam de forma intensiva o tema da predição das insolvências tendo como preditores a AFE, nomeadamente, o estudo Muñoz-Izquierdo *et al.* (2020), cujo modelo atinge uma capacidade preditiva de 84% (classifica de forma correta 79% das ativas e 89% das insolventes), pode concluir que o objetivo da criação de um modelo preditivo na presente investigação é alcançado.

5. Conclusões

Em Portugal, a partir de 2008, ano do começo da crise financeira, o número de insolvências de empresas aumentou significativamente e, desde então, estes valores têm se mantido elevados face aos anos anteriores à crise. Estas insolvências têm impactos irreversíveis, não só nas empresas que entram em insolvência mas, também, nos diferentes *stakeholders*, nomeadamente, clientes, fornecedores, credores, investidores e nos próprios colaboradores, gerentes e proprietários das empresas. De uma forma geral, as insolvências das empresas afetam a economia nacional como um todo.

Tendo em conta esta problemática, o presente estudo identifica uma relação entre o trabalho em AFE e as características da auditora e a insolvência das empresas. A CLC/RAFE é a única forma de comunicação entre os auditores e os utilizadores da informação e é onde todas as conclusões do trabalho realizado são mencionadas. De forma a identificar esta relação recorreu-se a técnicas de análise de dados assistidas por computador, tais como *text mining* e árvores de decisão, de forma a determinar perfis de empresas associados às insolvências.

Sendo o objetivo deste estudo gerar conhecimento que permita analisar de que forma a auditoria financeira contribui para a predição das insolvências das empresas, e com base nos resultados obtidos, pode concluir-se que os objetivos delineados para esta investigação foram cumpridos e, à exceção de uma parte da terceira hipótese de investigação, as duas outras foram validadas. De facto, o trabalho em AFE e as características da auditora estão relacionados com o *status* das empresas e que a CLC/RAFE apresenta características que diferem as empresas ativas das empresas insolventes.

Do modelo baseado em árvores de decisão com o algoritmo CART, pode concluir-se que a variável do tipo de opinião é a mais importante na predição das insolvências. Outra variável com importância significativa é a variável comentário, uma vez que a presença de comentários na CLC/RAFE está mais associada às empresas insolventes. Para além destas, comprovou-se que a dimensão da auditora é, também, bastante relevante na predição das insolvências, uma vez que, quando inserida nos modelos aumenta a capacidade preditiva dos mesmos.

Apesar da identificação, através da técnica de *text mining*, de termos mais correlacionados e que tendem a ser salientados pelo auditor nos comentários das empresas insolventes comparativamente com as ativas, constatou-se que a variável tópico não tem uma importância significativa na predição das insolvências face às variáveis referidas anteriormente.

Desta forma e, bastando apenas a informação do ano anterior à insolvência, conclui-se que estão mais associadas às insolvências as empresas auditadas por firmas que pertencem ao grupo das não *Big 4*, cuja CLC/RAFE emitida tem uma opinião qualificada e comentários que tendem a ser mais extensos, sendo estes associados aos tópicos de investimento e imparidade, e também gastos e ativo.

Neste sentido, importa concluir que o trabalho em AFE e as características da auditora contribuem de forma relevante para a predição do *status* das empresas e disponibilizam informação importante para se preverem as insolvências das empresas. Desta forma, responde-se claramente à questão de investigação.

5.1. Contributos

O presente estudo vem contribuir tanto para a literatura como para as empresas e profissionais. No que diz respeito à literatura e ao estudo empírico, esta investigação pretende aumentar o grau de conhecimento no âmbito do tema dos modelos preditivos das insolvências através da utilização do trabalho em AFE e das características da auditora como preditores. De facto, a informação contida na CLC/RAFE é pouco utilizada na predição das insolvências, apesar do enorme potencial em diferenciar empresas ativas e insolventes. Outro contributo é a utilização de técnicas inovadoras de *text mining*, nunca antes utilizadas, no tratamento dos dados para construção dos modelos preditivos das insolvências, tendo o trabalho em AFE como preditor. Para além destes, o presente estudo também contribui para o conhecimento do impacto que a pandemia COVID-19 têm nas CLC/RAFE e o que o auditor tende a mencionar relativamente a este tema recente de enorme relevância.

