

## Repositório ISCTE-IUL

---

Deposited in *Repositório ISCTE-IUL*:

2022-08-29

Deposited version:

Accepted Version

Peer-review status of attached file:

Peer-reviewed

Citation for published item:

Ilhéu, J. M., Laureano, R. M. S. & Santos, M. R. C. (2022). Modelo preditivo das insolvências: O contributo da auditoria financeira externa. In Rocha, A., Bordel, B., Penalvo, F. G., and Goncalves, R. (Ed.), 2022 17th Iberian Conference on Information Systems and Technologies (CISTI). Madrid: IEEE.

Further information on publisher's website:

10.23919/CISTI54924.2022.9820058

Publisher's copyright statement:

This is the peer reviewed version of the following article: Ilhéu, J. M., Laureano, R. M. S. & Santos, M. R. C. (2022). Modelo preditivo das insolvências: O contributo da auditoria financeira externa. In Rocha, A., Bordel, B., Penalvo, F. G., and Goncalves, R. (Ed.), 2022 17th Iberian Conference on Information Systems and Technologies (CISTI). Madrid: IEEE., which has been published in final form at <https://dx.doi.org/10.23919/CISTI54924.2022.9820058>. This article may be used for non-commercial purposes in accordance with the Publisher's Terms and Conditions for self-archiving.

---

### Use policy

Creative Commons CC BY 4.0

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a link is made to the metadata record in the Repository
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

---

# Modelo preditivo das insolvências: o contributo da auditoria financeira externa

## *Predicting insolvencies: the contribution of the external financial audit*

João M. Ilhéu

Instituto Universitário de Lisboa  
ISCTE-IUL  
Av. das Forças Armadas, Lisboa  
1649-026 Lisboa, Portugal  
[joao\\_miguel\\_ilheu@iscte-iul.pt](mailto:joao_miguel_ilheu@iscte-iul.pt)

Raul M. S. Laureano

Instituto Universitário de Lisboa  
ISCTE-IUL, BRU, ISTAR  
Av. das Forças Armadas, Lisboa  
1649-026 Lisboa, Portugal  
[raul.laureano@iscte-iul.pt](mailto:raul.laureano@iscte-iul.pt)

Márcia R. C. Santos

Instituto Politécnico de Setúbal  
Escola Superior de Ciência Empresariais  
Instituto Universitário de Lisboa  
ISCTE-IUL, BRU, ISTAR  
Av. das Forças Armadas, Lisboa  
1649-026 Lisboa, Portugal  
[marcia\\_rafaela\\_santos@iscte-iul.pt](mailto:marcia_rafaela_santos@iscte-iul.pt)

**Resumo** — Em Portugal, desde a crise financeira, o número de Pequenas e Médias Empresas (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 das características do auditor e o conteúdo da sua opinião 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 Certificações Legais de Contas (CLC)/Relatórios de Auditoria Financeira Externa (RAFE), entre os anos de 2016 e 2020, de uma amostra de 2.040 empresas, 1.020 não insolventes e 1.020 insolventes. Os resultados obtidos permitem identificar uma relação entre as características do auditor e o conteúdo da sua opinião e a insolvência das empresas, prevendo-se uma Percentagem de Exemplos Corretamente Classificados (PECC) de 93%. O principal contributo empírico desta investigação é gerar melhor conhecimento sobre a inviabilidade das empresas através da atividade de Auditoria Financeira Externa (AFE), recorrendo-se a novas técnicas nunca antes utilizadas em modelos preditivos.

**Palavras Chave** – *Insolvência, Auditoria Financeira Externa, Text Mining, Árvores de decisão, Previsão.*

**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 characteristics of the audit firm and audit opinion content in predicting insolvency. To this end, more advanced data analysis techniques are used, namely text mining and decision trees with the CART algorithm to analyse the audit report between the years 2016 and 2020 of a sample of 2.040 companies, 1.020 non-insolvent and 1.020 insolvent. The results obtained indicate the existence of a relationship between the audit opinion content 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, Prediction.*

### I. INTRODUÇÃO

Portugal desde a crise financeira vivida entre 2008 e 2014, aumentou o número de processos de insolvência face ao período anterior à crise [1]. Apesar de este valor tender a diminuir, o número de empresas que não consegue liquidar as suas obrigações já vencidas é elevado e preocupante. 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.

