



# ORANGE DATA MINING

## UM ESTUDO DAS INFORMAÇÕES DO RELATO INTEGRADO DE UNIVERSIDADES FEDERAIS BRASILEIRAS

*Orange Data Mining*

*A study of information from the integrated reporting of brazilian federal universities*

### **João Eudes de Souza Calado**

Doutorando em Ciências Contábeis pela Universidade de Brasília (UnB), mestre em Contabilidade e Controladoria pela FUCEPE (ES) e especialista em Ensino Superior, Contemporaneidade e Novas Tecnologias pela UNIVASF (PE). Professor voluntário de Orçamento e Finanças Públicas – Mestrado PROFIAP e TAE da Seção de Prestação de Contas da UNIVASF (PE).

**Lattes:** <http://lattes.cnpq.br/7567017277489581>

**Orcid:** <https://orcid.org/0000-0002-5435-1178>

**E-mail:** [joao.calado@univasf.edu.br](mailto:joao.calado@univasf.edu.br)

### **José Matias-Pereira**

Doutor em Ciências Políticas pela Universidade Complutense de Madrid, pós-doutor em Administração pela FEA/USP, professor de Administração e Finanças Públicas e pesquisador sênior do Programa de Pós-Graduação em Ciências Contábeis da Universidade de Brasília (UnB).

**Lattes:** <http://lattes.cnpq.br/5838194984225945>

**Orcid:** <https://orcid.org/0000-0001-6320-026X>

**E-mail:** [matias@unb.br](mailto:matias@unb.br)

### **Abimael de Jesus Barros Costa**

Doutor em Transportes e mestre em Contabilidade pela Universidade de Brasília (UnB), pós-doutor em Contabilidade Pública pela Universidade de Valência (Espanha). Pró-Reitor de Administração e Professor do Programa de Pós-Graduação em Ciências Contábeis da Universidade de Brasília (UnB). Ex-Auditor-Chefe da UnB e do Cebraspe.

**Lattes:** <http://lattes.cnpq.br/6524204350805774>

**Orcid:** <https://orcid.org/0000-0003-0404-701X>

**E-mail:** [acosta@unb.br](mailto:acosta@unb.br)



## RESUMO

O objetivo do estudo é analisar, sob o enfoque da mineração de dados, as informações do Relato Integrado de Gestão (RIG) em algumas Unidades Prestadoras de Contas (UPCs) brasileiras por meio da ferramenta *Orange Data Mining* (ODM). Para tanto, foi realizado um estudo qualitativo, documental e exploratório por meio de práticas de análise textual de dados financeiros e não financeiros do RIG de quinze universidades federais brasileiras. São apresentados dois exemplos de análises, com foco em um único exercício, 2019, o que poderá ser mais explorado em estudos futuros, considerando-se a expectativa de adoção da Estrutura Internacional do Relato Integrado (EIRI) pelas UPCs nos Estados e Municípios. Os resultados da abordagem prática na execução da ODM demonstraram que as características intuitivas da ferramenta podem facilitar as análises por diversos perfis de pesquisadores, do iniciante ao mais experiente, e que se configura como oportunidade para diferentes análises dos referidos dados do RIG. As contribuições da pesquisa poderão integrar a literatura sobre o uso da ODM no setor público, no estímulo a estudantes, servidores públicos, inclusive os do controle, na prática da avaliação das informações do RIG, sobretudo no controle social e na transparência.

**Palavras-chave:** mineração de dados; orange; análises; relato integrado; universidades.

## ABSTRACT

*The purpose of this study is to analyze, from a data mining perspective, the information from the Integrated Report (IR) in some Brazilian Accounting Units (UPCs) using the Orange Data Mining (ODM) tool. To this end, a qualitative, documentary and exploratory study was carried out through textual analysis practices of financial and non-financial data from the IR of fifteen Brazilian federal universities. Two examples of analyses are presented, focusing on a single fiscal year, 2019, which could be further explored in future studies, considering the expectation of adoption of the International Integrated Reporting Framework (IIRF) by the UPCs in the States and Municipalities. The results of the practical approach in the execution of the ODM demonstrated that the intuitive characteristics of the tool can facilitate the analyses by different profiles of researchers, from beginners to the most experienced, and constitutes an opportunity for different analyses of the aforementioned IR data. The contributions of the research may integrate the literature on the use of ODM in the public sector, in encouraging students, public servants, including those in control, in the practice of evaluating IR information, especially in social control and transparency.*