Para os gestores das empresas, este modelo permite alertar, atempadamente, e através da CLC/RAFE para problemas de viabilidade, de modo que estes consigam delinear estratégias para evitar as insolvências. Por outro lado, no que diz respeito aos utilizadores da informação, nomeadamente os investidores e credores, este modelo preditivo permite-lhes identificar as empresas com propensões em entrar em insolvência de forma a tomarem melhores decisões de investimento e financiamento. Para além disso, o trabalho da AFE passa a ganhar uma maior visibilidade e relevância por parte dos utilizadores da informação.

Em suma, este estudo releva a complementaridade de duas atividades fundamentais para as empresas, a da contabilidade e da auditoria, que se exercidas de forma responsável podem trazer mais e melhor conhecimento sobre a realidade económica/financeira das empresas e, assim, detetar precocemente problemas de inviabilidade das empresas.

5.2. Limitações e pistas para investigações futuras

Importa evidenciar que a falta de informação, nomeadamente, dos comentários emitidos pelo auditor nas empresas insolventes é a principal limitação desta investigação. Por um lado, no que respeita a análise da identificação dos tópicos salientes presentes nos comentários através da técnica de *text mining*. Por outro, esta falta de informação também poderá ter implicações na variável comentário, uma vez que não se consegue perceber se uma empresa com opinião não qualificada não tem qualquer comentário na CLC/RAFE porque de facto o auditor não o mencionou ou se é a base de dados que não

o divulga. Para além disto, a falta de informação do nome da auditora também limita a análise efetuada.

No que diz respeito a investigações futuras e de forma a colmatar as limitações evidenciadas anteriormente, sugere-se a realização do mesmo estudo, mas com um menor número de não respostas e aplicado a empresas de outros países. Para além disso, também será interessante a adoção de outras variáveis resultantes do trabalho em AFE, tais como, o número de comentários emitidos e a natureza dos mesmos, de forma a perceber se a inclusão destas variáveis aumenta a capacidade preditiva do modelo.

Outra das pistas futuras é analisar a evolução dos impactos da pandemia COVID-19 nas empresas e no trabalho da auditoria.

Bibliografia

- Abrantes, C. (2020). *Os modelos preditivos do sucesso de candidaturas a fundos europeus: o papel da manipulação de resultados*. [Dissertação de mestrado, Iscte – Instituto Universitário de Lisboa]. Repositório Iscte. <http://hdl.handle.net/10071/21827>
- Altman, E. I. (1968). Financial ratios, discriminant analysis, and the prediction of corporate bankruptcy. *The Journal of Finance*, 23(4), 589. <https://doi.org/10.2307/2978933>
- Altman, E. I., Iwanicz-Drozowska, M., Laitinen, E. K., & Suvas, A. (2016). Financial and non-financial variables as long-horizon predictors of bankruptcy. *Journal of Credit Risk*, 12(4), 60-90. <https://doi.org/10.2139/ssrn.2669668>
- Altman, E. I., & Sabato, G. (2013). Modelling credit risk from SMEs: Evidence from the US Market. *Abacus*, 43(6), 332–357. https://doi.org/10.1142/9789814417501_0009
- Altman, E. I., Sabato, G., & Wilson, N. (2010). The value of non-financial information in SME risk management. *The Journal of Credit Risk*, 6(2), 95–127. <https://doi.org/10.21314/jcr.2010.110>
- Assis, A. S. (2018). *Recuperação financeira de empresas*. [Dissertação de mestrado, Faculdade de economia da Universidade do Porto]. Repositório aberto. <https://hdl.handle.net/10216/117195>
- Bauweraerts, J. (2016). Predicting bankruptcy in private firms: Towards a stepwise regression procedure. *International Journal of Financial Research*, 7(2). <https://doi.org/10.5430/ijfr.v7n2p147>
- Beaver W.H. (1966). Financial ratios as predictors of failure. *Journal of Accounting Research*, 4, 71–111. <https://doi.org/10.2307/2490171>
- Berglund, N. R. (2020). Do client bankruptcies preceded by clean audit opinions damage auditor reputation? *Contemporary Accounting Research*, 37(3), 1914–1951. <https://doi.org/10.1111/1911-3846.12575>
- Boone, J. P., Khurana, I. K., & Raman, K. K. (2010). Do the big 4 and the second-tier firms provide audits of similar quality? *Journal of Accounting and Public Policy*, 29(4), 330–352. <https://doi.org/10.1016/j.jaccpubpol.2010.06.007>
- Cabete, N. P., Cardoso, M. G. M. S. (2006). Algoritmo cart : previsão do desempenho na matemática do secundário. *Revista de Ciências da Computação*, 1(1), 27-55. <https://doi.org/10.34627/rcc.v1i0.50>
- Carson, E., Fargher, N. L., Geiger, M. A., Lennox, C. S., Raghunandan, K., & Willekens, M. (2013). Audit reporting for going-concern uncertainty: A research synthesis. *Auditing*, 32(1), 353–384. <https://doi.org/10.2308/ajpt-50324>
- Chen, K. C. W., & Church, B. K. (1996). Going concern opinions and the market's reaction to bankruptcy filings. *Accounting Review*, 71(1), 117–128.