Assim, investigadores têm apostado em estudos que permitam prever as insolvências das empresas. Os primeiros avanços na criação de modelos preditivos foram concebidos tendo como preditores os indicadores financeiros [2]. Contudo, a gestão dos resultados e a manipulação dos mesmos colocam em causa a fiabilidade e credibilidade das Demonstrações Financeiras (DF), pelo que os indicadores financeiros podem não representar a realidade económico-financeira das empresas. [3]. Por este motivo, e também pela necessidade de se comparar

estes indicadores com a informação não financeira é que a análise isolada dos rácios não garante a complexidade exigida e necessitaria para prever as insolvências [4].

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 prever as insolvências [5] [6]. Perante a dificuldade em prever problemas de inviabilidade das empresas, a informação proveniente da Auditoria Financeira Externa (AFE), começou a ocupar um papel crucial nestes modelos tendo contribuído para a criação de modelos com boa capacidade preditiva [7][8].

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 [9]. A Certificação Legal de Contas (CLC) / Relatório de Auditoria Financeira Externa (RAFE) é o resultado final do trabalho em auditoria e onde constam, para além de outras informações, os desacordos e limitações de âmbito material ou potencialmente relevantes detetados no decorrer da auditoria.

O auditor não é responsável pela viabilidade a longo prazo das empresas, apesar de o ser na avaliação anual da continuidade das empresas [10]. Não obstante, e embora a informação da CLC/RAFE seja pouco utilizada como preditor das insolvências, esta tem um enorme potencial na medida em que apresenta características que ajudam a diferenciar as empresas insolventes das não insolventes [7]. Deste modo, no estudo [7] introduziu-se, pela primeira vez, nos modelos preditivos, a análise do conteúdo da opinião do auditor, sejam comentários de reservas, ênfases ou parágrafos relativos ao pressuposto da continuidade, de forma a perceber se o tema abordado nesses comentários permite diferenciar as empresas insolventes das não insolventes. De facto, a utilização de determinados termos pode ser um claro sinal de alerta, permitindo ajudar na prevenção das insolvências [11]. Uma das vantagens apontadas na literatura [11] é a metodologia utilizada, uma vez que a identificação dos temas mencionados no conteúdo da opinião do auditor é feita de forma manual por peritos em contabilidade, o que torna o processo dispendioso, demorado e sujeito a diferentes interpretações.

Assim, o foco da presente investigação é a predição das insolvências das empresas, tendo como preditores as características do auditor e o conteúdo da sua opinião. Para isso, e em termos metodológicos, são utilizadas técnicas de análise de dados multivariada, nomeadamente árvores de decisão, e também técnicas alternativas à análise manual de conteúdo da opinião do auditor, em particular, o *text mining*.

No que se refere aos estudos prévios 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 o conteúdo da opinião do auditor de uma grande amostra de empresas, cuja análise manual seria inviável.

Deste modo, este estudo visa gerar conhecimento que permita apoiar profissionais e académicos a antecipar com maior precisão problemas de viabilidade das empresas, bem como contribuir para melhorar a reputação das firmas de auditoria que ficou fragilizada desde a crise financeira de 2008. Neste sentido, a presente investigação responde à seguinte questão: Como é que as características do auditor e o conteúdo da sua opinião permitem prever as insolvências das empresas?

## II. A AUDITORIA NA PREDIÇÃO DAS INSOLVÊNCIAS

Face ao trabalho desenvolvido pelos auditores é razoável prever que a CLC/RAFE tenha informações relevantes para prever e alertar para potenciais insolvências das empresas [12]. Entre as variáveis mais importantes relacionadas com o trabalho efetuado pela AFE, e utilizadas nos poucos modelos preditivos das insolvências abordados na literatura, destacam-se o tipo de opinião, o número de reservas e ênfases mencionadas nos relatórios e a mudança e a dimensão do auditor [7][8][13][14].

O tipo de opinião, uma das variáveis mais utilizadas nos estudos analisados, aumenta a capacidade de prever a insolvência das empresas e é uma variável diferenciadora entre as empresas insolventes e não insolventes, mostrando que a qualificação da opinião é um claro sinal de alerta das empresas que enfrentam dificuldades financeiras [13][15]. Contudo, relatórios com opiniões não qualificadas, mas com parágrafos de ênfases, também estão relacionados com as empresas que declaram insolvência [7], o que vai ao encontro de estudos de outros investigadores que concluem 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 [12].

A análise do conteúdo da opinião do auditor é uma inovação no âmbito dos modelos preditivos das insolvências, tendo sido pela primeira vez incluída nestes modelos em 2019 [7]. De facto, o estudo revelou que o conteúdo é um bom preditor das insolvências e que os auditores tendem a mencionar nas CLC/RAFE das empresas em insolvência temas relacionados com gastos do período, investimentos e incertezas quanto à continuidade das operações, o que não acontece nas empresas não insolventes.

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 [16]. Cerca de 80% dos dados globais estão num formato não estruturado, pelo que estas técnicas podem ser ferramentas bastante valiosas permitindo um melhor conhecimento na tomada de decisão [17].

O *text mining* 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. 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 no conteúdo

da opinião do auditor. 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 [18].

No que diz respeito à informação das características da firma de auditoria, existe uma tendência das empresas que entram em insolvência serem auditadas por firmas de auditoria mais pequenas, enquanto as não insolventes por firmas de maior dimensão, as designadas por *Big 4* [8].

Outra das características abordadas na literatura é a mudança de auditoria, 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 [4]. Em concordância, outro estudo também demonstra que as empresas insolventes tendem a mudar mais vezes de auditor, o que não acontece nas empresas não insolventes [15]. 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, não há evidência estatística que permita afirmar que a mudança do auditor está relacionada com a insolvência das empresas [12].

### III. METODOLOGIA

Este estudo consubstancia um estudo exploratório, uma vez que ao abordar o tema da predição das insolvências tendo como preditores as características do auditor e o conteúdo da sua opinião, contribuiu para obter conhecimento deste tema pouco abordado na literatura. Deste modo, a análise inclui dois níveis: i) identificar os tópicos salientes presentes no conteúdo da opinião na CLC/RAFE, e ii) criar um modelo preditivo das insolvências.

#### A. Dados 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 amostra é obtida através de um processo com várias fases, sendo a amostra constituída por PME portuguesas sujeitas a revisão oficial de contas e que não apresentam dados omissos na variável do tipo de opinião em ambos os anos de análise (ano N e N-1). Foram, também, excluídas da amostra todas as empresas com forma jurídica distintas de sociedades anónimas (S.A.) e sociedade por quotas de responsabilidade limitada (LDA), restringindo-se a amostra a 2,040 empresas, 1,020 insolventes e 1,020 não insolventes. Importa referir que a seleção final das empresas não insolventes tem em conta o número e perfil das empresas insolventes selecionadas, de modo que a amostra final seja composta pelo mesmo número de empresas não insolventes e insolventes cujos perfis (dimensão e setor de atividade) sejam idênticos.

A amostra final, constituída por PME, contempla empresas essencialmente de Lisboa e Vale do Tejo (43%) e Norte (36%), sendo o setor da indústria aquele com maior representatividade (56,6%), seguindo-se o setor dos serviços (21,9%).

#### B. Variável dependente

A variável dependente é uma variável dicotómica, de forma a distinguir o *status* das empresas: não insolvente ou insolvente. Para efeitos do presente estudo são consideradas insolventes as

empresas que judicialmente foram decretadas como tal, sendo que se encontram impossibilitadas de cumprir com as suas obrigações já vencidas. Este tipo de variável é comum na maioria dos estudos em que o objetivo é a elaboração de modelos preditivos das insolvências [7][8].

#### C. Variáveis independentes

As variáveis independentes do modelo preditivo das insolvências dividem-se em três dimensões (D1 e D2 e D3):

- **D1:** Dimensão do conteúdo da opinião do auditor, onde se considera a variável do tipo de opinião (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), a variável da mudança do tipo de opinião (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 nesta primeira dimensão, considera-se a variável comentário (diferencia as empresas que têm comentário na CLC/RAFE, das empresas que não têm qualquer comentário). Importa mencionar que neste estudo não há distinção do número nem da natureza do comentário emitido, pelo que será interessante a integração destas variáveis em investigações futuras. Considera-se, também, a variável da dimensão do comentário e a variável tópicos 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.

- **D2:** Dimensão das características do auditor, onde se considera a variável dimensão da auditoria 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 da mudança de auditoria, que diferencia as empresas que não mudam de auditoria entre N-1 e N, das empresas que mudam para uma não *Big 4*, das que mudam para uma *Big 4*, das que mudam mas continuam a ser auditadas por uma não *Big 4* e das que mudam mas continuam a ser auditadas por *Big 4*.
- **D3:** Características demográficas da empresa, sendo selecionadas a dimensão, sector (comércio, indústria, serviços e agricultura) e a região NUTS II (Alentejo, Algarve, Centro, Lisboa e Vale do Tejo, Norte e regiões autónomas da Madeira e Açores) e antiguidade.

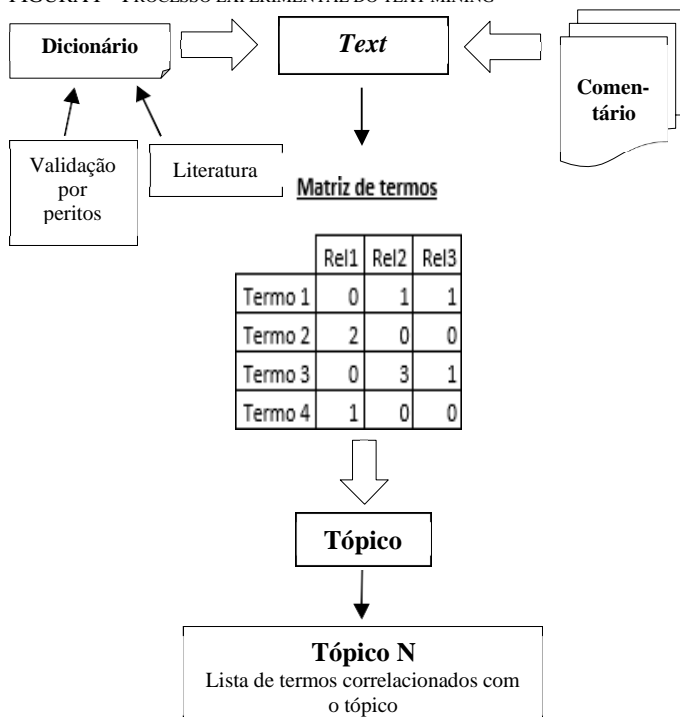
#### D. Técnicas de análise de dados

As análises foram realizadas no *SPSS Statistics* e também no *software Orange*. Em primeiro lugar, procedeu-se à análise descritiva uni e bivariada dos dados recolhidos para testar a sua qualidade e avaliar a relação entre as variáveis independentes e

o *status* da empresa [19]. Importa referir que 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 os anos de análise o 2019 que corresponde ao ano N e o 2018 ao N-1.

Recorreu-se a técnicas de *text mining* para a criação dos tópicos salientes presentes no conteúdo da opinião do auditor, 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 [7]. O objetivo da aplicação desta técnica de análise de conteúdo assistida por computador é identificar tópicos, de forma a perceber se existe uma tendência dos termos mencionados, que permita diferenciar as empresas insolventes das não insolventes. O procedimento da análise desses comentários está descrito na Figura I.

FIGURA I - PROCESSO EXPERIMENTAL DO TEXT MINING



São necessários dois *inputs*, um dicionário de termos relevantes que possam ser mencionados pelo auditor e os comentários. O dicionário é criado tendo em conta investigações anteriores [12] e o plano de contas onde constam os códigos e classificações utilizados na contabilidade das empresas, sendo revisto por peritos em contabilidade e auditoria. Após todo o 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. Posteriormente, a aplicação da técnica de *topic modeling* permite gerar diferentes combinações de termos associados a tópicos, aos quais estão associadas 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.

Por fim, recorre-se técnicas de estatística multivariada, nomeadamente as árvores de decisão com o algoritmo CART para a criação dos modelos preditivos do *status* da empresa. Os modelos concebidos através destes algoritmos não requerem a verificação de pressupostos e lidam bem com todos os tipos de variáveis, *outliers* e não respostas [3], o que se adequa ao estudo em causa. Adicionalmente, são muito facilmente interpretáveis ao gerarem um conjunto de regras de classificação.

Para analisar a validação dos modelos preditivos optou-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. De forma a melhorar a capacidade preditiva dos modelos são definidos parâmetro que permitem contrariar o sobre e sub ajustamento 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.

Optou-se por se criarem modelos para as diferentes dimensões de variáveis. Primeiramente, contendo apenas as variáveis do conteúdo da opinião do auditor sendo, posteriormente, acrescentado a esses modelos as variáveis das características do auditor. Por fim, foram também inseridas nos modelos as variáveis da terceira dimensão (demográficas). A seleção do melhor modelo teve em consideração o modelo que apresentava maior PECC, garantindo que a sensibilidade, isto é, a percentagem das empresas insolventes que o modelo acertou e estão de facto em insolvência e a especificidade que representa a percentagem de casos corretamente classificados de empresas não insolventes, apresentam valores igualmente elevados.

#### IV. RESULTADOS E DISCUSSÃO

##### A. Identificar os tópicos salientes presentes nos comentários emitidos na CLC/RAFE

Numa primeira fase, o estudo analisa as duas subamostras (não insolventes e insolventes) separadamente, de forma a identificar a frequência absoluta dos termos nos comentários dos auditores. Verifica-se que, de facto, existem determinados termos que permitem distinguir os relatórios das empresas insolventes das não insolventes. Pode verificar-se que o termo continuidade é o mais mencionado pelo auditor nas empresas insolventes (117 vezes), enquanto nas empresas não insolventes só é mencionado 32 vezes. Para além deste, termos como acionistas, artigo 35º do Código das Sociedades Comerciais (CSC), financiamento, insolvência e Plano Especial de Revitalização (PER) também tendem a ser mencionados com maior frequência pelos auditores nas CLC/RAFE das empresas insolventes face às não insolventes (Tabela I).

Importa mencionar que são analisados apenas 162 comentários em N (81 de empresas não insolventes e 81 de empresas insolventes) e 168 em N-1 (84 de empresas não insolventes e 84 de empresas insolventes). Opta-se por analisar o mesmo número de comentários das empresas não insolventes e insolventes para que a análise não seja distorcida.

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 insolventes e 12 vezes nas empresas solventes. De uma análise aleatória aos comentários que contêm este termo verificou-se que o auditor tende a emitir uma ênfase relativa à continuidade das operações.

TABELA I. FREQUÊNCIA ABSOLUTA DE CERTOS TERMOS

Termos	Insolventes	Solventes
Continuidade	117	32
Acionistas	47	16
Artigo 35º	44	6
Financiamento	36	13
PER	34	2
Insolvência	32	2

Da análise do *topic modelling* (Tabela II), 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 II. TERMOS CORRELACIONADOS COM OS TÓPICOS

Tópico 1	$\beta$	Tópico 2	$\beta$
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	$\beta$	Tópico 4	$\beta$
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  $\beta$

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. 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 CSC. Através da análise descritiva bivariada, pode verificar-se que, nas empresas não insolventes, os tópicos

seguem uma distribuição mais ou menos equitativa. Já no que diz respeito às empresas insolventes, o tópico 1 capital e continuidade destaca-se claramente dos restantes, estando presente em, aproximadamente, 50% das empresas (Tabela III).

TABELA III. DISTRIBUIÇÃO DOS TÓPICOS PELAS EMPRESAS

		Status			
		Não insolventes		Insolvência	
		Empresas	%	Empresas	%
Tópicos_N	Tópico 1	24	29,6	39	48,1
	Tópico 2	17	21,0	14	17,3
	Tópico 3	25	30,9	22	27,2
	Tópico 4	15	18,5	6	7,4
	Total	81	100	81	100
Tópicos_N1	Tópico 1	17	20,2	42	50
	Tópico 2	26	31	21	25
	Tópico 3	30	35,7	20	23,8
	Tópico 4	11	13,1	1	1,2
	Total	84	100	84	100

#### B. Resultados dos modelos preditivo das insolvências

Primeiramente importa analisar que todos os modelos apresentam bons resultados (Tabela IV). A tabela evidencia os três modelos com melhores resultados, sendo que o modelo (D1 e D2) 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 IV. RESULTADOS DOS MELHORES MODELOS PREDITIVOS

Resultados modelos	Modelos		
	D1	D1 e D2	D1, D2 e D3
PECC (%)	89,1	92,6	90,8
Dimensão das variáveis	AFE (Ano N)	AFE Auditora (Ano N)	AFE Auditora Demográficas (Ano N e N-1)
Variáveis finais Top 3	Tipo opinião_N	Tipo opinião_N	Tipo_opinião_N
	Comentário_N	Comentário_N	Comentário_N1
	Dim. Comentário_N	Dim. Auditora_N	Tipo_opinião_N1

No que diz respeito ao Top 3 das variáveis com maior importância para cada modelo, evidencia-se que as variáveis que surgem com maior frequência são as variáveis do conteúdo da opinião do auditor, nomeadamente, Tipo\_opinião e Comentário. Importa referir que, quando as variáveis das características do auditor são incluídas nos modelos, a variável da dimensão da auditora aparece no Top 3 das variáveis mais importantes, substituindo a variável da Dimensão do comentário, como se pode observar no modelo (D1 e D2). 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 [8]. De uma forma geral, tanto as variáveis do conteúdo da opinião do auditor como as variáveis das características do auditor são bons preditores das insolvências.

A Tabela V permite perceber com maior detalhe como é que as variáveis diferenciam o *status* previsto do melhor modelo (D1 e D2). O modelo prevê que para as empresas com opinião não qualificada, 98% sejam não insolventes e 2% insolventes. Por outro lado, evidencia-se que para as empresas com opinião qualificada na CLC/RAFE, 10% sejam não insolventes e 90% insolvente. As opiniões qualificadas são claros sinais de alerta de dificuldades financeiras [13]. No que concerne à variável da dimensão da auditora, o modelo prevê que as empresas auditadas por uma firma *Big 4* 100% sejam não insolventes. No que diz respeito à variável comentário, o modelo prevê que para as empresas que não têm comentário 100% sejam não insolventes e 0% insolventes. Já para as empresas que têm comentário, o modelo prevê que 15% sejam não insolventes e 85% insolventes. Estas conclusões vão ao encontro de outro estudo que evidencia que as empresas insolventes tendem a ter comentários na sua CLC/RAFE [12].

TABELA V. DISTRIBUIÇÃO DO STATUS PREVISTO PARA AS CATEGORIAS DAS VARIÁVEIS MAIS IMPORTANTES.

Variáveis	Descrição	Status previsto	
		Insolventes (%)	Insolventes (%)
Tipo opinião	ONQ	98	2
	OQ	10	90
Dim. Auditora	BIG 4	100	0
	N BIG 4	72	28
Comentário	Há	15	85
	Não há	100	0

Nota: ONQ: Opinião não qualificada; OQ: Opinião qualificada

## V. CONCLUSÕES

Tendo em conta esta problemática, o presente estudo identifica uma relação entre as características do auditor e o conteúdo da sua opinião e a insolvência das empresas, concluindo que a CLC/RAFE permite distinguir as empresas insolventes das não insolventes.

O presente estudo permite aumentar o grau de conhecimento no âmbito do tema dos modelos preditivos das insolvências através da AFE como preditor, bem como colmatar a lacuna no conhecimento na disciplina de contabilidade e auditoria referente ao desenvolvimento de modelos preditivos das insolvências que apliquem técnicas de *text mining* para a criação destes modelos. Por outro lado, os *insights* gerados permitem que o trabalho em auditoria ganhe uma maior visibilidade e relevância por parte dos utilizadores da informação, melhorando a sua reputação. A principal limitação é o número excessivo de não respostas, nomeadamente, dos comentários emitidos pelos auditores.

Em termos de pistas para investigações futuras e de forma a colmatar as limitações mencionadas acima, sugere-se a realização do mesmo estudo, mas com um maior número de

comentários. Para além disso, e como já referido anteriormente, a adoção de outras variáveis, como a natureza e o número de comentários emitidos na CLC/RAFE de forma a tentar aumentar a capacidade preditiva dos modelos.

## AGRADECIMENTOS

À FCT, pelo apoio no âmbito dos projetos estratégicos: UID/GES/00315/2020, UIDB/04466/2020 e UIDP/04466/2020. Ao Instituto Politécnico de Setúbal [programa RAADRI].

## REFERÊNCIAS BIBLIOGRÁFICAS

- [1] "PORDATA - Estatísticas, gráficos e indicadores de Municípios, Portugal e Europa," 2018. <http://www.pordata.pt/>
- [2] W.H. Beaver, "Financial ratios as predictors of failure," *Journal of Accounting Research*, vol. 4, pp. 71–111, 1966.
- [3] J. M. M. Constantino, "Gestão de resultados como preditor de insolvências: evidência nas empresas portuguesas," 2020.
- [4] J. H. Cheng, C. H. Yen, and Y. W. Chiu, "Improving business failure predication using rough sets with non-financial variables," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 4431 LNCS, no. PART 1, pp. 614–621, 2007.
- [5] J. Bauweraerts, "Predicting Bankruptcy in Private Firms: Towards a Stepwise Regression Procedure," *International Journal of Financial Research*, vol. 7, no. 2, 2016.
- [6] P. du Jardin, "Bankruptcy prediction using terminal failure processes," *European Journal of Operational Research*, 2015.
- [7] N. Muñoz-Izquierdo, M. D. M. Camacho-Miñano, M. J. Segovia-Vargas, and D. Pascual-Ezama, "Is the external audit report useful for bankruptcy prediction? Evidence using artificial intelligence," *International Journal of Financial Studies*, vol. 7, no. 2, 2019.
- [8] N. Muñoz-Izquierdo, E. K. Laitinen, M. del M. Camacho-Miñano, and D. Pascual-Ezama, "Does audit report information improve financial distress prediction over Altman's traditional Z-Score model?," *Journal of International Financial Management and Accounting*, vol. 31, no. 1, pp. 65–97, 2020.
- [9] C. D. Liggio, "Expectation gap - accountants legal Waterloo," *Journal of contemporary business*, vol. 3, no. 3, pp. 27–44, 1974.
- [10] International Auditing and Assurance Standards Board, "Manual das normas internacionais de controlo de qualidade, auditoria, revisão, outros trabalhos de garantia de fiabilidade e serviços relacionados", 2015.
- [11] J. C. das Neves, *Análise e Relato Financeiro*. 2012.
- [12] N. Muñoz-Izquierdo, M.-M. Camacho-Miñano, and D. Pascual-Ezama, "The content of the audit report in the year prior to bankruptcy filing. Empirical evidence from Spain," *Spanish Journal of Finance and Accounting / Revista Española de Financiación y Contabilidad*, vol. 46, no. 1, pp. 92–126, 2017.
- [13] K. van Peursem and Y. C. Chan, "Forecasting New Zealand corporate failures 2001-10: Opportunity lost?," *Australian Accounting Review*, vol. 24, no. 3, pp. 276–288, 2014.
- [14] E. I. Altman, G. Sabato, and N. Wilson, "The value of non-financial information in SME risk management," *The Journal of Credit Risk*, vol. 6, no. 2, pp. 95–127, 2010.
- [15] C. P. Sánchez, P. de L. Monelos, and M. R. López, "A parsimonious model to forecast financial distress, based on audit evidence," *Contaduría y Administración*, vol. 58, no. 4, pp. 151–173, 2013.
- [16] M. R. C. Santos, R. M. S. Laureano, and S. Moro, "Unveiling Research Trends for Organizational Reputation in the Nonprofit Sector," *Voluntas*, vol. 31, no. 1, pp. 56–70, 2020.
- [17] International Business Machines Corporation, "Text mining," 2020.
- [18] A. Zorio-Grima and P. Carmona, "Narratives of the Big-4 transparency reports: country effects or firm strategy?," *Managerial Auditing Journal*, vol. 34, no. 8, pp. 951–985, 2019.
- [19] R. Laureano, *Testes de Hipóteses e Regressão: O Meu Manual de Consulta Rápida*. 2020.