**Keywords:** data mining; orange; analysis; integrated report; universities.



## 1. INTRODUÇÃO

A mineração de dados é a extração de padrões de informações a partir de dados brutos e tem sido utilizada como ferramenta de gestão da informação nas mais diversas áreas do conhecimento, como auditoria, saúde, sistemas bancários (Alexandre; Balsa, 2023; Camilo; Silva, 2009; Jovic; Brkic; Bogunovic, 2014; Otia; Bracci, 2022). Com o avanço da mineração de dados nos últimos anos, muitas ferramentas de análise foram criadas e disponibilizadas, algumas gratuitamente, para auxiliar na mineração de grandes volumes de informações disponíveis na Web ou impressas, por exemplo: Weka, RapidMiner, Orange, Tanagra (Wahbeh *et al.*, 2011).

A *Orange Data Mining* (ODM) é uma dessas ferramentas, com estrutura abrangente para aprendizagem e mineração de dados destinada a usuários e pesquisadores experientes ou iniciantes com interesse em realizar análises por meio de recursos visuais intuitivos (Demsar *et al.*, 2013; Wahberh *et al.*, 2011). A interface de programação da ODM oferece uma visão estruturada das funcionalidades agrupadas em diversas categorias, visualmente representadas por diferentes Widgets para análises baseadas em instâncias, regressão linear e logística, arquivos de textos, entre outras (Demsar; Zupan; Leban; Curk, 2004; Jovic; Brkic; Bogunovic, 2014).

Nos últimos anos, a ODM tem sido utilizada nas investigações em diferentes áreas, com estudos publicados no âmbito de gestão, educação, saúde, comportamento, entre outros (Bagagi, 2022; Hozairi; Anwari; Alim, 2021; Kodati; Vivekanandam, 2018). No entanto, os trabalhos de pesquisa na ciência contábil e no setor público são incipientes, o que representa uma oportunidade para os pesquisadores, tendo em vista o seu modelo intuitivo de combinar métodos conhecidos de análise e visualização em aplicativos (Curk *et al.*, 2005; Otia; Bracci, 2022; Wiguna e Rifai, 2021).

Uma oportunidade de pesquisa relevante no setor público é a análise do *Integrated Reporting*, que é o modelo de Estrutura Internacional de Relatório Integrado (EIRI), criado por um conselho internacional com objetivo de comunicar a geração de valor pelas organizações (IIRC, 2013). Assim, o Tribunal de Contas da União (TCU), por meio da Decisão Normativa nº 170/2018, implementou o uso da EIRI nas prestações de contas das Unidades Prestadoras de Contas (UPCs) brasileiras mediante o Relatório Integrado de Gestão (RIG), que poderá ser estendido também a unidades no âmbito Estadual e Municipal (Brasil, 2018; Castro; Costa, 2022). UPC é uma unidade ou arranjo de unidades da administração pública federal, cujos administradores estão sujeitos ao dever de prestar contas (Brasil, 2020).

Nesse sentido, este estudo se propôs a responder à seguinte questão: é possível analisar as informações financeiras e não financeiras do Relato Integrado de Gestão (RIG) de universidades federais brasileiras por meio da ferramenta *Orange Data Mining*?

Para verificar tal possibilidade, foi estabelecido o objetivo de analisar, sob o enfoque da mineração de dados, as informações do RIG nas UPCs brasileiras utilizando a referida ferramenta.



Para tanto, foi realizada uma revisão da literatura sobre a ODM e uma análise qualitativa, documental e exploratória para exemplificar a prática no uso do software, utilizando Informações Orçamentárias, Financeiras e Contábeis (IOFC), e informações não financeiras (textuais) do RIG de universidades federais brasileiras, disponíveis em dados abertos na página web do TCU.

As investigações sobre a implementação e as contribuições práticas da adoção do RIG, assim como a formação profissional dos contadores, são consideradas questões relevantes da pesquisa contábil para maior transparência e controle social das informações no setor público (Castro; Costa, 2022; Guthrie *et al.*, 2019; TCU, 2018). Portanto, espera-se que os resultados desta investigação possam estimular estudantes, servidores públicos, inclusive os do controle, a usar a ODM e produzir suas pesquisas por meio de uma ferramenta intuitiva e acessível.

Cabe destacar nesta pesquisa que a discussão não será conduzida sob a ótica da Inteligência Artificial (IA) no setor público, considerando que os países ainda estão traçando suas estratégias para a modernização, o controle e a governança dos sistemas, entre outras; apesar dos altos investimentos de EUA e China, e seus interesses potenciais de aplicação da IA nas mais diversas áreas (Fatima; Souza; Dawson, 2020; Wirtz; Weyerer; Geyer, 2018).

