


## Notas explicativas explicam? Análise da comunicação do gerenciamento de risco a partir de técnicas de text mining

Do explanatory notes explain? Risk management disclosure analysis using text mining

¿Notas explicativas explican? Análisis de la comunicación de la gestión de riesgos a partir de técnicas de text mining

**Carla Bonato Marcolin\***

Doutora em Administração (UFRGS) e Pós-Doutoranda na Escola de Administração de Empresas de São Paulo (FGV), Professora da Faculdade de Gestão e Negócios (UFU), Uberlândia/MG, Brasil  
 cbmarcolin@gmail.com

<https://orcid.org/0000-0003-0260-5073> 

**Fernanda da Silva Momo**

Doutora em Administração (UFRGS) Professora da Faculdade de Ciências Econômicas (UFRGS), Porto Alegre/RS, Brasil  
 fernanda.smomo@gmail.com

<http://orcid.org/0000-0002-6512-5280> 

**Enrico Dalla Riva**

Mestre em Controladoria e Contabilidade (USP) e Doutorando na Escola de Administração de Empresas de São Paulo (FGV), São Paulo/SP, Brasil  
 enrico.riva@bcb.gov.br

<https://orcid.org/0000-0001-7032-6039> 

**João Luiz Becker**

Doutor em Management Science na University of California at Los Angeles (UCLA) Professor Titular de Tecnologia e Ciência de Dados na EAESP (FGV), São Paulo/SP, Brasil  
 beckerjoaluz@gmail.com

<https://orcid.org/0000-0003-4176-7374> 

**Endereço do contato principal para correspondência\***

Av. João Naves de Ávila, 2121, Bairro Santa Mônica, CEP: 38408-144 – Uberlândia/MG, Brasil

### Resumo

O estudo objetivou analisar a efetividade da comunicação do gerenciamento de risco nas notas explicativas considerando técnicas de *text mining*. A partir da análise de 241 trechos de texto de 32 instituições financeiras, utilizou-se a similaridade de cosseno como medida de distância para avaliar a relação da variabilidade dos textos com os índices de risco Basileia e Inad90. Os resultados indicam que as notas analisadas não estão em conformidade com a OCPC 07, pois seu conteúdo não reflete a flutuação dos índices de risco. A utilidade para o usuário em processos de tomada de decisões pode ser prejudicada, dado o indício de desconexão entre os textos e a política de riscos. Apresentam-se casos de repetição integral dos trechos de texto. Como contribuição, esse artigo avança nos estudos textuais das demonstrações financeiras, indo além de análises léxicas e volumétricas, trazendo o conteúdo semântico em volume e sua relação com indicadores financeiros.

**Palavras-Chave:** Notas explicativas; *Text mining*; Gerenciamento de risco

### Abstract

The study aimed to analyze the effectiveness of risk management communication in the explanatory notes considering text mining techniques. Based on the analysis of 241 text excerpts from 32 financial institutions, we used the cosine similarity as a distance measurement to assess the relationship between the variability of the texts and the Basel and Inad 90 risk indices. The results indicate that the analyzed notes are not in compliance with OCPC 07, as their content does not reflect the fluctuation of risk indices. The benefit of the notes for the user in decision-making processes may be impaired, given the evidence of disconnection between the texts and the risk policy. Cases of complete repetition of the text excerpts are presented. As a contribution, this article advances in the textual studies of financial statements, going beyond lexical and volumetric analyses, bringing the semantic content in volume and its relationship with finance indicators.

**Keywords:** Explanatory notes; Text mining; Risk management

### Resumen

El estudio tuvo como objetivo analizar la efectividad de la comunicación de la gestión de riesgos en las notas explicativas considerando técnicas de minería de textos. A partir del análisis de 241 extractos de texto

de 32 instituições financeiras, se utilizó la similitud del coseno como medida de distancia para evaluar la relación entre la variabilidad de los textos y los índices de riesgo de Basilea e Inad90. Los resultados indican que las notas analizadas no están en conformidad con OCPC 07, ya que su contenido no refleja la fluctuación de los índices de riesgo. La utilidad para el usuario en los procesos de toma de decisiones puede verse afectada, dada la evidencia de desconexión entre los textos y la política de riesgos. Se presentan casos de repetición completa de extractos del texto. Como contribución, este artículo avanza en los estudios textuales de los estados financieros, yendo más allá de los análisis léxicos y volumétricos, trayendo el contenido semántico en volumen y su relación con los indicadores financieros.

**Palabras-Clave:** Notas explicativas; *Text mining*; Gestión de riesgos

## 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. 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 (Fisher et al., 2016; Sawanobori, 2013; Yang et al., 2018).

A área de mineração de texto (*Text Mining*) tem contribuído 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 analisar a efetividade da comunicação do gerenciamento de risco de instituições financeiras a partir do conteúdo das notas explicativas, considerando técnicas de mineração de texto.

Assim, pretende-se analisar a possibilidade de evidenciar a efetividade de comunicação de gerenciamento de risco de instituições financeiras a partir de uma análise que tenha como fonte constitutiva o conteúdo textual das notas explicativas. Isso porque as notas explicativas são o espaço de detalhamento 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 relevantes para a compreensão da informação financeira da organização (CPC, 2011; 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. Assim como há estudos que usam técnicas de *data analytics* para aprimoramento de estimativas e escrita, análise de discursos de gestores, previsão de ações fraudulentas, entre outros (Ding et al., 2019; Loughren & McDonald, 2016), esse estudo, ao analisar a comunicação sobre o posicionamento de risco presente nas notas explicativas, visa identificar se essa fonte de texto pode servir de apoio para tomada de decisão para os diversos *stakeholders* em relação às instituições financeiras, ou se as notas explicativas ainda não representam fidedignamente o posicionamento das instituições frente suas práticas de gestão de risco o que implicaria, nesse sentido, que as notas explicativas não estariam explicando.

Logo, o uso de notas explicativas para análise de risco pode fornecer uma nova visão sobre a área de risco para órgãos de controles e investidores. Além disso, traz-se a discussão sobre o uso do espaço das notas explicativas como fonte de promoção da *accountability* e forma de representar fidedignamente as práticas empresariais quanto a gestão de risco, ou se este espaço ainda é utilizado de forma a cumprir normativos, sem uma maior preocupação em explicar e se comunicar sobre os posicionamentos da organização sobre o tema com os *stakeholders* (Hendriksen & Van Breda, 2010; Nakagawa et al., 2008).

Após esta introdução, a próxima seção apresenta um referencial teórico sobre divulgação contábil, 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 apresentam-se três seções em que se aborda o processo de divulgação contábil (2.1), as informações divulgadas sobre gerenciamento de riscos nas instituições financeiras (2.2) e as notas explicativas (2.3).

## 2.1 Divulgação Contábil

O processo de evidenciação da informação contábil encontra amparo teórico nas motivações para a adoção de padrões de transparência nas organizações. Verrecchia (2001) sugere na literatura três categorias de estudo de fatores de incentivo para a divulgação contábil: (a) baseada em associação, fundamentada em estudos sobre os efeitos da informação na tomada de decisão de investidores; (b) baseada em julgamento, em função do poder dos gestores em divulgar ou omitir informações internas; e (c) baseada em eficiência, que indica que qualquer informação é melhor que a ausência dela. Verrecchia (2001) sugere o desenvolvimento de uma teoria que envolve todos estes fatores, a partir do processo de redução de assimetrias informacionais.

Um exemplo da primeira abordagem é a análise da relação entre divulgação de informações contábeis e alterações nos valores das ações das empresas. A segunda abordagem busca investigar os incentivos dos gestores ou empresas para promover a divulgação, a exemplo de uma eventual redução de valor da empresa quando da não-divulgação de informações relevantes. Por fim, na abordagem com base em eficiência, as empresas buscam a redução da assimetria informacional no sentido de reduzir esse fator do seu custo de capital (Yamamoto & Salotti, 2006).

Dantas et al. (2005) investigaram a dualidade entre os benefícios da divulgação e a relutância das organizações em divulgar informações internas ao público. Não obstante a melhora na valorização das entidades e confiabilidade de credores e investidores, percebe-se uma aversão ao aumento do grau de evidenciação com o argumento de proteção da informação estratégica. Com relação às instituições financeiras, particularmente, os autores consideram que a exigência de uma divulgação apropriada se faz ainda mais necessária, no sentido de dar conhecimento aos usuários dos riscos aos quais a instituição está exposta (Dantas et al., 2005).

O fenômeno da divulgação contábil para o mercado pode ter característica voluntária ou obrigatória. Defensores da divulgação de caráter obrigatório consideram que as entidades não têm disposição em aumentar o nível de divulgação sem imposição legal (Yamamoto & Salotti, 2006). O presente estudo considera aspectos da segunda proposição de Verrecchia (2001), baseada na discricionariedade dos gestores sobre a divulgação da informação contábil, diante de uma obrigatoriedade de prestação de informações definida pelo regulador.

## 2.2 Informações sobre Gestão de Riscos

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 para promover boas práticas corporativas de procedimentos internos, tornando-se referência. O COSO (*Committee of Sponsoring Organizations of the Treadway Commission*), que assessora a *Security Exchange Commission* (SEC) nos Estados Unidos, foi uma dessas organizações que produziu material pertinente ao tema (Benjamin Jr., 2005). Na sua definição de gestão de riscos apresenta-a como 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 de 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 voltados 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 adequação de capital ao seu perfil de risco (BCBS, 1997).

No Brasil, as instituições financeiras são regulamentadas pelo Banco Central do Brasil (BCB), que segue as diretrizes emanadas pelo Comitê de Basileia. Atualmente, as exigências previstas na regulamentação sobre gestão de riscos estão dispostas na Resolução 4.557/17, que exige das instituições financeiras uma adequada estrutura de gerenciamento integrado de riscos compatível com suas atividades. A regulação exige, em seu artigo 56, a divulgação de sua estrutura de 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).

Torres e Galdi (2013) já exploraram o nível de aderência de divulgação de informações produzidas pelas instituições financeiras sobre gestão de riscos em relação aos normativos emanados pelo regulador no Brasil, e notaram possibilidades significativas de melhoria nos níveis de *disclosure*. Ryan (2012)

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.

Observa-se que na literatura já é possível encontrar evidências de que o *disclosure* de informações sobre gerenciamento de riscos de instituições financeiras apresenta relação positiva com: valorização das ações dos bancos (Jizi & Dixon, 2017), resultados financeiros (Oino, 2019), estabilidade (Del Gaudio et al., 2020) e governança (Nahar et al., 2020). Ademais, o nível de divulgação sobre riscos tem poder preditivo sobre a classificação de risco das instituições financeiras (Elamer et al., 2021) permitindo dirimir incertezas (Kim et al., 2019). Inclusive a comunicação do processo de gestão de riscos é parte integrante determinante de um processo de implementação de cultura de riscos e governança nas instituições bancárias (Schmitt, 2017).

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 et al., 2020; Dyer et al., 2017; Huang et al., 2018).

### 2.3 Notas Explicativas

As notas explicativas, segundo o Pronunciamento Técnico 26 - Apresentação das Demonstrações Contábeis (CPC 26 (R1)), “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, 2011), o que evidencia o caráter complementar mas ao mesmo tempo distintivo das notas explicativas. Da mesma maneira, o *International Financial Reporting Standards* (IFRS) 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). Destacando a importância das informações apresentadas nas notas explicativas como fonte de transparência e *accountability*.

Ao identificar situações de presença de informações irrelevantes e ausência de informações relevantes nas notas explicativas no Brasil, o CPC emitiu orientação técnica (OCPC 07) na qual conclui que já existem diretrizes para evidenciação de notas explicativas, tanto no CPC 26 como no Pronunciamento Conceitual Básico - Estrutura Conceitual para Elaboração e Divulgação de Relatório Contábil-Financeiro (CPC, 2014), normativo recepcionado pelo Banco Central do Brasil pela Resolução 4.144/12 (BCB, 2012), como regulador do sistema financeiro. A OCPC 07 destaca a importância da referida peça contábil no provimento de informações materiais e relevantes que sejam úteis aos usuários da informação quando da tomada de decisões nas entidades, ressaltando a necessidade de se evitar excesso de informações irrelevantes que possam desviar do objetivo de informar, confundido o usuário (CPC, 2014). Nesse sentido, já foi identificado no Brasil um processo de institucionalização de notas explicativas específicas por meio de regulamentação, no sentido de aumentar a capacidade de controle e fiscalização de mercados de riscos (Santos et al., 2010).

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 (Fisher et al., 2010; Porter, 2005).

O avanço em técnicas de mineração de texto tem auxiliado a mudar esse cenário. Fisher et al. (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 (Nielsen & Fuertes-Oliveira, 2013; Swandi et al., 2017).

Apesar de existirem diversos trabalhos nesta corrente, há críticas que apontam a falta da análise semântica quantitativa e em volume, reduzindo a capacidade de integração destes dados com outras informações contábeis (Gomes et al., 2018). Esta crítica 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 o texto para realizar análises e inferências atuais, como preço das ações e previsão de performance e fraude (Loughran & McDonald, 2016; Rahrovi et al., 2019). 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 para gestores e *stakeholders*.

Contudo, apesar de haver uma área de pesquisa madura no exterior (ver Fisher et al. (2010) para uma revisão completa de técnicas), o campo ainda é incipiente no Brasil, considerando especificamente o texto das notas explicativas. Os trabalhos encontrados são mais focados em análises léxicas e volumétricas em relação ao tamanho das notas e quantidade de palavras (Gomes et al., 2018), e não no seu conteúdo, como proposto nesta pesquisa. Ainda, alguns trabalhos aprofundaram a questão da legibilidade, que envolve uma análise quantitativa do conteúdo (palavras, frases e parágrafos) para cálculo do índice de Flesch (Gomes et al., 2018; Holtz & Santos, 2020).

Diferentemente de uma análise sintática, no presente artigo, avaliou-se o conteúdo semântico, relacionado ao significado das palavras. Buscou-se explorar elementos relacionados com as palavras no seu contexto, considerando assim o significado das frases, e não apenas a quantidade ou frequência das palavras. Desta forma, pode-se dizer que há uma aproximação da experiência de leitura de um analista e dos resultados da técnica utilizada. Ao buscar considerar as palavras no seu contexto, textos que possuam significados semelhantes (ainda que utilizem palavras distintas) serão considerados semelhantes, o que difere dos trabalhos anteriores que analisaram notas explicativas. O principal mecanismo que permite essa abordagem é o cálculo da matriz de termos e documentos ponderada pelo índice TF-IDF, detalhado na próxima seção.

Ademais, observa-se que apesar de ser madura a área que trabalha com a análise de relatórios anuais divulgados pelas empresas e sua relação com prevenção e previsão de riscos, poucos trabalhos trazem esta abordagem considerando especificamente as notas explicativas. Para preencher estas lacunas de pesquisa, este artigo busca propor uma análise que considere o conteúdo semântico das notas explicativas para instituições financeiras. O objetivo é contribuir e expandir trabalhos anteriores que tenham analisado as notas explicativas, porém considerando uma abordagem diferente. Para isso, utilizou-se o método descrito na próxima seção.

### 3 Método

A presente pesquisa caracteriza-se como quantitativa e exploratória. O objetivo é analisar a efetividade da comunicação sobre o gerenciamento de risco de instituições financeiras, considerando o conteúdo das notas explicativas. Para tal, foram coletados trechos das notas explicativas que dissertam sobre a estrutura de gerenciamento de riscos, usualmente localizados ao final das demonstrações, sob o título “Gerenciamento de Riscos” ou “Processos de Gestão”. 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 (<https://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. Todas as demonstrações estavam em português.

A escolha da amostra se deu por dois motivos principais. 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 BCB (<https://www3.bcb.gov.br/efdada/>). O objetivo da comparação foi obter uma visão quantitativa e objetiva da política de gerenciamento de risco de cada instituição, uma vez que estes índices expressam o apetite ao risco.

Das 32 instituições integrantes da amostra final, algumas não possuíam os 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 (trechos de texto), 133 de Financeiras e 108 de Agências de Fomento.

Para analisar o conteúdo sobre risco nas notas explicativas, 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 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. O objetivo com essa análise foi de compreender a capacidade do trecho da nota explicativa ilustrar o gerenciamento de risco, considerando que uma variabilidade nas medidas de risco poderia ser evidenciada por uma variabilidade entre o trecho do texto e a resolução. 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. Para gerar os vetores normalizados, os textos foram transformados através de ferramentas de *text mining* em R através do RStudio. O primeiro passo foi transformar o conjunto de 242 textos (241 trechos de notas explicativas e 1 trecho da normativa) em um *corpus*, para que fosse possível o pré-processamento. Este pré-processamento tem como objetivo diminuir a dimensionalidade dos dados e o ruído que ocorre ao trabalhar com dados em texto, uma vez que conforme a Lei de Zipf, grande parte da informação é expressa com uma pequena parte do vocabulário (Manning et al., 2008). O pré-processamento incluiu a remoção de *stopwords* (palavras muito frequentes, mas com baixo significado semântico, como artigos e preposições), a transformação para letras minúsculas e a remoção de espaços em branco adicionais, conforme melhores práticas reportadas em literatura (Aggarwal & Zhai, 2012).

O segundo passo consistiu em gerar uma matriz  $X$  de termos e documentos, contendo nas linhas, os termos únicos do conjunto de textos e nas colunas, um índice indicando cada um dos documentos. De forma a respeitar as particularidades de cada grupo, optou-se por gerar duas matrizes, uma para Financeiras, de dimensão  $2.272 \times 134$ , e uma para Agências de Fomento, de dimensão  $2.365 \times 109$ , sendo a última coluna de cada matriz correspondente ao trecho da resolução, igual para ambas matrizes, e as linhas indicando os termos únicos de cada *corpus*.

Para preenchimento da célula  $a_{ij}$  de cada matriz, foi utilizado o índice TF-IDF (do inglês, *term-frequency* e *inverse document-frequency*). Este índice pondera a matriz através da valorização de termos mais exclusivos e desvalorização de termos muito frequentes, que não são capazes de distinguir os documentos 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). O índice TF-IDF é calculado conforme segue (Aggarwal & Zhai, 2012):

$$TF - IDF = \frac{f_{t,d}}{\max_{t,d}} \times \frac{N}{n_t}$$

A primeira parte da fórmula calcula a frequência relativa de cada termo, sendo  $f_{t,d}$  a frequência de cada termo  $t$  em cada documento  $d$  e  $\max_{t,d}$  o máximo de frequência de um único termo. A segunda parte da fórmula calcula o log da razão do total de documentos  $N$  e da quantidade de documentos que contém aquele termo ( $n_t$ ). O objetivo é ponderar a frequência de cada termo (primeira parte) com sua capacidade de identificar grupos de documentos (segunda parte), uma vez que termos muito frequentes entre os

documentos recebem valores menores, e termos muito frequentes em um determinado conjunto de documentos tendem a valores maiores.

Os vetores  $U$  considerados para o cálculo da similaridade de cosseno para as Financeiras consistiram nas colunas 1 a 133 da respectiva matriz termos e documentos, e o vetor  $V$  a coluna 134 da mesma matriz, que representa o trecho da resolução. Para as Agências de Fomento, seguiu-se similar processo, sendo as colunas 1 a 108 os vetores  $U$ , e a coluna 109 o vetor  $V$ . Dessa forma, os vetores possuíam o mesmo tamanho e consideravam a presença ou ausência de palavras mais relevantes dado a ponderação TF-IDF utilizada. Cada similaridade foi calculada de forma individual, ou seja, as colunas  $1 \times 134, 2 \times 134, \dots, 133 \times 134$  para as Financeiras e  $1 \times 109, 2 \times 109, \dots, 108 \times 109$  para as Agências de Fomento. A próxima seção apresenta a análise dos resultados.

## 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). Apesar da natureza de atividade de cada uma dessas classes ser distinta, pois enquanto as Financeiras atuam mais voltadas ao crédito, as Agências de Fomento possuem uma proposta financeira mais social, para os objetivos do trabalho considerou-se importante observar os dados em conjunto. Assim, esta seção está dividida em duas subseções. A primeira apresenta uma abordagem mais descritiva dado o conteúdo dos textos, e a segunda uma análise da relação dos textos com medidas de risco.

### 4.1 Análise Descritiva

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 (BCB, 59,66), 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. Operam somente com a captação de recursos por meio de emissão de títulos próprios ou letras de câmbio (Viera et al., 2012). Os 134 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.

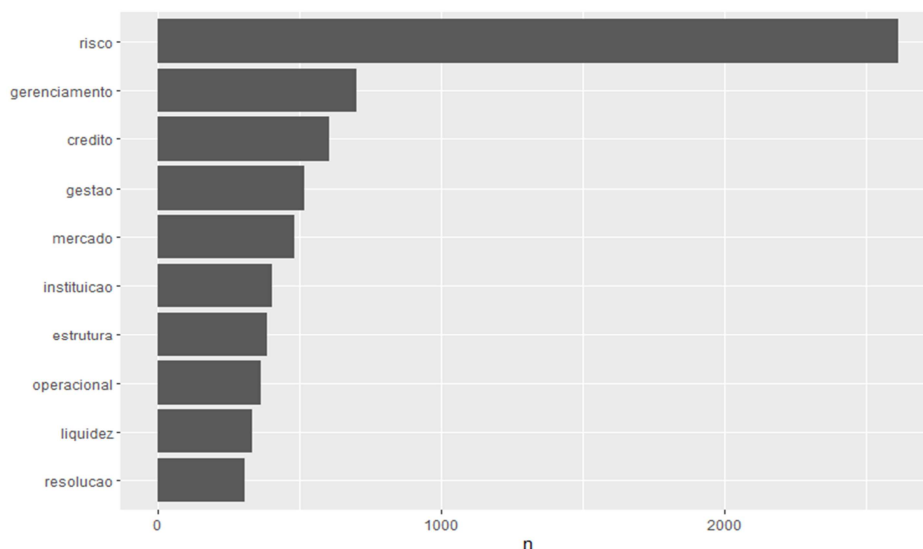


Figura 1 - Top 10 palavras mais frequentes no *corpus* Financeiras

Fonte: Dados da Pesquisa

Apesar da Resolução 4.557/17 (BCB, 2017) mencionar seis tipos de riscos diferentes (crédito, mercado, variação taxas de juros, operacional, liquidez e socioambiental) e 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, foi realizada uma análise de  $n$ -gramas, com  $n$  variando de dois até cinco. Essa análise permite visualizar os termos mais frequentes nas frases, complementando análises do tipo *bag of words*, que não consideram a ordem das palavras. A Tabela 1 apresenta as bi e penta-gramas mais frequentes para Financeiras.

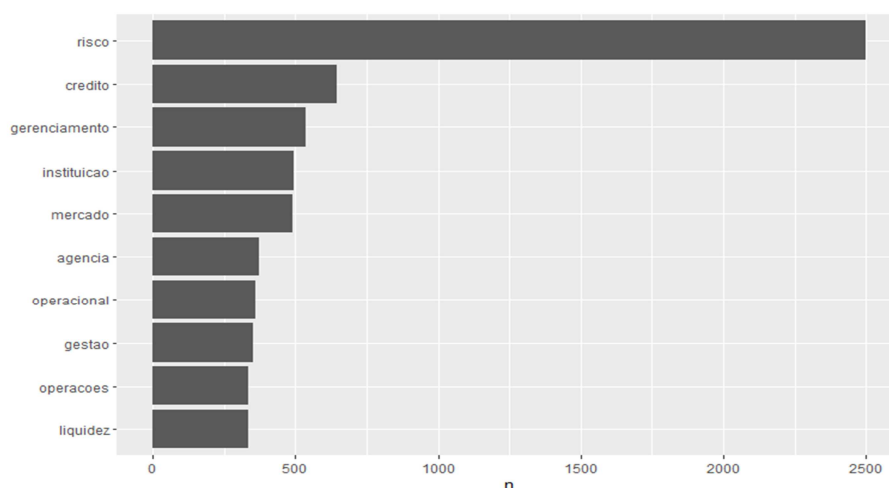
**Tabela 1:**  
**Conjunto de n-termos (n=2 e n=5) em ordem mais frequentes para Financeiras**

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 CMN	109	(...) identificar avaliar monitorar controlar mitigar (...)	22
Banco Central	98	(...) risco mercado liquidez operacional crédito (...)	22
Controles Internos	93	(...) risco liquidez o risco liquidez (...)	21

Fonte: Dados da Pesquisa

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

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 perceber 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 nomenclatura 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 (BCB, 2017). Os resultados da análise de n-gramas, que aprofunda a percepção sobre os termos mais frequentes, encontram-se na Tabela 2.

**Tabela 2:**  
**Conjunto de n-termos (n=2 e n=5) em ordem mais frequentes para Agências de Fomento**

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
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

Fonte: Dados da Pesquisa



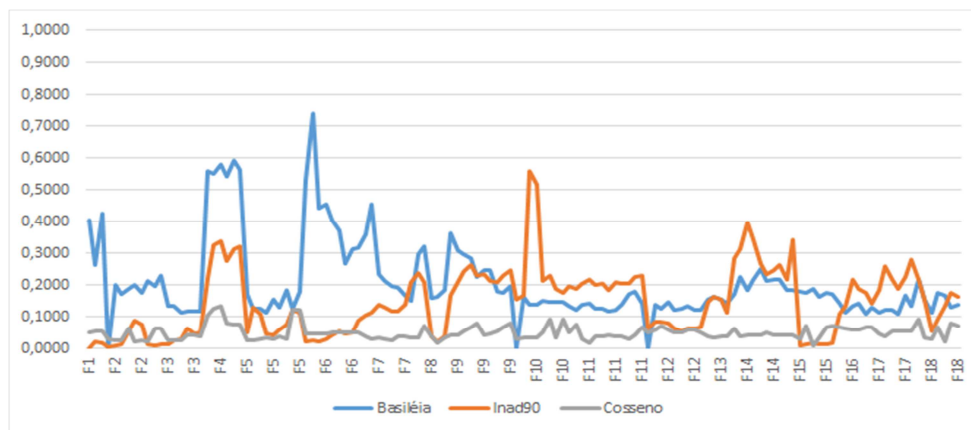
A maior variabilidade de termos encontrada na Tabela 2 em relação à Tabela 1 denota que este grupo (Agências de Fomento) possui uma discussão mais ampla presente nas notas explicativas, no que condiz à estrutura de gerenciamento de risco. Apesar de poder ser um reflexo das diferentes atuações das instituições 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.

#### 4.2 Variabilidade Risco versus Texto

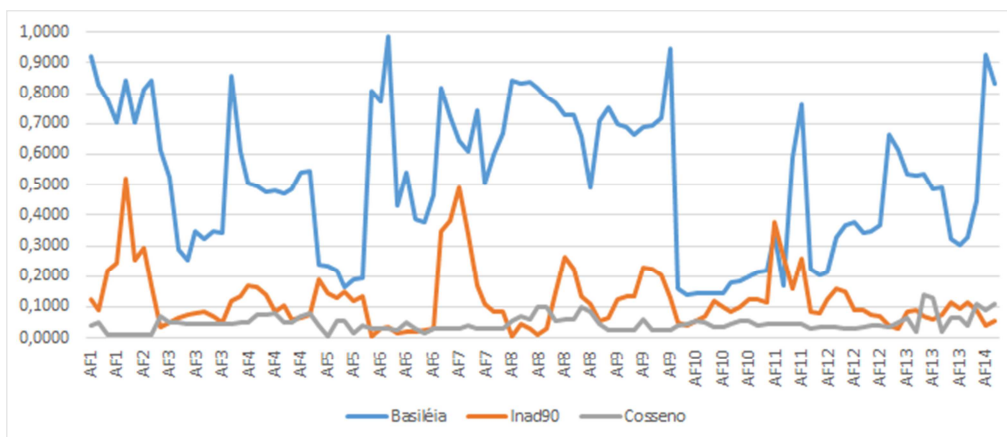
Nesta subseção, avaliam-se os textos das notas explicativas das Financeiras e Agências de Fomento como fonte de dados para entender o posicionamento dessas instituições em suas políticas de risco. O pressuposto é que as notas explicativas apresentam informações fidedignas que fornecem uma maior transparência sobre os posicionamentos das organizações frente a gestão de riscos, uma vez que as notas explicativas devem ser uma descrição narrativa das políticas adotadas pelas empresas (CPC, 2020; IFRS, 2020). Nesse sentido, espera-se que, ao haver uma mudança nessa política, independente de seus motivadores serem internos ou externos, deve haver mudanças substanciais nas notas explicativas, visto que nas notas explicativas devem-se prover informações materiais e relevantes aos usuários da informação (CPC, 2014) e; portanto, se o apetite ao risco aumenta ou diminui não se espera um mesmo texto nas notas explicativas.

Para compreender a relação entre os textos e o comportamento de risco, uma primeira análise foi verificar se havia uma variação conjunta entre a similaridade e as medidas de risco. A Figura 3 apresenta o índice de Basiléia, o Inad90 e a similaridade de cosseno entre os trechos de texto de cada uma das Financeiras (F1 até F18) e a resolução, no eixo x. No eixo y, encontra-se uma escala percentual, respeitando o intervalo dos índices e da similaridade de cosseno (de 0 até 1). Pode-se observar que, com exceção de poucas instituições (especialmente F3, F4 e F5) nas quais há uma tendência conjunta, os índices não parecem estar conectados com o conteúdo dos textos em relação à resolução.

O mesmo pode ser observado em relação à Figura 4, que apresenta a mesma análise para as Agências de Fomento. Percebe-se que, sem exceções, as medidas de risco estão alinhadas quanto à variabilidade, porém descoladas do movimento da similaridade de cosseno, que parece oscilar muito pouco em comparação com o risco em si. Essa baixa relação levanta a hipótese de que os textos não variam tanto quanto os próprios índices de risco.



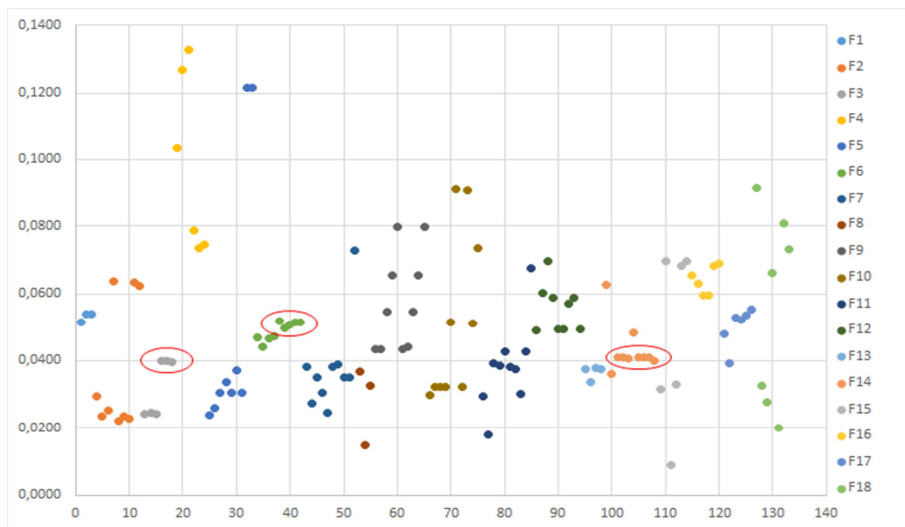
**Figura 3 - Índice de Basiléia, Inad90 e Similaridade de Cosseno para Financeiras**  
Fonte: Dados da Pesquisa



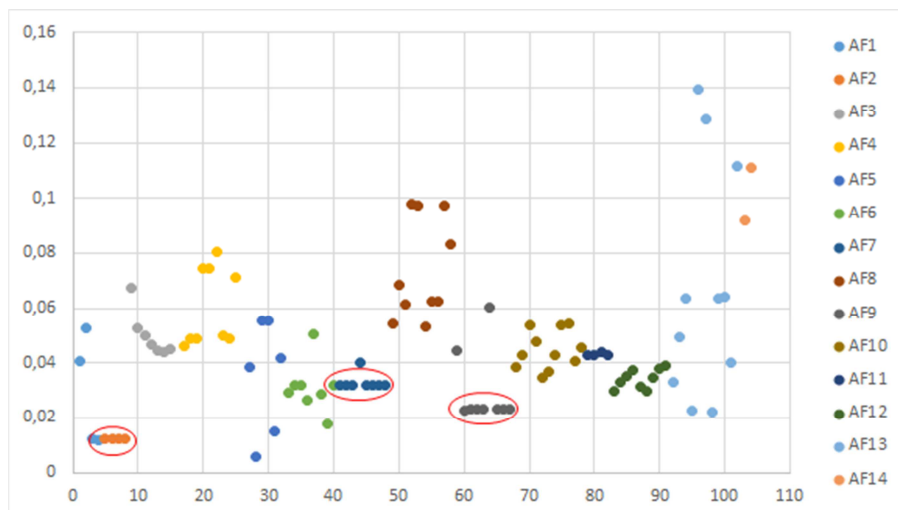
**Figura 4 - Índice de Basiléia, Inad90 e Similaridade de Cosseno para Agências de Fomento**  
Fonte: Dados da Pesquisa

Para aprofundar essa hipótese, analisou-se a dispersão entre as similaridades. Considerando que o texto deve informar sobre a política de riscos, e sabendo que este risco é variável ao longo do tempo, esperava-se que houvesse dispersão intra-empresa, ou seja, entre os trechos de texto de uma mesma empresa para o trecho de texto da resolução. As Figuras 5 e 6 apresentam essa análise para Financeiras e Agências de Fomento, respectivamente. Na Figura 5, o eixo x representa cada um dos 133 trechos de texto de Financeiras. Na Figura 6, o eixo x representa cada um dos 108 trechos de texto de Agências de Fomento. Para ambos, o eixo y representa a similaridade de cosseno, considerando que todos os valores ficaram abaixo de 0,2.

Como pode ser percebido pelas Figuras 5 e 6, o *output* esperado não pode ser confirmado para todas as instituições. Com poucas exceções, há no geral baixa dispersão entre as diferentes instituições, o que pode denotar um texto de cunho mais genérico utilizado para informar sobre o gerenciamento e a política de risco. Além disso, considerando que o trecho do texto da norma que referencia gerenciamento de riscos é sempre igual, estão destacados ainda nas Figuras grupos de texto de uma mesma instituição com baixíssima variabilidade, indicando um texto altamente similar. Essas duas evidências reforçam a hipótese de que o texto não representa a realidade da política de gerenciamento de risco, tendo em vista a evidência de variabilidade dos índices.



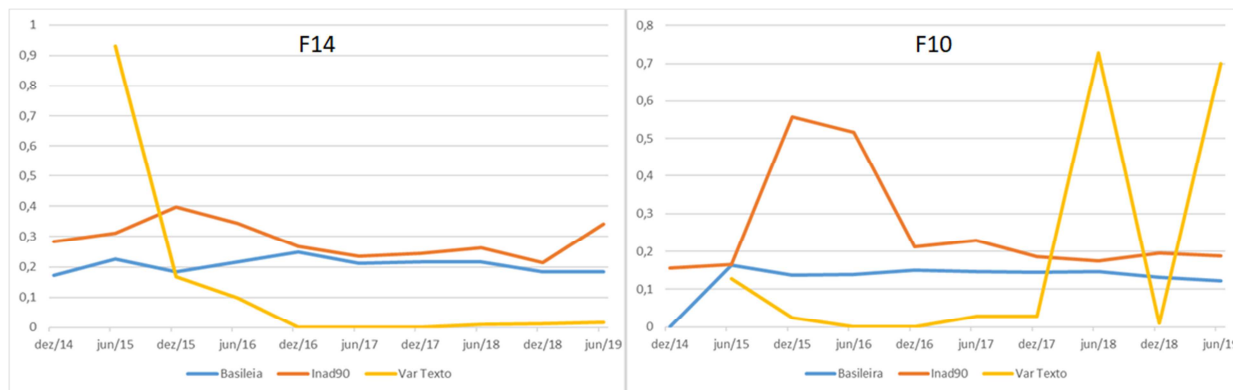
**Figura 5 - Gráfico de Dispersão dos valores de Similaridade de Cosseno nas Financeiras**  
(As cores representam as 18 instituições analisadas, com destaque para instituições com baixíssima variabilidade)  
Fonte: Dados da Pesquisa



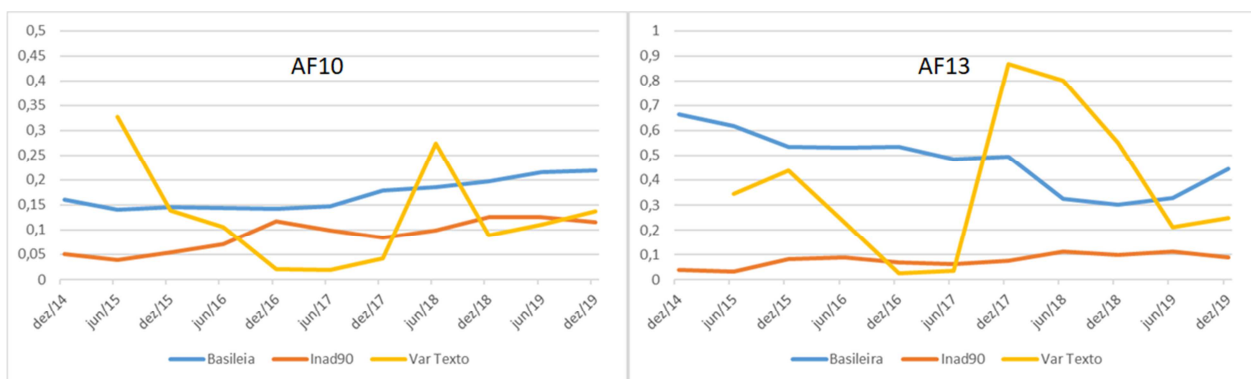
**Figura 6 - Gráfico de Dispersão dos valores de Similaridade de Cosseno nas Agências de Fomento**  
(As cores representam as 14 Agências analisadas, com destaque para instituições com baixíssima variabilidade)  
Fonte: Dados da Pesquisa

Para aprofundar na análise da relação entre a estrutura dos trechos de texto das notas explicativas e as medidas de risco, apresentam-se as Figuras 7 e 8. Foram selecionadas as 2 instituições de cada grupo que possuíam maior quantidade de registros (trechos de texto e medidas de risco) e, dentre estas, uma que possuía a maior variabilidade total (maior distância entre os trechos de texto) e uma que possuía a menor variabilidade total, de forma a evitar vieses.

O objetivo foi compreender como a variabilidade das medidas de risco estava (ou não) associada com a variabilidade entre os textos de uma mesma instituição. Esperava-se que quanto maior a mudança entre um período e outro nas medidas de risco, maior seria a distância entre os trechos de textos das respectivas demonstrações, uma vez que as notas explicativas são o espaço para melhor descrever esse tipo de alteração (IFRS, 2020). Contudo, tal relação não pode ser observada: as figuras demonstram que apesar de haver variabilidade nas medidas de risco entre um período e outro, o texto se manteve muito similar.



**Figura 7 - Variabilidade Medidas de Risco versus Texto das Financeiras F14 e F10**  
(A linha amarela representa a variabilidade do texto, ou seja,  $1 - \text{cosseno}(U, V)$ )  
Fonte: Dados da Pesquisa

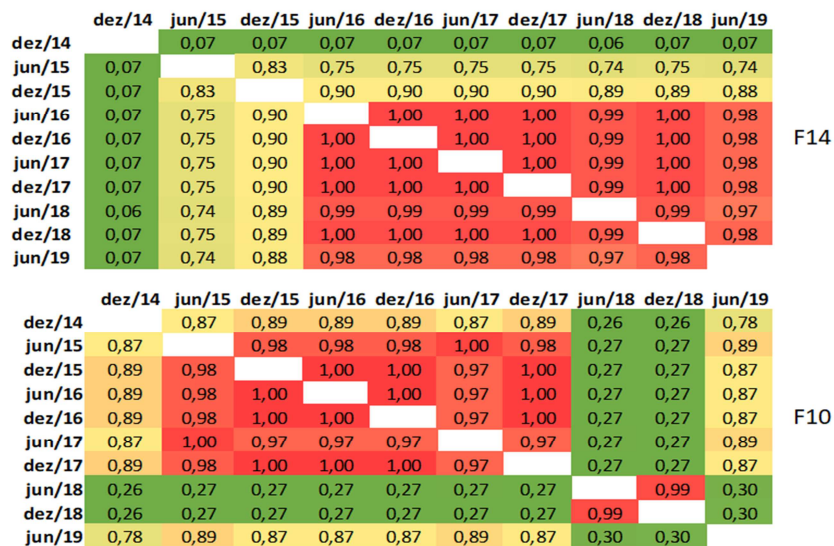


**Figura 8 - Variabilidade Medidas de Risco versus Texto das Agências de Fomento AF10 e AF13**  
(A linha amarela representa a variabilidade do texto, ou seja,  $1 - \text{cosseno}(U, V)$ )  
Fonte: Dados da Pesquisa

Na Figura 7, é possível perceber que ambas as instituições possuem variabilidade nas medidas de risco, porém não possuem respaldo na variabilidade do texto. A instituição F14 apresentou variabilidade nos seus trechos entre dez/14 e dez/16. Contudo, de dez/16 até jun/19, o texto foi muito similar, com variabilidade praticamente nula. Já a instituição F10 apresentou textos muito similares entre jan/14 e jun/17 mesmo com a alta variabilidade na medida de risco Inad90, neste mesmo período.

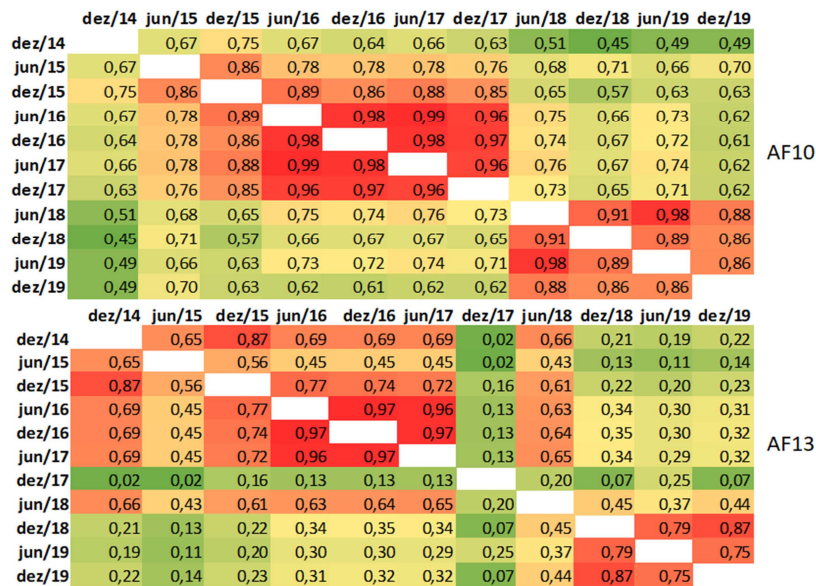
Na Figura 8 ocorre semelhante fenômeno. Destaca-se que o relativo equilíbrio das medidas de risco contrasta com a oscilação dos textos, o que não é em si um indicador de prejuízo da informação, uma vez que a política de gerenciamento de risco pode ser alterada com o objetivo de estabilidade do próprio risco. Contudo há períodos de variabilidade de risco, como na Agência AF10, entre dez/16 e dez/17, que não se refletem no texto.

Considerando a diferença nas variabilidades observadas, optou-se por aprofundar alguns casos individuais. Assim, as mesmas instituições das Figuras 7 e 8 foram analisadas através de um mapa de calor (Figuras 9 e 10). O mapa de calor apresenta as semelhanças dos trechos de texto entre os períodos. Como números mais próximos de 1 apresentam maior perigo de reuso de texto, estes foram sinalizados com cores quentes, ao passo que valores menores (textos mais distantes entre si) estão na escala de verde.



**Figura 9 - Similaridade entre os textos das Financeiras F14 (acima) e F10 (abaixo)**  
 (A diagonal foi removida para melhor visualização, e continha o valor de 1, indicando máxima semelhança)  
 Fonte: Dados da Pesquisa

Na Figura 9 percebe-se uma tendência contrária das duas instituições. Enquanto a F10 parece estar modificando seus textos mais recentemente, a F14 parece ter seus textos cada vez mais semelhantes entre si. Ambas, contudo, apresentam valores de 1,00 fora da diagonal, o que indica trechos de texto exatamente iguais. Apesar de ser razoável considerar que a política de gerenciamento de risco não sofrerá alterações semestrais, um texto exatamente igual não apresenta variabilidade, portanto, não há ali informação relevante para o tomador de decisão, o que contraria o papel desse tipo de dado em demonstrações financeiras (Fisher et al., 2010).



**Figura 10 - Similaridade entre os textos das Agências de Fomento AF10 (acima) e AF13 (abaixo)**  
 (A diagonal foi removida para melhor visualização, e continha o valor de 1, indicando máxima semelhança)  
 Fonte: Dados da Pesquisa

A Figura 10 ressalta uma semelhança ainda maior nas Agências selecionadas. No caso da AF10, apesar de ao longo do tempo os textos se tornarem mais distantes (maior presença de indicadores verdes nas bordas), os textos entre os períodos são muito semelhantes (concentração de vermelhos na diagonal), especialmente entre jun/16 e dez/17. Pode ser levantada a hipótese de que períodos com a mesma empresa externa assessorando ou auditando as demonstrações possam estar na origem da maior similaridade entre períodos próximos. O mesmo acontece com a AF13, ainda que sem o mesmo padrão longitudinal, mas sim concentrado em alguns períodos, como de dez/15 à jun/17 e de dez/18 à dez/19.

Os achados permitem concluir que há uma baixa relação do texto com as métricas de risco. As Figuras 5 e 6 permitiram levantar a hipótese de baixa variabilidade do texto de diferentes períodos de cada instituição, confirmada pelos casos analisados nas Figuras 7 e 8. Percebe-se que as notas analisadas não estão em total conformidade com a OCPC 07 (CPC, 2014). A característica da utilidade para o usuário em

processos de tomada de decisões fica prejudicada, uma vez que há uma baixa relação com as métricas de risco. Ademais, percebe-se o excesso de informações irrelevantes, dado a semelhança dos textos entre períodos, o que pode causar um desvio do objetivo finalístico de informar o usuário.

Estes resultados vão também 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 Basileia e Inad90, houve variação na política de risco das instituições ao longo do período, porém essa variação não pode ser evidenciada nas notas, 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. De forma a melhor ilustrar essa evidência, foram selecionados alguns trechos com alta similaridade, para realizar uma contra-prova da baixa variabilidade evidenciada. Os resultados estão na Tabela 3.

**Tabela 3:**  
**Trechos de texto repetidos nas Notas Explicativas**

Cod	Data	Trecho de Texto da Nota Explicativa
F14	Jun 16	“Estrutura de Gerenciamento de Risco Operacional, Mercado e Crédito: O Conglomerado Econômico Financeiro da [F14], atendendo às disposições da Resolução do Conselho Monetário Nacional nºs 3.380/06, 3.464/07 e 3.721/09 e suas alterações, possui estrutura de Gerenciamento de risco Operacional, Mercado e Crédito, respectivamente, segregada das áreas de negócio e auditoria interna, bem como estão capacitadas a identificar, avaliar, monitorar, controlar e mitigar seus risco, inclusive aqueles decorrentes de serviços terceirizados. Todas as estruturas de Gerenciamento de risco são direcionadas (...)”
F14	Dez 16	“Estrutura de Gerenciamento de Risco Operacional, Mercado e Crédito: O Conglomerado Econômico Financeiro da [F14], atendendo às disposições da Resolução do Conselho Monetário Nacional nºs 3.380/06, 3.464/07 e 3.721/09 e suas alterações, possui estrutura de Gerenciamento de risco Operacional, Mercado e Crédito, respectivamente, segregada das áreas de negócio e auditoria interna, bem como estão capacitadas a identificar, avaliar, monitorar, controlar e mitigar seus risco, inclusive aqueles decorrentes de serviços terceirizados. Todas as estruturas de Gerenciamento de risco são direcionadas (...)”
F14	Jun 17	“Estrutura de Gerenciamento de Risco Operacional, Mercado e Crédito: O Conglomerado Econômico Financeiro da [F14], atendendo às disposições da Resolução do Conselho Monetário Nacional nºs 3.380/06, 3.464/07 e 3.721/09 e suas alterações, possui estrutura de Gerenciamento de risco Operacional, Mercado e Crédito, respectivamente, segregada das áreas de negócio e auditoria interna, bem como estão capacitadas a identificar, avaliar, monitorar, controlar e mitigar seus risco, inclusive aqueles decorrentes de serviços terceirizados. Todas as estruturas de Gerenciamento de risco são direcionadas (...)”
AF10	Dez 16	“(…) Síntese das ações desenvolvidas no âmbito da gestão de risco: a. Risco de mercado Revisada, anualmente, a Política de Gerenciamento de Risco de Mercado da Instituição tem o objetivo de identificar, avaliar, monitorar e controlar a exposição desses risco, garantindo a existência de mecanismos eficientes para a supervisão e acompanhamento. O risco de mercado não é significativo para [AF10], considerando que a maioria de suas operações têm as taxas “casadas” no ativo e no passivo por decorrerem de operações de repasses, o que mitiga o risco de exposição. b. Risco de crédito O risco de crédito decorre da possibilidade de perdas (...)”
AF10	Jun 17	“(…) Síntese das ações desenvolvidas no âmbito da gestão de risco: a. Risco de mercado Revisada, anualmente, a Política de Gerenciamento de Risco de Mercado da Instituição tem o objetivo de identificar, avaliar, monitorar e controlar a exposição desses risco, garantindo a existência de mecanismos eficientes para a supervisão e acompanhamento. O risco de mercado não é significativo para [AF10], considerando que a maioria de suas operações têm as taxas “casadas” no ativo e no passivo por decorrerem de operações de repasses, o que mitiga o risco de exposição. b. Risco de crédito O risco de crédito decorre da possibilidade de perdas (...)”
AF10	Dez 17	“(…) Síntese das ações desenvolvidas no âmbito da gestão de risco: a. Risco de mercado Revisada, anualmente, a Política de Gerenciamento de Risco de Mercado da Instituição tem o objetivo de identificar, avaliar, monitorar e controlar a exposição desses risco, garantindo a existência de mecanismos eficientes para a supervisão e acompanhamento. O risco de mercado não é significativo para [AF10], considerando que a maioria de suas operações têm as taxas “casadas” no ativo e no passivo por decorrerem de operações de repasses, o que mitiga o risco de exposição. b. Risco de crédito O risco de crédito decorre da possibilidade de perdas (...)”

Fonte: Dados da Pesquisa

A clara repetição exata dos trechos evidencia o baixo valor informacional de uma fonte que deveria auxiliar processos de tomada de decisão. A repetição indica que, além de não haver a correta curadoria do conteúdo sobre gerenciamento de riscos, há um descolamento em relação à prática, uma vez que não estão refletidas as oscilações encontradas nos índices quantitativos. As notas explicativas, que deveriam auxiliar na compreensão das políticas adotadas que levaram aos resultados encontrados nas demonstrações, nos casos analisados, estão muitas vezes com um papel figurativo, o que pode prejudicar os usuários dessa informação. Dessa forma, percebe-se que as notas explicativas não podem ser vistas, nos casos analisados, como um espaço de ampliação da *accountability* e transparência nos relatórios contábeis de forma que estes sejam representações fidedignas dos processos empregados pelas empresas no que tange a gestão de risco (Hendriksen & Van Breda, 2010; Nakagawa et al., 2007).

Dado as evidências de que o trecho das notas explicativas dedicado ao esclarecimento da política de riscos não apenas não está relacionado a índices associados a apetite de risco, como por vezes não apresenta nenhuma nova informação, percebe-se a existência de discricionariedade dos gestores na

divulgação (e omissão) de informações sobre o gerenciamento de risco (Verrecchia, 2001). A repetição detectada com o índice máximo de similaridade traz para discussão uma necessidade latente não apenas de exigência de divulgação, como também de incentivos ou exigências de utilidade dessa informação. Sendo as instituições financeiras, há um aumento dessa preocupação, uma vez que o desconhecimento dos riscos pode ser muito prejudicial aos usuários (Dantas et al., 2005).

## 5 Considerações Finais

O artigo teve como objetivo analisar a efetividade da comunicação do gerenciamento 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.

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 foi identificar se havia alguma relação com as medidas de risco tradicionais. A intuição era que o texto poderia evidenciar a política de gerenciamento de riscos, que impacta por sua vez nos índices quantitativos relacionados.

Para realizar uma comparação do texto com as políticas praticadas de gerenciamento de risco, a similaridade de cosseno foi analisada 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. Adicionalmente, foi realizada uma análise sobre a variabilidade do texto em si, e os resultados demonstraram que a frequência de textos repetidos é alta, com alguns trechos sendo equivalentes, 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 relação 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, pode ser percebido que o texto é reutilizado ou levemente modificado, o que causa uma baixa variabilidade, dentro de uma mesma instituição, entre os períodos, diferentemente 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*. Assim, no caso analisado, o conteúdo das notas explicativas não reflete o seu objetivo de descrever fidedignamente as políticas de gestão de risco dessas organizações, aumentando a transparência entre empresa e usuários da informação contábil-financeira.

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 euclidiana e pré-vetorização dos textos) está no horizonte de estudos futuros. 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. Por fim, a ampliação das fontes de texto (como relatórios de administração ou *releases*) pode auxiliar a melhor investigar a relação entre a informação divulgada e os índices de riscos.

## Referências

Aggarwal, C. C., & Zhai, C. (2012). A survey of text clustering algorithms. In *Mining text data* (pp. 77-128). Springer, Boston, MA. [https://doi.org/10.1007/978-1-4614-3223-4\\_4](https://doi.org/10.1007/978-1-4614-3223-4_4)

BCB – Banco Central do Brasil. (2012). *Resolução nº 4.144*, de 2012. 27 de setembro de 2012. Available at [https://www.bcb.gov.br/pre/normativos/busca/downloadNormativo.asp?arquivo=/Lists/Normativos/Attachments/49098/Res\\_4144\\_v1\\_O.pdf](https://www.bcb.gov.br/pre/normativos/busca/downloadNormativo.asp?arquivo=/Lists/Normativos/Attachments/49098/Res_4144_v1_O.pdf)

BCB – Banco Central do Brasil. (2017). *Resolução nº 4.557*, de 23 de fevereiro de 2017. Available at [https://www.bcb.gov.br/pre/normativos/busca/downloadNormativo.asp?arquivo=/Lists/Normativos/Attachments/50344/Res\\_4557\\_v1\\_O.pdf](https://www.bcb.gov.br/pre/normativos/busca/downloadNormativo.asp?arquivo=/Lists/Normativos/Attachments/50344/Res_4557_v1_O.pdf)

BCB – Banco Central do Brasil. (1966). *Resolução nº 45*, de 30 de dezembro de 1966. Available at [https://www.bcb.gov.br/pre/normativos/busca/downloadNormativo.asp?arquivo=/Lists/Normativos/Attachments/40040/Res\\_0045\\_v8\\_P.pdf](https://www.bcb.gov.br/pre/normativos/busca/downloadNormativo.asp?arquivo=/Lists/Normativos/Attachments/40040/Res_0045_v8_P.pdf)

BCB – Banco Central do Brasil. (1959). *Portaria do Ministério da Fazenda nº 309*, de 30 de novembro de 1959. Available at [https://www.bcb.gov.br/pre/acordos\\_e\\_convenios/Portaria\\_MF\\_309-1959.pdf](https://www.bcb.gov.br/pre/acordos_e_convenios/Portaria_MF_309-1959.pdf)

BCBS - Basel Committee on Banking Supervision. (1997). Core principles for effective banking supervision. *Basileia, Suíça: Bank for International Settlements*. Available at <https://www.bis.org/publ/bcbs30a.pdf>

Bergamini Junior, S. (2005). Controles internos como um instrumento de governança corporativa. *Revista do BNDES*, 12(24), 149-188. Available at: <http://web.bndes.gov.br/bib/jspui/handle/1408/13474>.

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. <https://doi.org/10.1111/1475-679X.12294>

Committee of Sponsoring Organizations of the Treadway Commission - COSO (2004). *Enterprise Risk Management - Integrated Framework*. AICPA: New York, USA. Available at <https://www.coso.org/documents/Framework%20Reference%20Secured.pdf>

CPC – Comitê de Pronunciamentos Contábeis. *CPC 26 (R1)*. (2011) Available at: [http://static.cpc.aatb.com.br/Documentos/312\\_CPC\\_26\\_R1\\_rev%2014.pdf](http://static.cpc.aatb.com.br/Documentos/312_CPC_26_R1_rev%2014.pdf). Accessed July 23<sup>rd</sup>, 2020.

CPC – Comitê de Pronunciamentos Contábeis. *OCPC 07*. (2014) Available at: [http://static.cpc.aatb.com.br/Documentos/488\\_OCPC\\_07\\_Orienta%C3%A7%C3%A3o.pdf](http://static.cpc.aatb.com.br/Documentos/488_OCPC_07_Orienta%C3%A7%C3%A3o.pdf). Accessed March 25<sup>th</sup>, 2021.

Crockford, G. N. (1982). The bibliography and history of risk management: some preliminary observations. *Geneva Papers on Risk and Insurance*, 169-179. Available at <https://www.jstor.org/stable/41950036>

Dantas, J. A., Zendersky, H. C., Santos, S. C. & Niyama, J. K. (2005). A dualidade entre os benefícios do disclosure e a relutância das organizações em aumentar o grau de evidência. *Revista Economia & Gestão*, 5(11), 56-76. Available at <http://periodicos.pucminas.br/index.php/economiaegestao/article/view/40/34>

Del Gaudio, B. L., Megaravalli, A. V., Sampagnaro, G., & Verdoliva, V. (2020). Mandatory disclosure tone and bank risk-taking: Evidence from Europe. *Economics Letters*, 186, 108531. <https://doi.org/10.1016/j.econlet.2019.108531>

Ding, K., Lev, B., Peng, X., Sun, T. & Vasarhelyi, M. (2018). On Creating Accounting Estimates using Machine Learning. *SSRN Electronic Journal*.

Dyer, T.; Lang, M.; & Stice-Lawrence, L. (2017) The Evolution of 10-K Textual Disclosure: Evidence from Latent Dirichlet Allocation. *Journal of Accounting and Economics* 64, 221–45. <https://doi.org/10.1016/j.jacceco.2017.07.002>

Elamer, A. A., Ntim, C. G., Abdou, H. A., Owusu, A., Elmagrhi, M., & Ibrahim, A. E. A. (2021). Are bank risk disclosures informative? Evidence from debt markets. *International Journal of Finance & Economics*, 26(1), 1270-1298. <https://doi.org/10.1002/ijfe.1849>

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. <https://doi.org/10.1002/isaf.1386>

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. <https://doi.org/10.2308/jeta.2010.7.1.1>

Gomes, M., 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. doi: <https://doi.org/10.4270/ruc.2018216>

Hendriksen, E. S., & Van Breda, M. F. (2010) *Teoria da Contabilidade*. 5. ed. São Paulo: Editora Atlas.

Holtz, L., & dos Santos, O. M. (2020). Legibilidade das notas explicativas das empresas brasileiras de capital aberto. *Enfoque: Reflexão Contábil*, 39(1), 57-73. <https://doi.org/10.4025/enfoque.v39i1.45275>

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). <https://doi.org/10.1145/3090354.3090369>

Huang, A.; Leheavy, R., Zang, A. & Zheng, R. (2018). Analyst Information Discovery and Information Interpretation Roles: A Topic Modeling Approach. *Management Science*, 64(6), 2473–2972. <https://doi.org/10.1287/mnsc.2017.2751>

IFRS. IAS1. Available at: <https://www.ifrs.org/issued-standards/list-of-standards/ias-1-presentation-of-financial-statements/>. Accessed July 23<sup>rd</sup>, 2020.

Jizi, M. I., & Dixon, R. (2017). Are risk management disclosures informative or tautological? Evidence from the US banking sector. *Accounting Perspectives*, 16(1), 7-30. <https://doi.org/10.1111/1911-3838.12134>

Kim, J., Kim, M., & Kim, Y. (2019). Bank transparency and the market's perception of bank risk. *Journal of Financial Services Research*, 1-28. <https://doi.org/10.1007/s10693-019-00323-7>

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. [https://doi.org/10.1007/978-3-642-41278-3\\_74](https://doi.org/10.1007/978-3-642-41278-3_74)

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

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. Available at <https://wiki.uib.no/info310/images/4/4c/McAfeeBrynjolfsson2012-BigData-TheManagementRevolution-HBR.pdf>

Manning, C. D., Schütze, H., & Raghavan, P. (2008). *Introduction to information retrieval*. Cambridge university press.

Nahar, S., Azim, M. I., & Hossain, M. M. (2020). Risk disclosure and risk governance characteristics: evidence from a developing economy. *International Journal of Accounting & Information Management*, 28(4). 577-605. <https://doi.org/10.1108/IJAIM-07-2019-0083>

Nakagawa, M., Sordi Relvas, T. R., & Dias Filho, J. (2008). Accountability: a razão de ser da Contabilidade. *Revista de Educação e Pesquisa em Contabilidade (REPeC)*, 1(3), 83-100. <http://dx.doi.org/10.1590/S1519-70772007000200001>

Nielsen, S., & Fuertes-Olivera, P. A. (2013). Development in lexicography: From polyfunctional to monofunctional accounting dictionaries. *Lexikos*, 23, 323-347. <https://doi.org/10.5788/23-1-1218>



Oino, I. (2019). Do disclosure and transparency affect bank's financial performance?. *Corporate Governance: The International Journal of Business in Society*. <https://doi.org/10.1108/CG-12-2018-0378>

Potter, B. N. (2005). Accounting as a social and institutional practice: Perspectives to enrich our understanding of accounting change. *Abacus*, 41(3), 265-289. <https://doi.org/10.1111/j.1467-6281.2005.00182.x>

Raghavan, R. S. (2003). Risk management in banks. *Chartered Accountant-New Delhi*-, 51(8), 841-851. Available at <http://un.uobasrah.edu.iq/lectures/1784.pdf>

Rahrovi, D. 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. <https://doi.org/10.1108/JAAR-01-2018-0016>

Ryan, S. G. (2012). Risk reporting quality: Implications of academic research for financial reporting policy. *Accounting and business research*, 42(3), 295-324. <https://doi.org/10.1080/00014788.2012.681855>

Ribeiro, P. A. N. (2004). Agências de fomento – ambientes institucional, legal e operacional. *Revista da Desenhahia*, Salvador, 1. Available at <https://www.desenhahia.ba.gov.br/publitao/arquivos/arquivos/171dcdd029e944b1b144791bd7b91e32.pdf>

Santos, N. N. F., Flores, E., Guimarães, I. P., Slomski, V. G., & Silva, A. F. (2010). Nível de evidencição obtido na divulgação das notas explicativas de instrumentos financeiros derivativos segundo a deliberação CVM n 550/08. *Revista Contemporânea de Contabilidade*, 7(13), 175-196. <https://doi.org/10.5007/2175-8069.2010v7n13p175>

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. <https://doi.org/10.1016/j.elerap.2018.03.008>

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. <https://doi.org/10.1108/JAOC-08-2010-0038>

Schmitt, E. (2017). Effective risk culture in banks: Responsibilities and boundaries of the risk management. *European Financial Systems 2017 - Proceedings of the 14 th International Scientific Conference*, 239-246. Available at [https://is.muni.cz/do/econ/sborniky/70896034/EFS2017-Proceedings\\_2\\_final.pdf](https://is.muni.cz/do/econ/sborniky/70896034/EFS2017-Proceedings_2_final.pdf)

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. <https://doi.org/10.12928/TELKOMNIKA.v15i4.6414>

Torres, M. C., & Galdi, F. C. (2013). Análise da aderência do disclosure da gestão de riscos pelas instituições financeiras à circular Bacen 3.477. *Revista Contemporânea de Contabilidade*, 10(21), 137-158. <https://doi.org/10.5007/2175-8069.2013v10n21p137>

Verrecchia, R. E. (2001). Essays on disclosure. *Journal of Accounting and Economics*, 32(1-3), 97-180. [https://doi.org/10.1016/S0165-4101\(01\)00025-8](https://doi.org/10.1016/S0165-4101(01)00025-8)

Yamamoto, M. M., & Salotti, B. M. (2006). *Informação contábil: estudos sobre a sua divulgação no mercado de capitais*. São Paulo: Editora Atlas.

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. <https://doi.org/10.1287/isre.2017.0727>

## NOTAS

### AGRADECIMENTOS

Não se aplica.

### CONTRIBUIÇÃO DE AUTORIA

Concepção e elaboração do manuscrito: C. B. Marcolin, E. D. Riva, F. S. Momo.

Coleta de dados: E. D. Riva

Análise de dados: C. B. Marcolin, J. L. Becker.

Discussão dos resultados: C. B. Marcolin, E. D. Riva, F. S. Momo, J. L. Becker.

Revisão e aprovação: C. B. Marcolin, E. D. Riva, F. S. Momo, J. L. Becker

### CONJUNTO DE DADOS DE PESQUISA

O conjunto de dados que dá suporte aos resultados deste estudo não está disponível publicamente.

### FINANCIAMENTO

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - e do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). Em conformidade com a Portaria nº 206, de 4 de setembro de 2018, "o presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001".

### CONSENTIMENTO DE USO DE IMAGEM

Não se aplica.

### APROVAÇÃO DE COMITÊ DE ÉTICA EM PESQUISA

Não se aplica.

### CONFLITO DE INTERESSES

Não se aplica.

### LICENÇA DE USO

Os Direitos Autorais para artigos publicados neste periódico são do autor, com direitos de primeira publicação para a Revista. Em virtude de aparecerem nesta Revista de acesso público, os artigos são de uso gratuito, com atribuições próprias, em aplicações educacionais, de exercício profissional e para gestão pública. A Revista adotou a licença [Creative Commons Atribuição 4.0 Internacional - CC BY NC ND](#). Esta licença permite acessar, baixar (download), copiar, imprimir, compartilhar, reutilizar e distribuir os artigos desde que com a citação da fonte, atribuindo os devidos créditos de autoria. Nesses casos, nenhuma permissão é necessária por parte dos autores ou dos editores. Autores têm autorização para assumir contratos adicionais separadamente, para distribuição não-exclusiva da versão do trabalho publicada nesta revista (ex.: publicar em repositório institucional ou um capítulo de livro).

### PUBLISHER

Universidade Federal de Santa Catarina. Curso de Ciências Contábeis e Programa de Pós-graduação em Contabilidade. Publicação no [Portal de Periódicos UFSC](#). As ideias expressadas neste artigo são de responsabilidade de seus autores, não representando, necessariamente, a opinião dos editores ou da universidade.

### EDITORES

Carlos Eduardo Facin Lavarda e Suliani Rover

### HISTÓRICO

Recebido em: 05/11/2020 – Revisado por pares em: 05/03/2021 – Reformulado em: 04/04/2021 – Recomendado para publicação em: 07/05/2021 – Publicado em: 30/06/2021

\* *Preprint do artigo apresentado no XXIII SEMEAD, 2020.*