- Chen, P. F., He, S., Ma, Z., & Stice, D. (2016). The information role of audit opinions in debt contracting. *Journal of Accounting and Economics*, 61(1), 121–144. <https://doi.org/10.1016/j.jacceco.2015.04.002>
- Chen, S., Su, X., & Wang, Z. (2005). An Analysis of auditing environment and modified audit opinions in China: Underlying reasons and lessons. *International Journal of Auditing*, 9(3), 165–185. <https://doi.org/10.1111/j.1099-1123.2005.255.x>
- Cheng, J. H., Yen, C. H., & Chiu, Y. W. (2007). Improving business failure predication using rough sets with non-financial variables. *International Conference on Adaptive and Natural Computing Algorithms*, 4431(1), 614–621. https://doi.org/10.1007/978-3-540-71618-1_68
- Constantino, J. (2020). *Gestão de resultados como preditor de insolvências: Evidência nas empresas portuguesas*, [Dissertação de Mestrado, Iscte – Instituto Universitário de Lisboa]. Repositório Iscte. <http://hdl.handle.net/10071/22264>
- COSEC – Companhia de Seguro de Créditos (2020, junho). *Empresas insolventes no primeiro semestre representam volume de negócios superior a €550 milhões*. https://www.cosec.pt/media/2144/200716_empresas-insolventes-representam-volume-de-neg%C3%B3cios-superior-a-500m.pdf
- DeFond, M. L., & Zhang, J. (2014). A Review of archival auditing research. *Journal of Accounting and Economics*, 58(1), 275–326. <http://dx.doi.org/10.1016/j.jacceco.2014.09.002>
- Direção-Geral da Política da Justiça (2020, julho). *DESTAQUE ESTATÍSTICO TRIMESTRAL – 1.º TRIMESTRE DE 2020*. https://estatisticas.justica.gov.pt/sites/siej/pt-pt/Destaques/20200731_D78_FalenciasInsolvencias_2020_T1.pdf
- Dodd, P., Dopuch, N., Holthausen, R., Leftwich, R. (1984). Qualified audit opinions and stock prices: Information content, announcement dates, and concurrent disclosures. *Journal of Accounting and Economics*, 6(1), 3–38. [https://doi.org/10.1016/0165-4101\(84\)90018-1](https://doi.org/10.1016/0165-4101(84)90018-1)
- Freire, M. (2016). O Reporte do Auditor. *Cadernos Do Mercado de Valores Mobiliários*. <https://www.cmvm.pt/pt/EstatisticasEstudosEPublicacoes/CadernosDoMercadoDeValoresMobilarios/Documents/O%20Reporte%20do%20Auditor%20-%20M%C3%A1rio%20Freire.pdf>
- Geiger, M. A., & Rama, D. v. (2006). Audit firm size and going-concern reporting accuracy. *Accounting Horizons*, 20(1), 1–17. <https://doi.org/10.2308/acch.2006.20.1.1>
- IAABS – International Auditing and Assurance Standards Board. (2015). *Manual das normas internacionais de controlo de qualidade, auditoria, revisão, outros trabalhos de garantia de fiabilidade e serviços relacionados*. Ordem dos revisores oficiais de contas
- IBM - International Business Machines Corporation (2020, junho). *Text mining*. <https://www.ibm.com/cloud/learn/text-mining>

- INE – Instituto Nacional de Estatística. (2018).
https://www.ine.pt/xportal/xmain?xpgid=ine_main&xpid=INE&xlang=pt
- Jacinto, S. M. (2019). *Determinantes da auditoria voluntária nas PME portuguesas*. [Dissertação de Mestrado, Iscte – Instituto Universitário de Lisboa]. Repositório Iscte.
<http://hdl.handle.net/10071/19432>
- Jardin, P. (2015). Bankruptcy prediction using terminal failure processes. *European Journal of Operational Research*, 242(1), 286-303. <https://doi.org/10.1016/j.ejor.2014.09.059>
- Laureano, R. (2020). *Testes de Hipóteses e Regressão: O Meu Manual de Consulta Rápida* (2ª Ed.) Edições Sílabo.
- Lopes, I. T. (2019). *Auditoria financeira do controlo interno ao controlo externo independente*. Edições Almedina.
- Lukason, O., & Hoffman, R. C. (2014). Firm bankruptcy probability and causes: An Integrated study. *International Journal of Business and Management*, 9(11), 80-91.
<https://doi.org/10.5539/ijbm.v9n11p80>
- Major, M. (2009). Reflexão sobre a investigação em contabilidade de gestão. *Revista de Gestão dos Países da Língua Portuguesa*, 8(1), 43–50.
- Major, M. J. (2017). O positivismo e a pesquisa “alternativa” em contabilidade. *Revista Contabilidade & Finanças*, 28(74), 173–178.
- MAZARS. (2016). *Overview of the new auditor’s audit report under the ISAs*.
<http://www.mazars.pt/Home/Noticias/Publicacoes/Abordagem-geral-ao-novo-relatorio-de-auditoria>
- Menon, K., & Williams, D. D. (2010). Investor reaction to going concern audit reports. *Accounting Review*, 85(6), 2075–2105. <https://doi.org/10.2308/accr.2010.85.6.2075>
- Muñoz-Izquierdo, N., Camacho-Miñano, M. D. M., Segovia-Vargas, M. J., & Pascual-Ezama, D. (2019). Is the external audit report useful for bankruptcy prediction? Evidence using artificial intelligence. *International Journal of Financial Studies*, 7(2), 1-23. <https://doi.org/10.3390/ijfs7020020>
- Muñoz-Izquierdo, N., Camacho-Miñano, M.-M., & Pascual-Ezama, D. (2017). The content of the audit report in the year prior to bankruptcy filing. *Spanish Journal of Finance and Accounting*, 46(1), 92–126. <https://doi.org/10.1080/02102412.2016.1245886>
- Muñoz-Izquierdo, N., Laitinen, E. K., Camacho-Miñano, M. del 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 and Accounting*, 31(1), 65–97. <https://doi.org/10.1111/jifm.12110>
- Neves, J. C. (2015). *Análise e Relato Financeiro* (8ª Ed.). Texto Editores.

- Ohlson, J. A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*, 18(1), 109. <https://doi.org/10.2307/2490395>
- Sánchez, C. P., Monelos, P. de L., & López, M. R. (2013). A parsimonious model to forecast financial distress, based on audit evidence. *Contaduría y Administración*, 58(4), 151–173. [https://doi.org/10.1016/s0186-1042\(13\)71237-3](https://doi.org/10.1016/s0186-1042(13)71237-3)
- Santos, M. R. C., Laureano, R. M. S., & Moro, S. (2020). Unveiling research trends for organizational reputation in the Nonprofit Sector. *International Journal of Voluntary and Nonprofit Organizations*, 31(1), 56–70. <https://doi.org/10.1007/s11266-018-00055-7>
- Schneider, A. (2018). Studies on the impact of accounting information and assurance on commercial lending judgments. *Journal of Accounting Literature*, 41(2), 63–74. <https://doi.org/10.1016/j.acclit.2018.03.005>
- Pestana, H. M. & Gageiro, J. N. (2014). *Análise categórica, árvores de decisão e análise de conteúdo em ciências sociais e da saúde com o SPSS* (6ª ed.). Lisboa: Edições Sílabo.
- Peurseem, K., & Chan, Y. C. (2014). Forecasting New Zealand corporate failures 2001-10: Opportunity lost? *Australian Accounting Review*, 24(3), 276–288. <https://doi.org/10.1111/auar.12029>
- PORDATA - Estatísticas, gráficos e indicadores de Municípios, Portugal e Europa. (2018 e 2020). <http://www.pordata.pt/>
- Tahinakis, P. and Samarinas, M. (2016) The incremental information content of audit opinion. *Journal of Applied Accounting Research*, 17, 139-169. <https://doi.org/10.1108/JAAR-01-2013-0011>
- Zorio-Grima, A., & Carmona, P. (2019). Narratives of the big-4 transparency reports: country effects or firm strategy? *Managerial Auditing Journal*, 34(8), 951–985. <https://doi.org/10.1108/MAJ-09-2018-1994>

Anexos

Anexo A - Medidas descritivas dos indicadores financeiros

Variável	Nº	Média	Desvio padrão	Mínimo	Mediana	Máximo
EBITDA_N ⁽¹⁾	2022	2858,01	9746,33	-33820,49	54,56	90341,57
EBITDA_N1 ⁽¹⁾	1990	2974,44	10620,32	-50911,91	88,05	76445,37
Solvabilidade_N	2016	276,04	1908,30	-100,00	21,10	21676,89
Solvabilidade_N1	1986	548,27	5161,76	-99,97	24,40	80508,35
Endividamento_N	2014	197,24	981,70	0,06	82,40	19658,15
Endividamento_N1	1986	180,57	1288,88	0,01	80,38	24749,18
AT_N ⁽¹⁾	2019	37903,67	115344,56	0,10	3820,38	816643,46
AT_N1 ⁽¹⁾	1990	38017,99	114429,77	0,01	3956,29	788288,34
Vendas_N ⁽¹⁾	1867	17860,72	50487,23	0,10	2792,46	370246,29
Vendas_N1 ⁽¹⁾	1853	17449,24	49403,18	0,17	2970,94	360364,55

Nota: (1) A unidade de medida é milhares de euros (x1000 euros)

Anexo B - Relação entre a *performance* financeira e o *status* da empresa

Variável	Status	Média	Desvio padrão	Teste t Eta / Eta ²
EBITDA_N ⁽¹⁾	Ativa	5854,79	13043,81	t(1026,3) = 14,7; p = 0,001 Eta = 0,997; Eta ² = 0,993
	Insolvência	-168,55	969,26	
EBITDA_N1 ⁽¹⁾	Ativa	5939,66	14165,09	t(1041,4) = 13,5; p = 0,001 Eta = 0,998; Eta ² = 0,996
	Insolvência	-118,64	1580,54	
Solvabilidade_N	Ativa	544,31	2658,07	t(1020,6) = 6,5; p = 0,001 Eta = 0,981; Eta ² = 0,962
	Insolvência	3,47	138,36	
Solvabilidade_N1	Ativa	1009,59	7121,76	t(1056,7) = 4,2; p = 0,001 Eta = 0,976; Eta ² = 0,952
	Insolvência	65,07	997,92	
Endividamento_N	Ativa	67,83	128,88	t(1014,2) = -6,0; p = 0,001 Eta = 0,980; Eta ² = 0,960
	Insolvência	329,00	1376,41	
Endividamento_N1	Ativa	68,06	130,84	t(978,4) = -3,9; p = 0,001 Eta = 0,966; Eta ² = 0,932
	Insolvência	298,41	1832,45	
AT_N ⁽¹⁾	Ativa	73039,47	154619,72	t(1019,9) = 14,6; p = 0,001 Eta 0,975= ; Eta ² = 0,942
	Insolvência	2312,47	7577,71	
AT_N1 ⁽¹⁾	Ativa	71710,93	152176,91	t(1032,2) = 14,4; p = 0,001 Eta = 0,968 ; Eta ² = 0,933
	Insolvência	2872,18	13700,68	
Vendas_N ⁽¹⁾	Ativa	32552,04	65916,16	t(1001,7) = 14,7; p = 0,001 Eta = 0,999; Eta ² = 0,997
	Insolvência	1453,76	5923,96	
Vendas_N1 ⁽¹⁾	Ativa	31333,58	64437,12	t(1011,9) = 14,2; p = 0,001 Eta = 0,999; Eta ² = 0,997
	Insolvência	1863,15	7883,04	

Nota: (1) A unidade de medida é milhares de euros (x1000 euros)

Anexo C – Dicionário dos termos relevantes (continua)

Termos	Termos similares
desacordo	desacordo
limitação de âmbito	limitação de âmbito
continuidade	continuidade
plano especial de revitalização	plano especial de revitalização, Processo Especial de Revitalização, Revitalização
regime Extrajudicial de Recuperação de Empresas	regime extrajudicial de recuperação de empresas, sistema de Recuperação de Empresas Via Extrajudicial, SIREVE,
processo judicial	processo judicial, processos judiciais, contencioso
insolvência	insolvência
artigo 35º - Código das Sociedades Comerciais - perda de metade do capital social	art.º 35.º, artigo 35.º, artigos 35.º, 35.º, artigo 35º, artigo 35, artigos 35º, artigos 35
liquidez	liquidez
risco	risco, riscos
controlo Interno	controlo interno
materialidade	materialidade, material
distorção	distorção, distorções
covid	covid, pandemia, pandémica, surto
caixa	caixa, depósitos à ordem, depósitos bancários
devedores	clientes, contas a receber de clientes, rúbrica de clientes, dividas de terceiros, devedores, devedor, saldos a receber
credores	fornecedores, contas a pagar de fornecedores, rúbrica de fornecedores, credores, credor
impostos diferidos	impostos diferidos
estado e outros entes públicos	Estado e outros entes públicos, EOEP
financiamento	financiamentos, financiamento
acionistas	accionistas, sócios, acionista, accionista, acionistas
diferimentos	diferimentos
provisões	provisões, provisão
inventário	inventário, compras, mercadorias, matérias-primas, existências, compras, mercadorias
investimento	investimentos, investimentos
método de equivalência patrimonial	método de equivalência patrimonial
ativos biológicos	ativos biológicos, ativo biológico, activos biológicos, activo biológico
goodwill	goodwill
ativo	ativo, activo, bens do património

(continua na página seguinte)

Anexo C – Dicionário dos termos relevantes (continuação)

Termos	Termos similares
capital Próprio	capital próprio, capital social, capitais próprios
resultados transitados	resultados transitados
passivo	passivo
gastos	gastos, custos, CMVMC, fornecimentos e serviços externos, FSE
peçoal	peçoal, gastos com o peçoal, remunerações, recursos humanos
depreciação	depreciação, depreciações, depreciado
amortização	amortização
imparidade	imparidade, imparidades
acrécimos	acrécimos
rendimentos	rendimentos, vendas, prestações de serviços, prestação de serviços, prestações de serviços
reversões	reversões, reverses
resultados	resultado líquido, resultado do período
mensuração	mensuração, valorização, revalorização
vida útil	vida útil, vidas úteis
locações	locação, locações
participações financeiras	participações financeiras, participação financeira
contas consolidadas	contas consolidadas
contas individuais	contas individuais

Fonte: Elaboração própria

Anexo D – Frequência absoluta dos termos nas subamostras das empresas ativas e insolventes

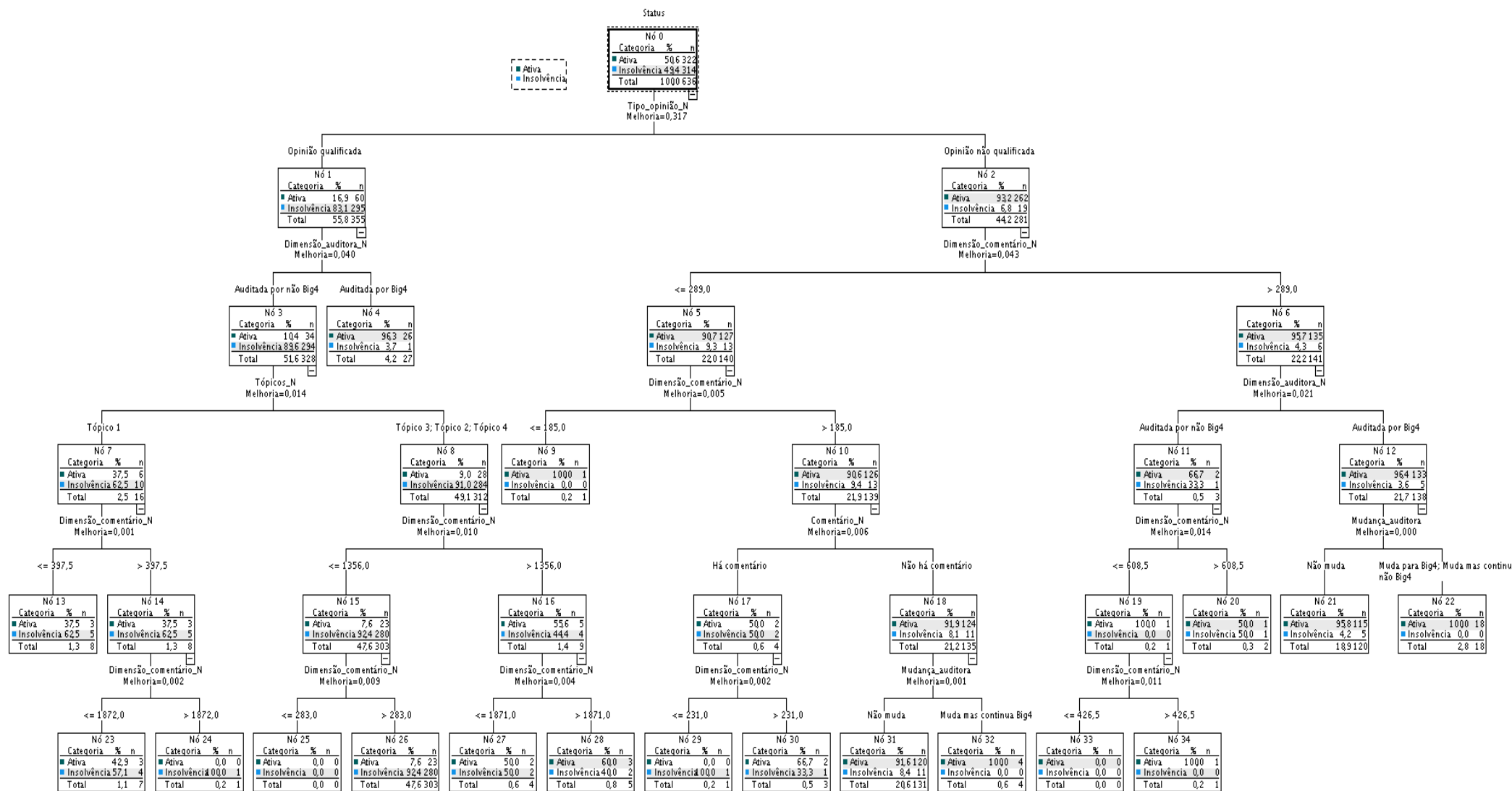
ID	Termos	Frequência absoluta
1	gasto	91
2	mensuração	81
3	imparidade	70
4	caixa	63
5	inventário	58
6	investimento	56
7	devedor	52
8	resultado líquido	44
9	ativo	44
10	rendimento	35
11	depreciação	34
12	credor	33
13	continuidade	32
14	passivo	29
15	participação financeira	27
16	capital	27
17	covid-19	24
18	resultado transitado	21
19	vida útil	19
20	risco	19
21	impostos diferidos	17
22	acionista	16
23	provisão	15
24	materialidade	14
25	financiamento	13
26	goodwill	12
27	MEP	12
28	limitação âmbito	9
29	art35	8
30	controlo interno	6
31	distorção	6
32	amortização	4
33	processos judiciais	4
34	reversão	3
35	diferimento	3
36	desacordo	2
37	liquidez	2
38	insolvência	2
39	PER	2
40	acréscimo	2
41	contas consolidadas	1
42	peçoal	1
43	locação	1

ID	Termos	Frequência absoluta
1	continuidade	117
2	ativo	101
3	devedor	95
4	imparidade	76
5	inventario	72
6	mensuração	50
7	credor	47
8	acionista	45
9	art35	44
10	passivo	39
11	gasto	39
12	financiamento	36
13	rendimento	35
14	PER	34
15	insolvência	32
16	materialidade	30
17	caixa	25
18	impostos diferidos	23
19	capital	21
20	investimento	19
21	resultado líquido	19
22	risco	15
23	provisão	14
24	resultado transitado	13
25	covid-19	12
26	peçoal	9
27	participação financeira	8
28	acréscimo	8
29	processos judiciais	8
30	amortização	7
31	MEP	7
32	goodwill	5
33	reversão	4
34	liquidez	3
35	diferimento	3
36	limitação de âmbito	3
37	vida útil	2
38	depreciação	2
39	desacordo	2
40	controlo interno	2
41	RERE	2
42	eoep	2
43	distorção	1

Fonte: Elaboração própria

Nota: A primeira tabela diz respeito às empresas ativas e a segunda às empresas insolventes.

Anexo E - Árvore de decisão para a amostra de teste



Anexo F - Árvore de decisão para a amostra de treino