## 2. REVISÃO DA LITERATURA

Nesta seção, apresentam-se as discussões sobre os principais aspectos da mineração de dados, as principais ferramentas para sua execução e diferentes tipos de análises. Complementarmente, são apresentados estudos anteriores, até mesmo alguns comparativos do uso da ODM com outras ferramentas.

### 2.1 Principais aspectos da mineração de dados

A mineração de dados é a principal etapa na descoberta de conhecimento em conjuntos de dados brutos e que integra diversos procedimentos de análise para extração de novas informações relevantes aos usuários interessados (Jovic; Brkic; Bogunovic, 2014).

Wahbeh *et al.* (2011) argumentam que uma grande quantidade de dados e informações estão disponíveis para todos e podem ser armazenados em diversos tipos de repositórios de informações, à disposição em redes ou impressos. No entanto, tamanha quantidade de informações necessita de técnicas poderosas para melhor interpretação dos dados que ultrapassam a capacidade humana de compreensão e tomada de decisão (Alexandre; Balsa, 2023; Wahbeh *et al.*, 2011).

Nesse sentido, o grande volume de dados disponibilizados pelas organizações contribuiu para que diversas ferramentas de análises fossem desenvolvidas, algumas com acesso aberto para auxiliar na classificação de informações relevantes e limpas para o processo decisório nas organizações (Jovic; Brkic; Bogunovic, 2014; Otia; Bracci, 2022; Zhang; Zhang; Yang, 2003).



Apesar dos avanços obtidos com a aplicação da mineração de dados, os desafios ainda são muitos, tais como a segurança dos dados, a privacidade dos indivíduos, o aumento das estruturas de armazenamento e as novas tendências, exemplo da Web semântica (Camilo; Silva, 2009). Os desafios da mineração de dados têm produzido cenários complexos, exigindo que sejam desenvolvidas variações nos algoritmos tradicionais, o que também é válido para os debates mais recentes acerca da IA (Otia; Bracci, 2022; Wirtz; Weyerer; Geyer, 2018).

A etapa de mineração inclui a preparação e modelagem de um conjunto de dados que podem ser obtidos de várias fontes, incluindo bancos de dados relacionais tradicionais, *Data Warehouses*, documentos da Web ou arquivos de texto simples (Jovic; Brkic; Bogunovic, 2014; Wiguna; Rifai, 2021).

Revisando a literatura, nota-se que a mineração de dados foi encaminhada para uma popularização se considerado que as ferramentas estão cada vez mais fáceis de serem usadas por usuários não especialistas, o que é fundamental para o crescimento e a consolidação da mineração de dados, até nas organizações do setor público. Um exemplo de ferramenta amigável para a realização de variadas análises sem que seja necessário ser um especialista em mineração de dados é a *Orange Data Mining*.

## 2.2 Ferramentas de mineração de dados

Discute-se mais especificamente, a seguir, a ferramenta *Orange Data Mining*, destacando a sua criação, execução e avanços, inclusive a relevância da análise de dados não financeiros por meio do *Text Mining*.

### 2.2.1 *Orange Data Mining*

A *Orange*, na definição de Wahbeh *et al.* (2011), é uma biblioteca de objetos e rotinas centrais que inclui uma grande variedade de *machine learning* padrão ou não tão padrão assim, de algoritmos de mineração de dados, com rotinas para entrada e manipulação de dados por meio de uma variedade de tarefas, como a impressão de árvores de decisão. Lançado em 2004, mas concebido no final dos anos 1990 por pesquisadores da Eslovênia, a *Orange* é uma estrutura de código aberto que apresenta *scripts* e programação visual e, por conta do seu *design* baseado em componentes integrados com o Python, poderá atrair pesquisadores em razão da velocidade de execução do sistema e facilidade de prototipagem de novos métodos (Demsar *et al.*, 2013).

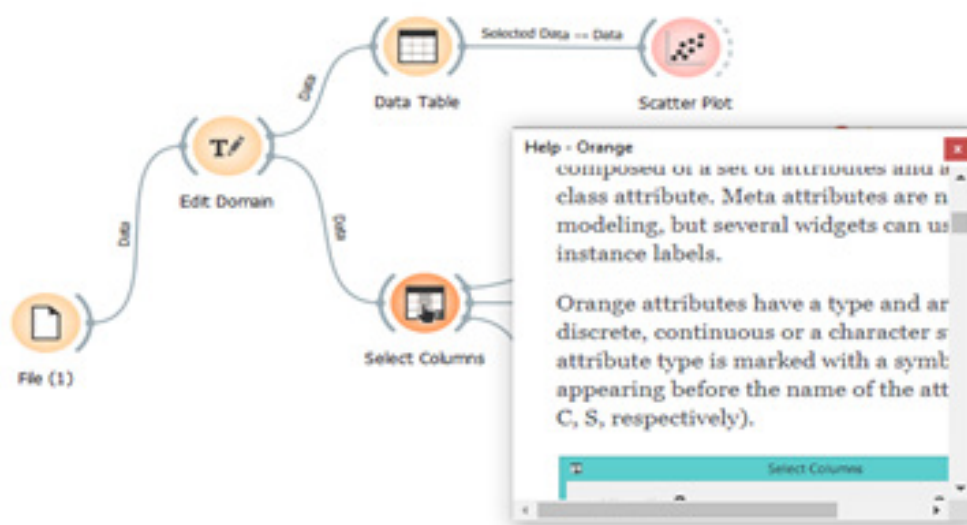
Segundo Curk *et al.* (2005), na *Orange Data Mining*, é possível que usuários que não sejam programadores gerenciem o fluxo de dados, de microarranjos e personalizem suas análises combinando ferramentas comuns de análise de dados de acordo com as suas necessidades. O sistema é publicamente disponível e a interface gráfica para o usuário é fornecida por meio de programação visual e *Widgets* cuidadosamente projetados, que suportam a exploração interativa de dados (Demsar *et al.*, 2013).



O processo de criação de um esquema com *Widgets* e suas conexões é feito visualmente por meio de uma interface gráfica de fácil utilização, na qual os esquemas podem ser salvos e compilados em *scripts* executáveis para reutilização posterior (Curk *et al.*, 2005). Ou seja, a árvore de análise criada poderá ser reutilizada para estudos similares posteriormente, em que o pesquisador precisará apenas substituir os dados a serem analisados.

A Figura 1 ilustra os diferentes *Widgets*, as conexões e o quadro de consulta com informações fornecidas pela ferramenta para orientação no desenvolvimento das análises.

**Figura 1** – Imagem da interface da *Orange Data Mining*



Fonte: *Orange Data Mining*, versão 3.34.

A “*Orange* é executada no Windows, Mac OS X e Linux, e pode ser instalada a partir do repositório *Python Package Index*. O instalador binário para Windows e o pacote de aplicativos para Mac OS X estão disponíveis no site do projeto”<sup>1</sup> (Demsar *et al.*, 2013, p. 2352), dentre as tarefas suportadas pela *Orange* – pré-processamento, amostragem, filtragem, regressão, regressão linear e logística, métodos descritivos, modelos de previsão, entre outras (Demsar *et al.*, 2004; Demsar *et al.*, 2013). Embora não destacado pelos autores, a tarefa de mineração de textos também é suportada pela ferramenta nas versões mais recentes (Wiguna; Rifai, 2021). O que talvez explique a ausência de menção a isso seja a funcionalidade, um dos avanços do sistema nas últimas versões.

### 2.2.2 Text Mining

Nas organizações, não somente os dados financeiros poderão ser observados, mas também dados dos relatórios anuais, postagens nas mídias sociais e os que podem ser relevantes para auxiliar na interpretação das informações financeiras. Por exemplo, informações negativas

<sup>1</sup> Disponível em: <https://orangedatamining.com/>. Acesso em: 4 jun. 2023.



ou positivas sobre a organização podem impactar os resultados dos negócios, e não estão atreladas apenas aos dados financeiros (Marwa; Kristanto, 2022).

Na contabilidade, muitas das informações da gestão são apresentadas por meio de dados não numéricos, como nas notas explicativas das demonstrações contábeis e nos relatórios da administração. No caso do RIG, a adoção da EIRI, exigida pelo TCU para todas as UPCs brasileiras, contempla: a) abordagem estratégica; b) conectividade da informação; c) relações com partes interessadas; d) materialidade; e) concisão; f) confiabilidade e completude; g) coerência e comparabilidade; e h) clareza (IIRC, 2013; Brasil, 2018).

Portanto, a informação textual tem grande potencial de gerar valor organizacional para as instituições públicas. Marwa e Kristanto (2022) argumentam que as informações não financeiras podem ser aquelas consultadas nos relatórios anuais ou até em postagens nas redes sociais e que são relevantes para interpretação das informações financeiras.

Considerando o avanço da mineração de dados, que contribuiu para o surgimento de diversas ferramentas de análises, faz-se necessário aprofundar as discussões para além de suas principais características, apontando vantagens, desvantagens e possíveis limitações de algumas delas.

## 2.3 Estudos anteriores

A seção de estudos anteriores apresenta outras ferramentas de análises, apontando estudos que comparam essas ferramentas com a ODM. Expõem-se também pesquisas que utilizaram a *Orange Data Mining* e as possibilidades de seu uso em estudos direcionados para o setor público.

### 2.3.1 Outras ferramentas

No estudo de Wahbeh *et al.* (2011), realizou-se uma análise comparativa entre algumas das ferramentas e pacotes de softwares de mineração de dados disponíveis gratuitamente. Os autores identificaram que o desempenho das ferramentas para a tarefa de classificação é afetado pelo tipo de conjunto de dados usado e pela forma como os algoritmos de classificação são implementados.

Audibert *et al.* (2020) realizaram experimentos utilizando dados proprietários da *Orange* para validar requisitos de escalabilidade, estabilidade, robustez, velocidade, entre outros. Segundo os autores, a supervisão automática dos sistemas de TI é um desafio para a *Orange*, pois o tamanho e a complexidade das operações na medição de comportamentos normais e anormais ao longo do tempo tornaram os métodos tradicionais lentos e propensos a erros. Contudo, os autores conseguiram detectar e resolver as anomalias do sistema (Audibert *et al.*, 2020).



Em 2022, os mesmos autores publicaram um estudo no qual avaliam comparativamente o desempenho de dezesseis algoritmos em cinco conjuntos de dados públicos por meio de métodos convencionais de machine learning e redes neurais profundas; sem especificar as ferramentas analisadas, concluíram não ser possível apontar uma superioridade aparente entre os dois métodos na detecção de anomalias (Audibert *et al.*, 2022).

Em análise comparativa entre as ferramentas Weka, *Orange*, Tanagra e KNIME, para avaliar as tarefas de maior desempenho na classificação e aplicabilidade dos dados de apoio a tomada de decisão, Wahbeh *et al.* (2011) indicaram nos resultados que a *Orange* ficou na segunda posição nas duas tarefas, atrás da ferramenta Weka (Wahbeh *et al.*, 2011).

Jovic, Brkic e Bogunovic (2014) analisaram os prós e contras no uso das ferramentas RapidMiner, R, Weka, KNIME, *Orange* e *Scikit-learn*. Os autores destacam a interface da *Orange* como polida e visualmente atraente, oferecendo aos usuários uma experiência agradável, mas ponderam que uma aparente desvantagem dessa ferramenta em relação às demais seja a limitação no número de *Widgets*. Concluem afirmando que não existe uma única ferramenta melhor, pois todas têm pontos fortes e fracos (Jovic; Brkic; Bogunovic, 2014).

Não é difícil encontrar na literatura estudos que discutam sobre as ferramentas de análises e mineração de dados, notando-se que não se esgotam as possibilidades de uso dessas ferramentas no desenvolvimento de estudos exploratórios, descritivos, com análises de dados qualitativos, quantitativos, entre outras. Outro ponto relevante para os interessados é o fato de serem ferramentas acessíveis, intuitivas e, muitas delas, sem custos para estudantes e pesquisadores.

### 2.3.2 Estudos realizados com a ODM e possibilidades de pesquisas no setor público

Ao consultar a literatura dos trabalhos mais citados nos últimos cinco anos com o descritor *Orange Data Mining* no título e nas palavras-chave, com o auxílio do software de buscas *Publish or Perish*, versão 8, na base de dados do *Google Scholar*, identificaram-se estudos nas áreas de saúde, gestão, educação, fruticultura, comportamental (sentimentos), redes sociais, entre outras (Hozairi; Anwari; Alim, 2021; Kodati; Vivekanandam, 2018; Peker; Özkaraca; Sasar, 2018; Vaishnav; Rao, 2020; Wiguna; Rifai, 2021).

Para Kodati e Vivekanandam (2018), as técnicas de mineração de dados ajudam a encontrar informações ocultas em grupo de dados sobre doenças, contribuindo em suas análises na classificação e predição de doenças cardíacas. No mesmo sentido, Peker, Özkaraca e Sasar (2018) desenvolveram um método eficaz de modelo de previsão usando um grande volume de dados e recursos para seleção e classificação na determinação de parâmetros efetivos no diagnóstico do diabetes.

Na educação, o uso da ferramenta de mineração de dados contribui para a identificação de alunos propensos a atrasar sua formação, comprometendo o credenciamento para ingresso em programas de pós-graduação; assim, os indicadores podem ser usados pela coordenação dos cursos para atacar o problema no seu início (Hozairi; Anwari; Alim, 2021).





A *Orange* também é utilizada para análise de sentimentos em textos, opiniões e comentários nas redes sociais. Wiguna e Rifai (2021), por exemplo, analisaram os dados de textos publicados no Twitter (atual X) sobre a Lei geral *Omnibus* de criação de empregos em 2020 na Indonésia. Os autores rastrearam dados de *tweets* com *hashtags* específicas e analisaram sentimentos e emoções apontados nas mensagens, destacando-se a alegria e a surpresa como os principais (Wiguna; Rifai, 2021).

Vaishnav e Rao (2020) aplicaram algoritmos de *machine learning* num conjunto de imagens de frutas para determinar que algoritmo tem a maior pontuação de precisão e exatidão de classificação. Assim, a tomada de decisão é baseada nas imagens usadas para treinar o algoritmo na identificação de características específicas de determinadas frutas.

No Brasil, Bagagi (2022) utilizou a mineração textual de codificação aberta, axial e seletiva da *Orange* para emergir categorias e subcategorias de dados coletados em entrevistas semiestruturadas com gestores de altos postos de decisão em hospitais. O estudo investigou a dinâmica de inovação em diferentes modelos de gestão hospitalar. Calado *et al.* (no prelo) utilizaram a análise textual da *Orange* para identificar a frequência de palavras e similaridades dos artigos sobre tópicos contemporâneos em contabilidade do setor público, identificando, por exemplo, numa mesma pesquisa, contribuições sobre a implementação dos sistemas de custos, modelagem e inovação aberta – todas temáticas relevantes da pesquisa contábil no setor público.

Como se observa nas características das pesquisas com o uso da *Orange*, é possível inferir diversas possibilidades de investigação no setor público, como: análise textual dos relatórios financeiros, contábeis, de gestão e governança das UPCs, análise de sentimentos nas redes sociais sobre as ações da gestão, imagens do próprio RIG. Nesse sentido, a adoção do modelo internacional do RI também se apresenta como um campo oportuno de estudo para avaliação das informações financeiras ou textuais por meio da *Orange*.

Oportunidades para avançar no conhecimento da ferramenta e da análise textual da *Orange* podem ser vistas nas videoaulas do professor Abdinardo Oliveira<sup>2</sup>, da UTFPR do Mestrado Profissional em Administração Pública, durante as aulas remotas no período da pandemia da covid-19 e disponíveis na Web.

### 3. METODOLOGIA

Considerando o objetivo do estudo, a pesquisa pode ser caracterizada como pesquisa qualitativa, documental e exploratória. O percurso metodológico se deu a partir da revisão da literatura sobre a *Orange Data Mining* e a exploração da ferramenta para testes e análises de informações financeiras e não financeiras do setor público constantes do RIG de algumas universidades brasileiras após a exigência normativa pelo TCU. Adicionalmente foram utilizadas técnicas de coleta de dados a partir dos recortes e tabulação de textos e dados financeiros constantes nos RIG referentes ao exercício de 2019, considerando se tratar do único exercício

---

2 Disponível em: <https://www.youtube.com/watch?v=mKug7ZnWhqs>. Acesso em: 20 abr. 2023.



em que as informações estão disponibilizadas na página web do TCU. Os RIG dos exercícios seguintes deverão estar disponíveis nas páginas web de cada uma das UPCs (universidades).

### 3.1 Perfil da amostra

Para ilustrar o uso da ODM, as análises serão realizadas nos dados do RIG de quinze universidades federais, escolhidas aleatoriamente, três de cada uma das cinco regiões brasileiras. O motivo para a escolha do RI foi o fato de os estudos de caso sobre a adoção do modelo internacional do Relato Integrado e a formação dos profissionais serem incentivados nas pesquisas contábeis (Guthrie *et al.*, 2019). Complementarmente, o RIG das UPCs deve apresentar declaração do contador, opinião dos auditores externos, notas explicativas e informações sobre o desempenho orçamentário, financeiro e patrimonial da gestão (Brasil, 2018). Portanto, representa uma oportunidade para aplicação prática da ferramenta, avaliação da transparência e maior controle das informações das UPCs no país.

### 3.2 Procedimentos de análises

As análises a partir da abordagem prática utilizando a ferramenta ODM foram realizadas em duas etapas. A primeira textual e a segunda financeira, ambas exploratórias, apresentadas nas análises e discussões. A Tabela 1 apresenta informações prévias, resumidas e tabuladas dos dados abertos disponíveis na página do TCU.

**Tabela 1** – Informações prévias dos RIG das UPCs

Instituição	Região	Total de páginas RIG	Total de páginas IOFC	Total de páginas Declaração do Contador	Ilustrações nas IOFC
FUBSB	CO	138	19	1	SIM
FUABC	SE	98	10*	2	SIM
UFAMZ	N	155	51	2	SIM
FUVSF	NE	106	22	2	NÃO
UFBA	NE	137	11	-	SIM
UFGO	CO	205	70	7	SIM
UFMG	SE	125	9	-	SIM
UFSC	S	164	9	-	SIM
UFAC	N	108	6	1	SIM
UFCE	NE	270	8	1	NÃO



Instituição	Região	Total de páginas RIG	Total de páginas IOFC	Total de páginas Declaração do Contador	Ilustrações nas IOFC
UFMT	CO	102	39	3	SIM
UFPA	N	97	5	1	SIM
UFPR	S	371	79	2	SIM**
UFRJ	SE	440	67	3	NÃO
UFRS	S	216	8	-	NÃO

Fonte: os autores (2023).

Legenda: \* as IOFC foram distribuídas ao longo do RIG e indicados *links* de acesso às Demonstrações.

\*\* Apenas o organograma da instituição.

- Não localizadas as declarações dos contadores.

As siglas FUBSB, FUABC, UFAMZ, FUVSF e UFCE apresentadas na Tabela 1, diferem das siglas e abreviações correspondentes à Universidade de Brasília (UnB), Universidade Federal do ABC (UFABC), Universidade Federal do Amazonas (UFAM), Universidade Federal do Vale do São Francisco (UNIVASF) e Universidade Federal do Ceará (UFC), respectivamente. Como pode ser observado na tabela, há divergências entre as instituições no que tange à quantidade de páginas do RIG, das IOFC e da Declaração dos Contadores. Nesse sentido, parece indicar a necessidade de adequações ao Modelo Internacional de RI exigido pelo TCU por parte de algumas instituições. Trata-se do primeiro exercício após a Decisão Normativa e único ano disponibilizado na página do TCU até então. Assim, espera-se que na divulgação dos RIG dos anos seguintes pelas UPC, em suas páginas próprias na Internet, haja maior aderência ao modelo exigido.

Antes da apresentação das análises, faz-se necessário esclarecer que o foco estará na ferramenta e não na análise estatística dos dados. Portanto, para que os leitores não se frustrem, informa-se que não serão discutidos índices, como *P-value*, significância, ou outros. Essas avaliações serão feitas quando das análises aplicadas efetivamente pelos pesquisadores e leitores interessados na ferramenta a partir deste estudo.

Outra questão importante: as análises práticas são duas das variadas possibilidades de uso da ferramenta, ou seja, são análises básicas para os que ainda não a conhecem, principalmente. Portanto, é apenas uma introdução, “a ponta do *iceberg*” – base para os interessados, que terão a oportunidade de aprofundar as análises e explorar a ferramenta.

## 4. ANÁLISES E DISCUSSÕES

A seção de análises e discussões foi dividida em duas etapas. A primeira apresenta a análise textual, não financeira. Na segunda, são apresentadas as análises das informações financeiras dos dados coletados nos RIGs.



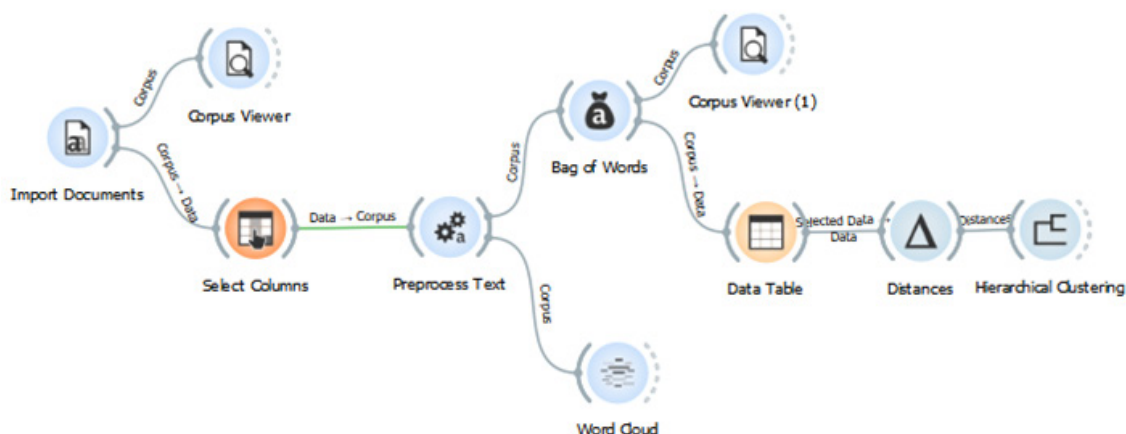
## 4.1 ANÁLISE TEXTUAL

Para demonstrar, na prática, o uso da ODM nesta primeira etapa, mais especificamente o *Text Mining*, foi utilizado o recorte das IOFC e da Declaração dos Contadores para análise da similaridade, frequência de palavras, hierarquização, entre outros aspectos. Bagagi (2022) classifica essas análises na *Orange* como estágios de limpeza, identificação e formatação numa estrutura de dados organizados – tokenização, visualização geral em *Word Cloud* (nuvem de palavras), normalização, agrupamento hierárquico de textos, palavras relevantes e agrupamento hierárquico dos tokens de palavras relevantes.

A característica principal da *Orange* observada na literatura é a sua intuitividade. Demsar *et al.* (2013) argumentam que cada *Widgets* apresenta informações sobre o preenchimento necessário para a análise pretendida. Sendo assim, o pesquisador é orientado sobre as configurações necessárias em cada etapa.

Os *Widgets* utilizados para as análises das IOFC e Declaração dos contadores constam na Figura 2.

**Figura 2** – Widgets para análise textual das IOFC e Declaração dos contadores



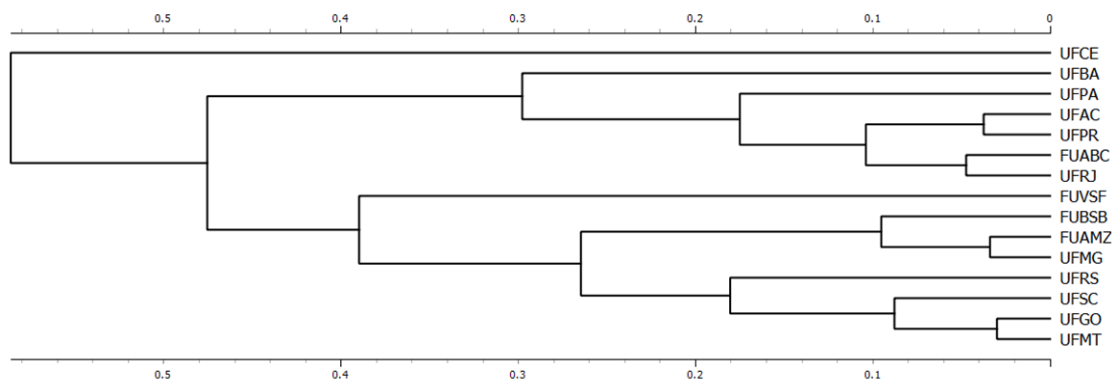
Fonte: os autores (2023) a partir da *Orange*, versão 3.34.0.

A análise textual da ODM é feita a partir do agrupamento de palavras, números e precede o cálculo da distância entre as linhas representadas pelos tokens, utilizando o *widget Distances* e aplicando os parâmetros: *Distances between – Rows* e *Distance Metric – Cosseno*, por funcionarem melhor com o *corpus* de documentos textuais (Bagagi, 2022).

Os resultados das análises de hierarquia e relevância das palavras nos textos das IOFC constam na Figura 3, a seguir.



**Figura 3** – Frequência e relevância das palavras nos textos das IOFC



Fonte: os autores (2023) a partir da *Orange*, 3.34.0.

A Figura 3 apresenta a UFCE, sigla que representa a Universidade Federal do Ceará (UFC), no topo da hierarquia de frequência e relevância de palavras ligadas a dois grandes grupos. O primeiro agrupando seis instituições de quatro regiões brasileiras, e o segundo agrupando outras oito instituições distribuídas nas cinco regiões do país. Nessa figura, é possível inferir que, apesar de existirem dois grupos, há pequenos subgrupos que se alinham, ou seja, não há aparentemente uma similaridade de modelo de informações, o que comprometeria, por exemplo, a comparabilidade – característica qualitativa de melhoria da informação contábil; e que também representa um dos objetivos de adoção da EIRI pelo TCU no Brasil (Araújo; Callado, 2020; Brasil, 2018). