

NOTAS EXPLICATIVAS EXPLICAM? PROPOSTA DE UMA MEDIDA DE RISCO A PARTIR DE TÉCNICAS DE TEXT MINING

CARLA BONATO MARCOLIN

UNIVERSIDADE FEDERAL DE UBERLÂNDIA (UFU)

FERNANDA DA SILVA MOMO

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL (UFRGS)

ENRICO DALLA RIVA

ESCOLA DE ADMINISTRAÇÃO DE EMPRESAS DE SÃO PAULO (FGV-EAESP)

Agradecimento à órgão de fomento:

Os autores agradecem ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo apoio à este trabalho.

NOTAS EXPLICATIVAS EXPLICAM? PROPOSTA DE UMA MEDIDA DE RISCO A PARTIR DE TÉCNICAS DE TEXT MINING

1 INTRODUÇÃO

Os avanços computacionais permitem que ferramentas de inteligência artificial estejam cada vez mais acessíveis nos ambientes organizacionais. A maior capacidade de processamento e de memória expande as possibilidades de análise, trazendo outros formatos de dados antes indisponíveis. Ao mesmo tempo, o cenário de Big Data traz em uma de suas dimensões o V de Variedade, que compreende os dados não estruturados, objeto de análise deste artigo (McAfee et al., 2012).

Dados não estruturados são todos aqueles cujo formato não permite a organização tradicional em tabelas ou bancos de dados relacionais, e para os quais os processos de ETL (do inglês *Extract, Transform e Load*) não podem ser aplicados. São exemplos os dados de redes sociais como vídeos e fotos, dados em áudio e dados em texto. Este último formato, em texto, vem sendo o foco de diversas pesquisas na área contábil-financeira pela sua flexibilidade e riqueza de informação, podendo gerar importantes *insights* para processos decisórios de diferentes naturezas (Yang, Adomavicius, Burtch & Ren, 2018; Fisher, Garnsey & Hughes, 2016; Sawanobori, 2013).

A área de mineração de texto (*Text Mining*) tem contribuído com pesquisas no campo das ciências sociais aplicadas com o desenvolvimento de técnicas que permitem extrair, sumarizar e classificar textos, permitindo sua manipulação a partir de técnicas quantitativas. Neste artigo, **o objetivo é propor uma medida de risco para instituições financeiras a partir do conteúdo das notas explicativas, baseada em técnicas de mineração de texto.**

Assim, pretende-se analisar a possibilidade de evidenciar o impacto e a efetividade de gerenciamento de risco a partir de uma medida de risco que tenha como fonte constitutiva o conteúdo textual das notas explicativas. Isso porque as notas explicativas são o espaço de detalhamento informacional das demonstrações financeiras em que se apresenta, entre outros: a base para a elaboração das demonstrações e as políticas específicas utilizadas; informações complementares exigidas por normativas; informações adicionais que sejam relevante para a compreensão da informação financeira da organização (CPC, 2020; IFRS, 2020).

Tendo em vista a relevância dos conteúdos apresentados nas notas explicativas e a quantidade de fragmentos textuais existente nelas, há uma crescente busca para aplicação de técnicas de *data analytics* que possibilitam um melhor uso dessas informações para tomada de decisão, aprimoramento de auditorias e forma de apresentação das informações para os usuários da informação. Dessa forma, assim como há estudos que usam técnicas de *data analytics* para aprimoramento de estimativas e escrita (Ding, Lev, Peng, Sun, & Vasarhelyi, 2019), identificação de correlação entre a forma de divulgação de resultados, discurso dos gestores sobre as operações com valor futuro das ações, ganhos e possível ação fraudulenta dos gestores (Loughren & McDonald, 2016), esse estudo, ao propor uma medida de risco que considera as notas explicativas, visa identificar se esse índice pode suportar um melhor controle sobre a efetividade do gerenciamento de risco em instituições financeiras pelos diversos *stakeholders*. Logo, o uso de notas explicativas em medidas de risco pode fornecer uma nova visão sobre a área de risco para órgãos de controles e investidores, uma vez que não são analisadas dessa forma.

Após esta introdução, a próxima seção apresenta um referencial teórico sobre risco e notas explicativas, evidenciando o *gap* de pesquisa. A seção três aborda os aspectos e as escolhas

metodológicas, seguida da seção de análise dos resultados. Por fim, a seção cinco traz as conclusões, limitações e próximos passos deste estudo.

2 REFERENCIAL TEÓRICO

A seguir apresenta-se duas seções em que se aborda o tema do risco nas instituições financeiras (2.1) e das notas explicativas (2.2).

2.1 Risco

Instituições financeiras, particularmente, possuem no cerne de suas operações a presença de fatores de riscos e incertezas, inerentes aos negócios bancários. Logo, o processo de gestão de riscos é de fundamental importância para o entendimento dos riscos incorridos em suas rotinas operacionais, no sentido de garantir condições para que riscos sejam adequadamente enfrentados, controlados e gerenciados (Raghavan, 2003). De acordo com Crockford (1982), a efetividade de um gerenciamento de riscos está diretamente relacionada a um exitoso controle de perdas.

Algumas iniciativas internacionais surgiram no sentido de promover boas práticas corporativas de procedimentos internos e se tornaram referência mundial. O COSO (*Committee of Sponsoring Organizations of the Treadway Commission*), fundado em 1985 e que assessora a *Security Exchange Commission* (SEC) nos Estados Unidos, foi uma dessas organizações que produziu material pertinente ao tema (Benjamin Jr., 2005), e com relação a gestão de riscos, a definiu, em sua proposta de estrutura, como o processo aplicado à implementação da estratégia almejada em toda a empresa, desenhado para identificar eventos que pudessem afetar a entidade, e administrá-los dentro de seu apetite pré-estabelecido de risco, no sentido garantir o alcance dos objetivos da corporação (COSO, 2004).

A referência internacional em termos de sistemas financeiros é o Comitê de Basileia sobre Supervisão Bancária, órgão integrante do *Bank for International Settlements* (BIS), que por sua vez recomenda padrões de abordagens de gerenciamento de capital e riscos voltada aos reguladores bancários no mundo. No seu princípio fundamental de número 7, a entidade sugere que os supervisores dos sistemas bancários devem se assegurar que instituições supervisionadas possuam um amplo processo de gerenciamento de riscos que inclua: identificação, avaliação, monitoramento, controle e mitigação, e de adequação de capital ao seu perfil de risco (BCBS, 1997)

No Brasil, as instituições financeiras são regulamentadas pelo Banco Central do Brasil, que segue as diretrizes emanadas pelo Comitê de Basileia. Atualmente, as exigências previstas na regulamentação pertinentes a gestão de riscos estão dispostas na Resolução 4.557/17, a qual exige que as instituições financeiras mantenham uma adequada estrutura de gerenciamento integrado de riscos compatível com suas atividades. A regulação exige em seu artigo 56, a divulgação de sua estrutura gerenciamento de riscos e capital em relatório de acesso público, e resumo da descrição nas notas explicativas que acompanham as demonstrações financeiras semestrais (BCB, 2017).

Ryan (2015) identificou que as informações de risco divulgadas pelas instituições financeiras americanas estavam cada vez mais extensas e com integração opaca em relação às demonstrações financeiras, tornando-as de difícil identificação de relevância pelos usuários. Seu trabalho sugere que reguladores aprimorem a exigência de qualidade da informação, com padronização de divulgação de fatores de riscos, que permitam comparabilidade; condução de testes de estresse com evidenciação de resultados; estreitamento de laços com as informações contábeis, entre outros.

O aprofundamento das análises dos relatórios financeiros a partir de técnicas de *text mining* torna-se relevante devido ao volume de informações geradas nesses relatórios, sendo difícil a identificação de padrões, indícios de ações fraudulentas, informações sem qualidade quando analisados separadamente e/ou manualmente, havendo uma expansão de pesquisas que enfocam na análise das narrativas dos relatórios financeiros (Brown, Crowley, & Elliott, 2020; Huang et al., 2018; Dyer, Lang, & Stice-Lawrence, 2017). Na área de risco torna-se ainda mais relevante essa análise sobre se as informações divulgadas nos relatórios financeiros representam de forma efetiva as ações da organização no que tange ao enfrentamento, controle e gerenciamento dos riscos possibilitando um maior controle e aprimoramento das normas pelos órgãos reguladores e garantindo uma maior segurança a todos os *stakeholders* do mercado financeiro.

2.2 Notas Explicativas

As notas explicativas, segundo o Pronunciamento Técnico 26 (CPC26), “oferecem descrições narrativas ou decomposição (detalhamento) de itens apresentados [nas] demonstrações e informação acerca de itens que não se qualificam para serem reconhecidos nas demonstrações contábeis” (CPC, 2020), o que evidencia o caráter complementar mas ao mesmo tempo distintivo das notas explicativas. Da mesma maneira, o *International Financial Reporting Standards* destaca as notas explicativas como elementos relevantes para a existência de um relatório financeiro completo uma vez que estas apresentam as políticas adotadas, além de outras informações relevantes ao usuário da informação (IFRS, 2020).

Os dados em texto na área contábil são, portanto, significativos e importantes. Diferentes dos índices quantitativos, os dados em texto possuem uma subjetividade e variabilidade maior, não só em formato como também em conteúdo, o que faz com que possuam informações com um teor distinto. Essa especificidade acaba criando uma barreira de análise em uma área dominada por quantificações financeiras, o que muitas vezes torna esse tipo de dado invisível (Porter, 2005; Fisher et al., 2010).

O avanço em técnicas de mineração de texto tem auxiliado a mudar esse cenário. Fisher e outros (2010) identificaram que a área de contabilidade possui interface com a área de processamento de linguagem natural (NLP, do inglês *Natural Language Processing*) em duas direções. A primeira é mais conectada com a área de recuperação da informação (IR, do inglês *Information Retrieval*), e seu foco é compreender o léxico da área, bem como a formação de dicionários (Swandi, Suwardi & Maulidevi, 2017; Nielsen & Fuertes-Oliveira, 2013). Apesar de existirem diversos trabalhos nesta corrente, há críticas que apontam que a análise semântica no formato quantitativo e em volume não estaria presente, reduzindo a capacidade de integração destes dados com outras informações contábeis (Gomes, Ferreira & Martins, 2018).

Esta crítica nos conduz para a segunda direção, foco deste estudo, que conecta o dado em texto contábil com a área de mineração de texto. Ainda que ambas correntes contribuam para maior acessibilidade e compreensão destes textos, nesta direção existe o objetivo de utilizar este texto para realizar análises e inferências atuais, como valor de preço das ações e futuras e previsão de performance e fraude (Rahrovi Dastjerdi, Foroghi & Kiani, 2019; Loughren & McDonald, 2016). Desta forma, mais do que criar um léxico ou um conjunto linguístico específico da contabilidade, as técnicas de mineração de texto podem auxiliar a tornar este dado mais presente nas avaliações contábeis de diferentes naturezas para gestores e *stakeholders*.

Contudo, apesar de haver uma área de pesquisa madura no exterior (ver Fisher e outros (2010) para uma revisão de técnicas de mineração de texto semi-automatizadas e automatizadas),

o campo ainda é incipiente no Brasil, considerando especificamente o texto das notas explicativas. Os trabalhos encontrados ainda são mais focados em análises léxicas e volumétricas em relação a tamanho das notas e quantidade de palavras (Jales et al., 2020; Gomes, Ferreira & Martins, 2018), e não no seu conteúdo. Assim, observa-se que se encontra madura a área que trabalha com a análise de relatórios anuais divulgados pelas empresas e sua relação com prevenção e previsão de riscos contudo poucos trabalhos trazem abordagem considerando especificamente as notas explicativas. Para preencher estas lacunas de pesquisa, este artigo busca propor uma medida de risco que considere o conteúdo das notas explicativas para instituições financeiras. Para isso, realizou-se o estudo descrito na próxima seção.

3 MÉTODO

A presente pesquisa caracteriza-se como quantitativa e exploratória. O objetivo é propor uma medida de risco considerando o conteúdo das notas explicativas. Para tal, foram coletados trechos das notas explicativas que dissertam sobre a estrutura de gerenciamento de riscos. A amostra escolhida foram Sociedades de Crédito, Financiamento e Investimento (“Financeiras”) e Agências de Fomento (“Fomento”), classificadas no segmento 4 (S4), autorizadas a funcionar pelo Banco Central do Brasil (BC). As notas explicativas que acompanham os demonstrativos contábeis foram coletadas da Central de Demonstrações Financeiras do Sistema Financeiro (CDSFN), criada pela Circular 3.964/2019, disponível no site do BC (bcb.gov.br/estabilidadefinanceira/cdsfn), para as datas-bases junho e dezembro de 2019. Para as datas anteriores as informações foram coletadas nos sites de internet das instituições financeiras, diários oficiais dos Estados e via Google.

A escolha da amostra se deu por dois principais motivos. O primeiro foi a obrigatoriedade destes setores de cumprirem normativas de gerenciamento de risco apesar de sua menor área de abrangência considerando outras instituições financeiras como bancos comerciais. Dado a natureza exploratória desta pesquisa, optou-se por avaliar a efetividade da proposta a partir de empresas que teriam menos elementos de complexidade exigidos pela normativa. O segundo motivo foi o fato de serem dois grupos concisos. A amostra final contou com 18 Financeiras e 14 Agências de Fomento, que contempla quase a totalidade (86%) de instituições destes setores regulados pelo BC, o que permite ter uma abrangência mais completa para os testes realizados.

A captura dos dados a partir dos demonstrativos deu-se de forma manual, uma vez que os documentos não possuíam padrão para extração automatizada. Das 37 instituições, cinco não puderam ser incorporadas à amostra por indisponibilidade de captura de dados: os demonstrativos estavam ou protegidos ou em formato de imagem sem possibilidade de conversão. O período final da coleta compreendeu demonstrativos de Dezembro de 2014 até Dezembro de 2019, sendo dois demonstrativos anuais, um em Junho e um em Dezembro de cada ano.

O normativo tomado como base para medida de risco foi a Resolução 4.557, de 23 de fevereiro de 2017, que dispõe sobre a obrigatoriedade de implementação de estruturas de gerenciamento de riscos e de capital para os segmentos de 1 a 4 (S1 a S4), em conformidade com a Resolução 4.553/17, considerando a relevância e proporcionalidade das entidades supervisionadas. O regramento revogou as Resoluções 3.380/06, 3.464/07, 3.721/09, 4.090/12, e 3.988/11, que davam tratamento específico para gestão de riscos operacionais, de mercado, de crédito e de liquidez, e gestão de capital respectivamente, e as consolidou em um só regulamento, introduzindo o conceito de gestão integrada de riscos.

Como a proposta do presente artigo é inovadora e não possui comparativo, utilizou-se de medidas quantitativas de risco para avaliação e *benchmarking*. As medidas de risco coletadas para

comparação foram o índice de Basileia (relação entre o patrimônio de referência e os ativos ponderados pelo risco) e o índice Inad90 (porcentagem de operações da carteira de crédito da instituição, em atraso há mais de 90 dias, classificados em níveis de risco de D a H, *proxy* de ativos problemáticos), recuperados do site do BC (<https://www3.bcb.gov.br/ifdata/>).

Das 32 instituições integrantes da amostra final, algumas não possuíam algum dos índices em um ou mais dos 11 períodos, ou então não havia disponibilidade do texto sobre risco na nota explicativa, seja por indisponibilidade do demonstrativo, seja por não haver citação sobre essa estrutura. Desta forma, a amostra final contou com 241 registros, 133 de Financeiras e 109 de Agências de Fomento.

Para medida de risco, foi avaliada a distância entre os trechos de gerenciamento de risco das instituições e o trecho sobre estrutura de gerenciamento de risco da referida norma (Capítulo I, Seção III). As medidas de distância em textos buscam evidenciar a proximidade semântica entre diferentes documentos, e são utilizadas para tarefas como identificação de plágio e análise de similaridade de consumidores (Hourrane & Benlahmar, 2017; Saumya et al, 2018). Neste estudo, as medidas de distância buscam evidenciar o quão aderente as instituições são à normativa, considerando que quanto mais próximas, melhor atendem, e quanto mais distantes, menos o seu relato contempla aspectos da norma.

A medida de distância escolhida foi a distância, ou similaridade, de cosseno. Esta medida é largamente adotada na comunidade de mineração de textos, e está baseada em calcular o cosseno dos ângulos entre dois textos, projetados no mesmo espaço vetorial (Aggarwal & Zhai, 2012; Li & Han, 2013). Seu intervalo é de [0,1] e indica que a similaridade cresce quanto mais próximo de 1. A intuição é que textos cujo significado semântico seja similar, estarão projetados para a mesma direção, ainda que o tamanho dos textos, e por consequência, dos vetores, seja distinto. Essa particularidade da similaridade de cosseno torna sua aplicação adequada a este estudo, uma vez que existe variabilidade na quantidade de textos entre as diferentes instituições financeiras e a norma analisada, que não pode ser normalizada para não provocar perda de informação. A similaridade de cosseno entre cada par instituição-normativa (documentos U e V) foi calculada da seguinte maneira (Aggarwal & Zhai, 2012):

$$\text{cosseno}(U, V) = \frac{U \cdot V}{\|U\| \cdot \|V\|} = \frac{\sum_{i=1}^k U_i V_i}{\sqrt{\sum_{i=1}^k U_i^2} \sqrt{\sum_{i=1}^k V_i^2}}$$

Sendo U e V compostos pela frequência normalizada de seus i termos. Os vetores normalizados foram gerados através de uma matriz de termos e documentos, ponderadas pelo índice TF-IDF (do inglês, *term-frequency* e *inverse document-frequency*), que valoriza termos mais exclusivos e desvaloriza termos muito frequentes, que não são capazes de distinguir os documento por estarem presentes em todos eles. Esta ponderação permite uma análise do espaço semântico mais sofisticada e também objetiva melhorar o desempenho da similaridade de cosseno (Li & Han, 2013).

A preparação e pré-processamento dos dados contou com procedimentos indicados em literatura como remoção de *stopwords* (palavras de baixo significado semântico, como “de”, “para”, “o”, “por”, etc), a transformação para letras minúsculas e a eliminação de espaços em branco (Aggarwal & Zhai, 2012). O processamento e análise dos dados foi executado através do R, pelo software RStudio.

4 ANÁLISE DOS RESULTADOS

A amostra analisada contou com 241 textos retirados da seção de notas explicativas dos demonstrativos de 32 instituições autorizadas a funcionar pelo Banco Central (18 Financeiras e 14 Agências de Fomento). A natureza de atividade de cada uma dessas classes é distinta. Enquanto as Financeiras atuam mais voltadas ao crédito, as Agências de Fomento possuem uma proposta financeira mais social. Dado essa distinção, a análise foi dividida entre estes dois grupos, conforme as próximas seções.

4.1 Financeiras

As Sociedades de Crédito, Financiamento e Investimento (SCFI), também conhecidas e doravante denominadas Financeiras, são regulamentadas pela Portaria do Ministério da Fazenda 309/59 e pela Resolução do Banco Central 45/66, concedem empréstimos e financiamentos para a aquisição de bens, serviços e para capital de giro. Aquelas não ligadas a bancos, muitas vezes operam como braço financeiro de grupos comerciais ou industriais, como lojas de departamentos e montadoras de veículos. Geralmente atuam em nichos não atendidos por bancos, de risco mais elevado, como financiamento de veículos usados e convênios com estabelecimentos comerciais. E operam somente com a captação de recursos por meio de emissão de títulos - próprios ou letras de câmbio (Viera, Pereira & Pereira, 2012). Os 133 textos do grupo de Financeiras foram reunidos em um *corpus* (conjunto de documentos) com 2.272 termos únicos. Os termos mais frequentes identificam as características esperadas dos trechos das notas explicativas conforme Figura 1.

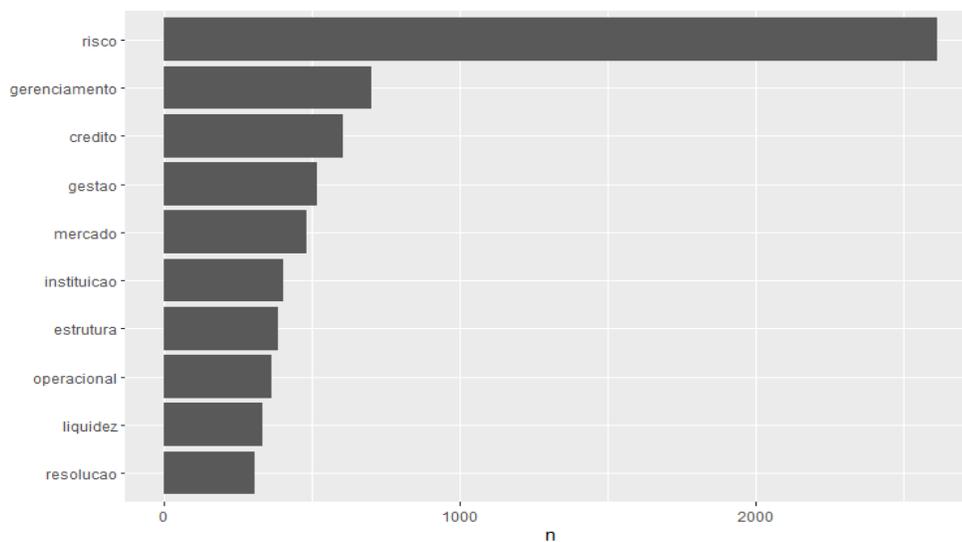


Figura 1: Top 10 palavras mais frequentes no *corpus* Financeiras

Fonte: Dados da Pesquisa

Apesar da Resolução 4.557/17 mencionar seis tipos de riscos diferentes (crédito, mercado, variação taxas de juros, operacional, liquidez e socioambiental), além de mencionar a importância de incluir demais riscos relevantes para cada instituição, apenas os riscos de crédito, de mercado, operacional e de liquidez estão entre as palavras mais frequentes. Para analisar essa questão com mais especificidade, foi realizada uma análise de n-gramas, com n variando de dois até cinco. Essa análise permite visualizar as n-palavras mais frequentes nas frases, complementando análises do

tipo *bag of words*, que não consideram a ordem das palavras. O Quadro 1 apresenta as bi e penta-gramas mais frequentes, considerando um limite inferior de 20 (mínimo de 20 ocorrências no *corpus*).

Através do Quadro 1 é possível verificar que os riscos de variação de taxa de juros e socioambiental não estão presentes nas notas explicativas. Adicionalmente, as penta-gramas apresentam recorrente menção à complexidade de produtos e inadequação de processos, o que pode indicar os principais demais riscos não mencionados na resolução, porém relevantes para esse conjunto de instituições.

bi-grama	Freq	penta-grama	Freq
Gerenciamento Risco	566	(...) compatíveis o tamanho complexidade produtos (...)	26
Risco Mercado	332	(...) deficiência ou inadequação processos internos (...)	26
Gestão Risco	330	(...) o tamanho complexidade produtos serviços (...)	26
Risco Crédito	329	(...) ou inadequação processos internos pessoas (...)	26
Risco Operacional	300	(...) tamanho complexidade produtos serviços oferecidos (...)	26
Estrutura Gerenciamento	215	(...) falha deficiência ou inadequação processos (...)	23
Risco Liquidez	177	(...) definido como possibilidade ocorrência perdas (...)	22
Resolução CNM	109	(...) identificar avaliar monitorar controlar mitigar (...)	22
Banco Central	98	(...) risco mercado liquidez operacional credito (...)	22
Controles Internos	93	(...) risco liquidez o risco liquidez (...)	21
Resolução n	86	(...) controles que permitem o acompanhamento (...)	20

Quadro 1: Conjunto de n-termos (n=2 e n=5) em ordem mais frequentes para Financeiras
Fonte: Dados da Pesquisa

O Gráfico 1 apresenta o Inad90 em conjunto com a similaridade de cosseno, sendo o eixo x cada demonstração das instituições financeiras (F1 até F18) pertencentes a este grupo. Pode-se observar que apesar dos valores estarem em grandezas diferentes, em alguns momentos (especialmente na primeira metade) há uma tendência conjunta, que acaba se dissipando, indicando que para algumas empresas existe uma correlação forte, e para outras não.

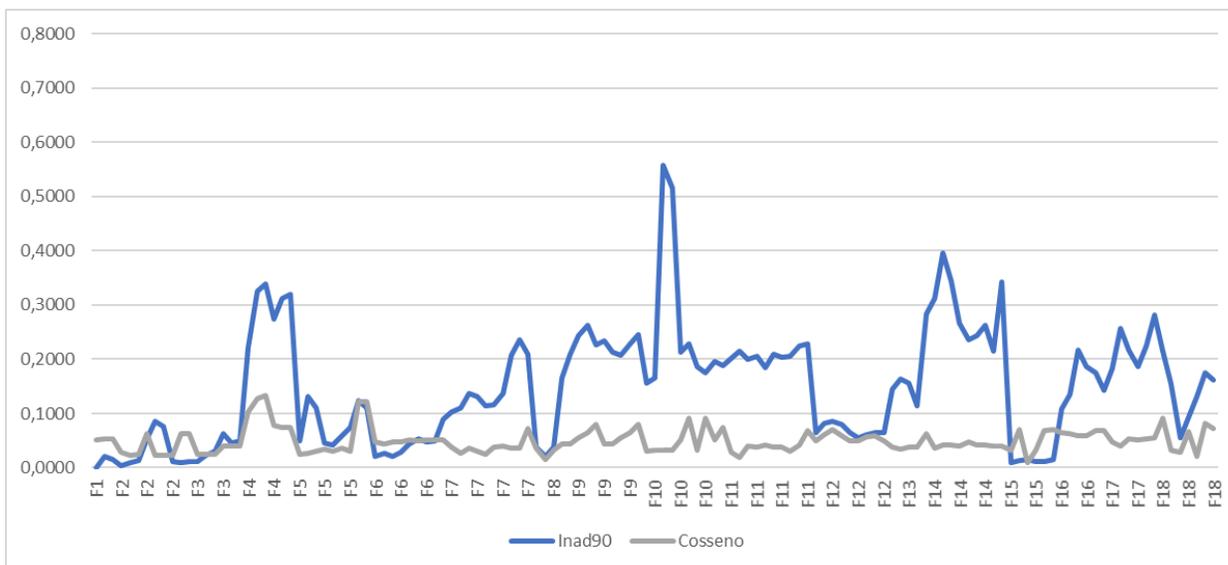


Gráfico 1: Inad90 e Similaridade de Cosseno para Financeiras

Fonte: Dados da Pesquisa

O mesmo pode ser observado em relação ao Gráfico 2, que apresenta a comparação entre o índice de Basiléia e a similaridade de cosseno. Apesar de serem medidas que demonstram diferentes direções de políticas de risco, o fato de algumas empresas possuírem ambas crescentes produziu uma tendência similar.

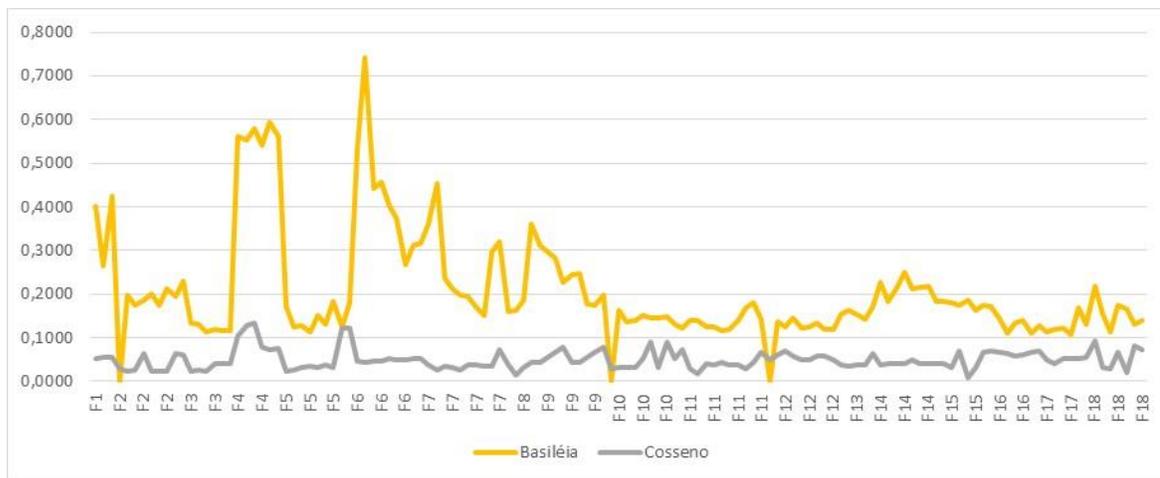


Gráfico 2: Índice de Basiléia e Similaridade de Cosseno para Financeiras

Fonte: Dados da Pesquisa

De forma a melhor compreender essa relação triangulada, o Gráfico 3 apresenta, por instituição, a correlação entre a similaridade de cosseno com o Inad90 (em azul) e com o índice de Basiléia (em laranja). A correlação de Pearson geral entre a similaridade de cosseno e o Inad90 foi de 0,2309, enquanto que com o índice de Basiléia foi de 0,3056.

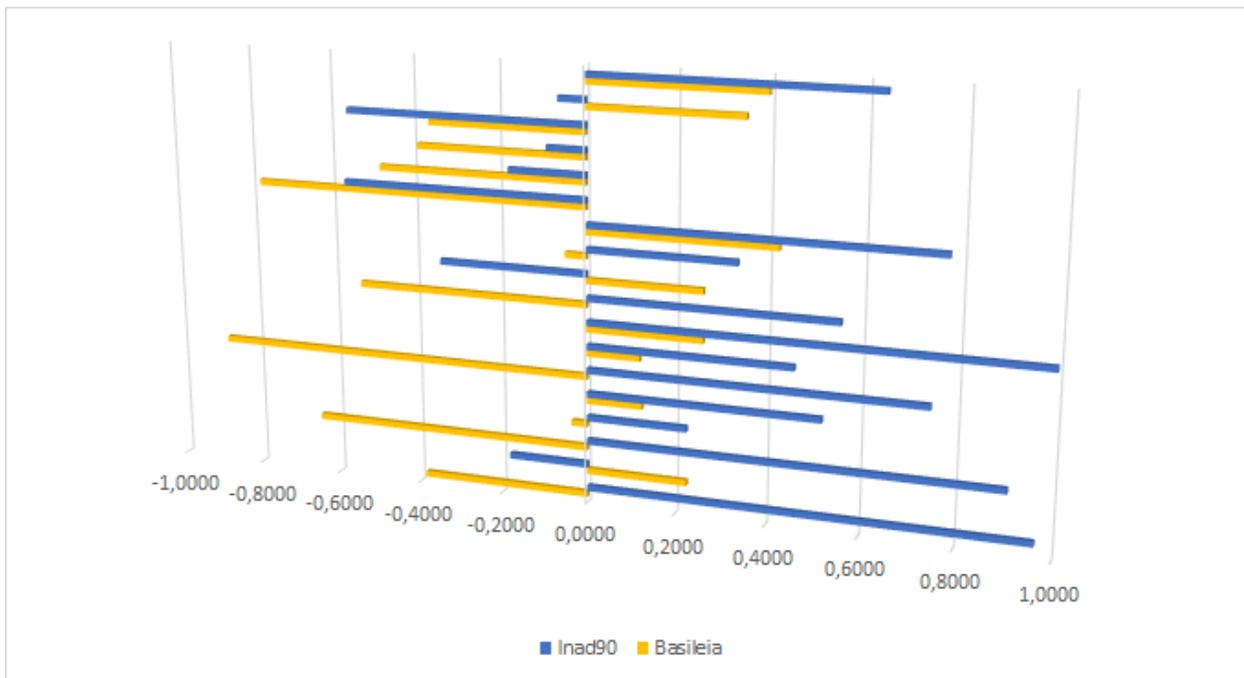


Gráfico 3: Correlação de Pearson, por instituição, com os índices de risco (Inad90 e Basileia)

Fonte: Dados da Pesquisa

Ainda que seja importante considerar as diferentes políticas de crédito, é possível verificar uma correlação fraca, em âmbito geral, entre cada um dos índices avaliados com a medida de cosseno. Percebe-se no Gráfico 3 que nove das dezoito instituições possuem uma correlação significativa (acima de 0,4) com o Inad90, e que destas nove, quatro também possuem correlação significativa (menor que -0,4), contudo contrária, com o índice de Basileia. Essa direção oposta é esperada dado que o índice de Basileia apresenta um objetivo contrário em relação ao Inad90: enquanto o primeiro demonstra menor apetite ao risco com crescimento, o segundo significa maior apetite ao risco conforme aumenta.

4.2 Agências de Fomento

As Agências de Fomento são instituições únicas por Estado, que atuam no financiamento de capital associado a projetos na Unidade da Federação, previstos em programas estaduais de desenvolvimento econômico e social, regulamentados pela Resolução 2.828/01 do Banco Central do Brasil (BCB, 2001). Segundo Ribeiro (2004), as agências se caracterizam por possuir fatores restritivos legais e econômicos, cujo sucesso está ligado às suas capacidades de promover novos projetos e investimentos. Os 109 textos deste grupo formaram um *corpus* com 2.366 termos únicos. A Figura 2 apresenta os termos mais frequentes.

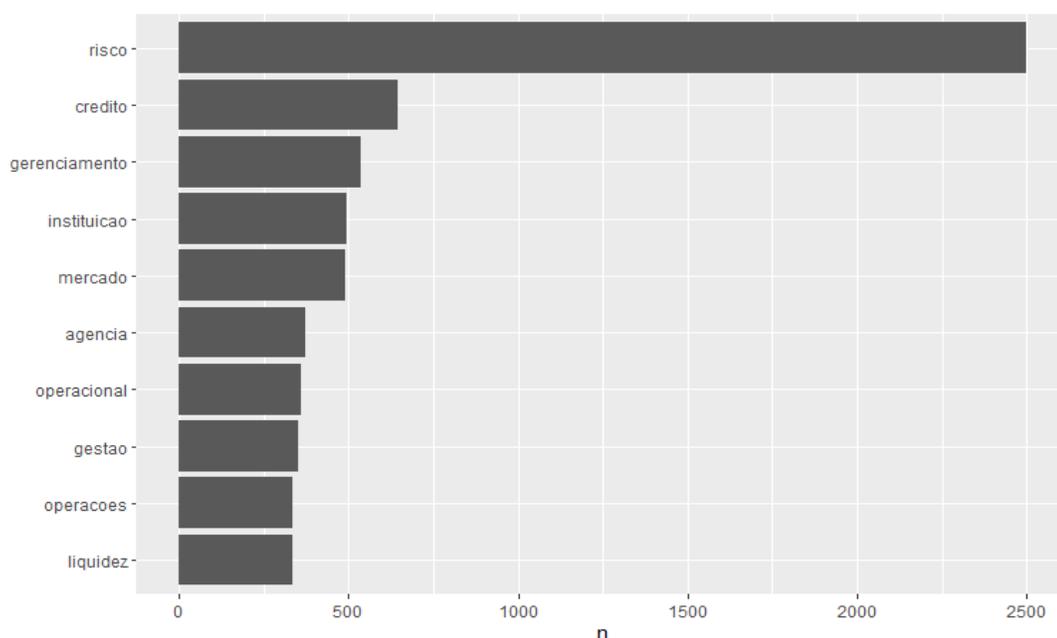


Figura 2: Top 10 palavras mais frequentes no *corpus* Agências de Fomento
 Fonte: Dados da Pesquisa

Pode-se a alta similaridade entre os dois grupos em relação aos termos mais frequentes: do top-10 termos, oito são iguais, ainda que a área de atuação dos dois grupos seja, em princípio, distinta. Chama atenção a palavra “agência”, termo esperado dado que a nominação das instituições que trabalham em fomento. Igualmente, também se nota a falta da menção ao risco socioambiental e ao risco de variação da taxa de juros, citados na Resolução 4.557/17. Os resultados da análise de n-gramas, que aprofunda a percepção sobre os termos mais frequentes, encontra-se no Quadro 2.

bi-grama	Freq	penta-grama	Freq
Gerenciamento Risco	429	(...) deficiência ou inadequação processos internos (...)	50
Risco Crédito	391	(...) falha deficiência ou inadequação processos(...)	50
Risco Operacional	315	(...) obrigações financeiras nos termos pactuados (...)	49
Risco Mercado	266	(...) ou inadequação processos internos pessoas (...)	49
Gestão Risco	208	(...) associadas ao cumprimento tomador ou (...)	47
Lavagem Dinheiro	190	(...) internos pessoas sistemas ou eventos (...)	47
Risco Liquidez	175	(...) perdas associadas ao cumprimento tomador (...)	47
Ocorrência Perdas	132	(...) pessoas sistemas ou eventos externos (...)	47
Possibilidade Ocorrência	121	(...) possibilidade ocorrência perdas resultantes falha (...)	47
Resolução n	113	(...) ocorrência perdas resultantes falha deficiência (...)	46
Banco Central	105	(...) prevenção ao crime lavagem dinheiro (...)	39

Quadro 2: Conjunto de n-termos (n=2 e n=5) em ordem mais frequentes para Agências de Fomento
Fonte: Dados da Pesquisa

A maior variabilidade de termos encontrada no Quadro 2 em relação ao Quadro 1 denota que este grupo possui uma discussão mais ampla presente nas notas explicativas, no que condiz à estrutura de gerenciamento de risco. Apesar de poder ser um reflexo das diferentes atuações das instituições deste grupo, percebe-se que a temática de lavagem de dinheiro está mais evidente. Ainda, há similaridade na questão de inadequação (ou deficiência termo neste grupo mais presente) de processos.

A relação entre o índice Inad90 e a similaridade de cosseno pode ser analisada pelo Gráfico 4, tendo no eixo x os valores obtidos para cada Agência de Fomento (AF1 até AF14) dos períodos cuja demonstração, com o texto de risco das notas explicativas, e o respectivo índice, estavam disponíveis. Percebe-se que a relação é ainda menor do que no grupo de Financeiras, o que é confirmado pelo menor índice de correlação de Pearson, que foi de -0,0187.



Gráfico 4: Inad90 e Similaridade de Cosseno para Agências de Fomento
Fonte: Dados da Pesquisa

O Gráfico 5 apresenta a relação entre a similaridade de cosseno e o índice de Basiléia. Esta análise teve quatro outliers removidos, de quatro agências diferentes, que foram reportados acima dos 150% nas demonstrações de dezembro de 2014, dado que a média de todos os períodos foi de 51,75%. Da mesma forma, pode-se perceber que o gráfico confirma a baixa relação entre as séries, com um índice de correlação de Pearson no valor de -0,2824, ainda que maior do que o índice Inad90.

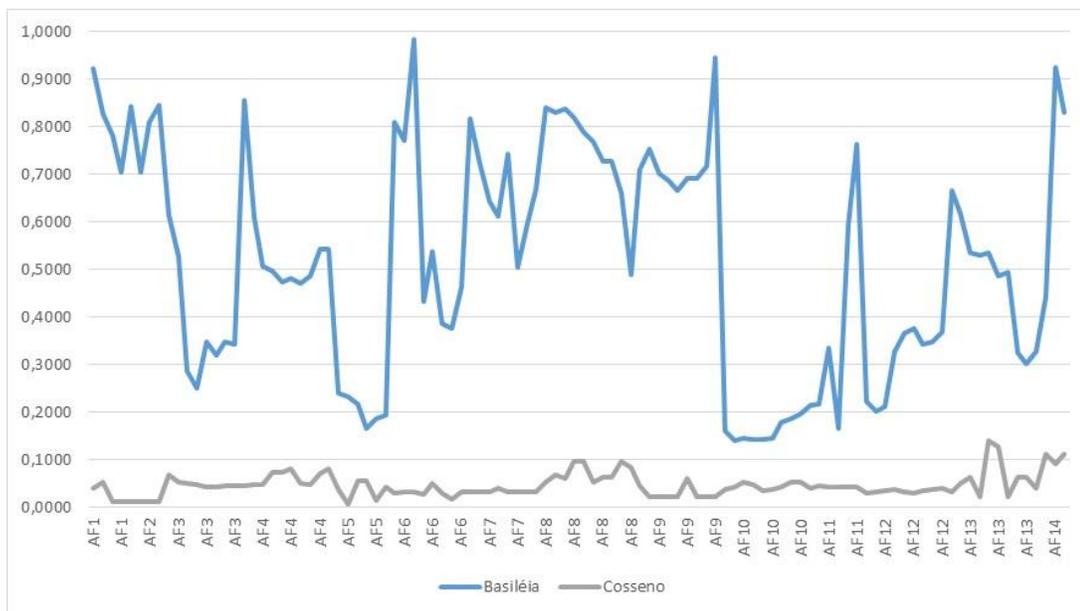


Gráfico 5: Índice de Basiléia e Similaridade de Cosseno para Agências de Fomento
 Fonte: Dados da Pesquisa

Considerando a correlação, para cada instituição, entre os índices e a similaridade de cosseno, não foi possível reproduzir para este grupo. Diferente do que ocorreu com as Financeiras, os valores da similaridade foram iguais (ou muito próximos) para diferentes períodos intra-instituição, e, portanto, a falta de variabilidade não permitiu realizar o cálculo do índice de correlação (uma vez que o desvio-padrão foi igual a zero). O Gráfico 6 apresenta os valores da similaridade de cosseno entre o texto das notas explicativas na demonstração de cada período de cada instituição deste grupo em relação ao trecho da resolução analisado.

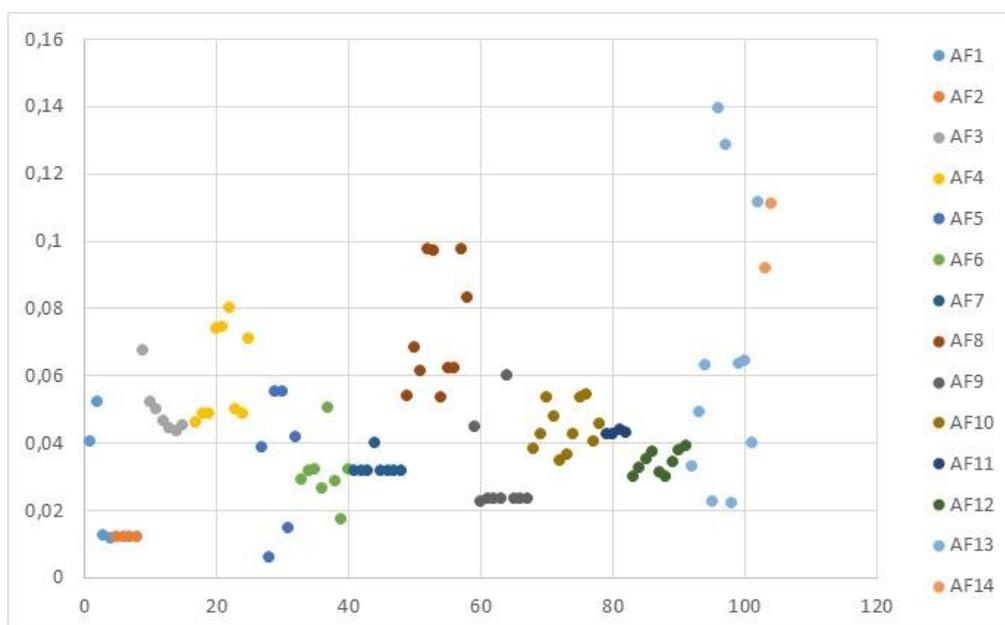


Gráfico 6: Gráfico de Dispersão dos valores de Similaridade de Cosseno nas Agências de Fomento (As cores representam as 14 Agências analisadas)
 Fonte: Dados da Pesquisa

A medida de similaridade de cosseno foi utilizada neste trabalho para mensurar a proximidade dos trechos sobre gerenciamento de riscos nas notas explicativas de dois grupos de instituições financeiras (Financeiras e Agências de Fomento) com a Resolução 4.557/17 do BC, que regulamenta a gestão integrada de risco. O objetivo era identificar se havia alguma relação desta medida de distância textual com as medidas de risco tradicionais. A intuição era que o texto poderia evidenciar a política de gerenciamento de riscos que impacta nos índices quantitativos relacionados.

Os achados permitem concluir que não há relação desta medida com as métricas de risco. Ainda, o Gráfico 6 permite levantar a hipótese de baixa variabilidade do texto entre os períodos de cada empresa. Estes resultados vão ao encontro dos estudos de Ryan (2015), que classificou as informações de risco qualitativas de instituições financeiras americanas como opacas e sem relevância. Conforme mostram os índices quantitativos, houve variação na política de risco das instituições ao longo do período, porém essa variação não pode ser evidenciada na similaridade de cosseno, indicando um afastamento entre o texto sobre gerenciamento de risco da nota explicativa e a política implementada de gerenciamento de risco, ou até mesmo algum nível de reuso do texto, desconectado dos acontecimentos e das decisões tomadas no período.

5 CONSIDERAÇÕES FINAIS

O presente artigo teve como objetivo propor uma medida de risco a partir do conteúdo das notas explicativas, baseada em técnicas de mineração de texto. O texto das notas explicativas tem como função apresentar dados adicionais de forma qualitativa, que auxiliam na compreensão e leitura das demonstrações financeiras. Considerando a área de risco, as notas explicativas podem auxiliar a esclarecer elementos como apetite ao risco, medidas de gerenciamento e estrutura de contingências. Assim, procurou-se expandir as possibilidades de pesquisa considerando as notas explicativas, para além de análises léxicas e volumétricas, utilizando o seu conteúdo.

Neste estudo, foram extraídos trechos das notas explicativas de Financeiras e Agências de Fomento que tratavam da temática de risco. A partir da montagem dessa base de dados, os textos foram pré-processados e foi calculada a distância entre cada trecho de cada demonstração para o trecho sobre gerenciamento de risco da Resolução 4.557/17 do Banco Central. Para o cálculo desta distância, foi utilizada a medida de similaridade de cosseno, pela sua tradição na comunidade de mineração de texto e adequação ao presente trabalho, mais focada no sentido semântico do que na variabilidade do tamanho dos textos.

Para realizar uma comparação da efetividade da medida e uma reflexão sobre seu potencial, a similaridade de cosseno foi comparada com outras duas métricas de risco tradicionais: o índice de Basiléia e o índice Inad90. De uma forma geral, não houve evidência que suportasse a relação entre a similaridade de cosseno e os índices de risco, o que traz algumas reflexões.

O texto selecionado para análise trata exclusivamente sobre risco. Por estar presente nas notas explicativas, esperava-se que ele refletisse o posicionamento e a justificativa do risco das organizações, que se apresenta variável ao longo do período analisado. Contudo, a baixa correlação

com os índices quantitativos denota uma possível distância entre essa informação e a atual prática de gestão de risco das organizações. De certa forma, as instituições podem estar relatando determinadas práticas e realizando outras. A partir dos resultados evidenciados especialmente nas Agências de Fomento, uma outra hipótese levantada pelos autores é que o texto seja reutilizado ou levemente modificado, o que causa uma baixa variabilidade, dentro de uma mesma instituição, entre os períodos, diferente do que ocorre com os índices de risco estudados. Neste cenário, o texto sobre risco nas notas explicativas teria uma função de cumprir um requisito, contudo sem informação relevante para os *stakeholders*.

Este artigo apresenta algumas limitações que buscaram ser contornadas sempre que possível. A indisponibilidade de algumas demonstrações e alguns indicadores fez com que a base ficasse desbalanceada, com algumas instituições com mais registros do que outras. A escolha da medida de distância possui um peso importante, e o teste com outras medidas e outras técnicas (como distância de Jaccard e pré-vetorização dos textos) está no horizonte de estudos futuros. Também como estudos futuros os autores pretendem ampliar a análise realizando um teste de similaridade entre os textos das notas explicativas de cada instituição, com o objetivo de verificar o nível de reuso de texto. Por fim, sugere-se também que outros índices de risco sejam incorporados, bem como outros grupos de instituições financeiras que estão sob a égide da Resolução 4.557/17.

REFERÊNCIAS

Aggarwal, C. C., & Zhai, C. (2012). A survey of text clustering algorithms. In *Mining text data* (pp. 77-128). Springer, Boston, MA.

Banco Central do Brasil (2011). *Resolução nº 2.828*, de 30 de março de 2011.

Banco Central do Brasil (2017). *Resolução nº 4.557*, de 23 de fevereiro de 2017.

Basel Committee on Banking Supervision. (1997). Core principles for effective banking supervision. *Basileia, Suíça: Bank for International Settlements*.

Bergamini Junior, S. (2005). Controles internos como um instrumento de governança corporativa. *Revista do BNDES*, 12(24), 149-188,

Brown, N. C., Crowley, R. M., & Elliott, W. B. (2020). What are you saying? Using topic to detect financial misreporting. *Journal of Accounting Research*, 58(1), 237-291.

Committee of Sponsoring Organizations of the Treadway Commission (2004). *Enterprise Risk Management - Integrated Framework*. AICPA: New York, USA.

CPC. *CPC 26*. Disponível em:

http://static.cpc.aatb.com.br/Documentos/312_CPC_26_R1_rev%2014.pdf Acesso em: 23 jul 2020

Crockford, G. N. (1982). The bibliography and history of risk management: some preliminary observations. *Geneva Papers on Risk and Insurance*, 169-179.

Ding, Kexing & Lev, Baruch & Peng, Xuan & Sun, Ting & Vasarhelyi, Miklos. (2018). On Creating Accounting Estimates using Machine Learning. *SSRN Electronic Journal*.

Dyer, T.; M. Lang; and L. Stice-Lawrence. (2017) The Evolution of 10-K Textual Disclosure: Evidence from Latent Dirichlet Allocation. *Journal of Accounting and Economics* 64 (2017), 221– 45.

Fisher, I. E., Garnsey, M. R., & Hughes, M. E. (2016). Natural language processing in accounting, auditing and finance: A synthesis of the literature with a roadmap for future research. *Intelligent Systems in Accounting, Finance and Management*, 23(3), 157-214.

Fisher, I. E., Garnsey, M. R., Goel, S., & Tam, K. (2010). The role of text analytics and information retrieval in the accounting domain. *Journal of emerging technologies in Accounting*, 7(1), 1-24.

Gomes, M. C., Ferreira, R. R., & Martins, V. A. (2018). O Impacto da OCPC 07 sobre o Tamanho e a Legibilidade das Notas Explicativas de Companhias Brasileiras. *Revista Universo Contábil*, 14(2), 162-184.

Hourrane, O., & Benlahmar, E. H. (2017). Survey of plagiarism detection approaches and big data techniques related to plagiarism candidate retrieval. In *Proceedings of the 2nd international Conference on Big Data, Cloud and Applications* (pp. 1-6).

Huang, A.; R. Lehavy; A. Zang; and R. Zheng. (2018). Analyst Information Discovery and Information Interpretation Roles: A Topic Modeling Approach. *Management Science* 64, 2833– 55.

IFRS. *IAS1*. Disponível em: <https://www.ifrs.org/issued-standards/list-of-standards/ias-1-presentation-of-financial-statements/> Acesso em: 23 jul 2020.

Jales, J. B. S. C., Soares, J. M. M. V., Godeiro, L. L., & de Souza, A. N. M. (2020). Aspectos de Evidenciação das Informações Divulgadas nas Notas Explicativas em Companhias Financeiras. *Revista Conhecimento Contábil*, 10(01).

Li, B., & Han, L. (2013). Distance weighted cosine similarity measure for text classification. In *International Conference on Intelligent Data Engineering and Automated Learning* (pp. 611-618). Springer, Berlin, Heidelberg.

Loughran, T., & McDonald, B. (2016). Textual analysis in accounting and finance: A survey. *Journal of Accounting Research*, 54(4), 1187-1230.

McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D. J., & Barton, D. (2012). Big data: the management revolution. *Harvard business review*, 90(10), 60-68.

Nielsen, S., & Fuertes-Olivera, P. A. (2013). Development in lexicography: From polyfunctional to monofunctional accounting dictionaries. *Lexikos*, 23, 323-347.

Potter, B. N. (2005). Accounting as a social and institutional practice: Perspectives to enrich our understanding of accounting change. *Abacus*, 41(3), 265-289.

Raghavan, R. S. (2003). Risk management in banks. *Chartered Accountant-New Delhi-*, 51(8), 841-851.

- Rahrovi Dastjerdi, A., Foroghi, D., & Kiani, G. H. (2019). Detecting manager's fraud risk using text analysis: Evidence from Iran. *Journal of Applied Accounting Research*, 20(2), 154-171.
- Ryan, S. G. (2012). Risk reporting quality: Implications of academic research for financial reporting policy. *Accounting and business research*, 42(3), 295-324.
- Ribeiro, P. A. N. (2004). Agências de fomento – ambientes institucional, legal e operacional. *Revista da Desenharia*, Salvador, 1.
- Saumya, S., Singh, J. P., Baabdullah, A. M., Rana, N. P., & Dwivedi, Y. K. (2018). Ranking online consumer reviews. *Electronic Commerce Research and Applications*, 29, 78-89.
- Sawanobori, C. (2013). Influence of fundraising infeasibility on accounting changes: An investigation into early financial reporting practices using text mining. *Journal of Accounting and Organizational Change*, 9(4), 471-489.
- Swandi, I., Suwardi, I. S., & Maulidevi, N. U. (2017). Automatic data interpretation in accounting information systems based on ontology. *Telkomnika (Telecommunication Computing Electronics and Control)*, 15(4), 1817-1829.
- Vieira, J. A. G., Pereira, H. F. S., & Pereira, W. N. A. (2012). Histórico do Sistema Financeiro Nacional. *Revista Científica e-Locução*, 1(02), 146-162.
- Yang, M., Adomavicius, G., Burtch, G., & Ren, Y. (2018). Mind the gap: Accounting for measurement error and misclassification in variables generated via data mining. *Information Systems Research*, 29(1), 4-24.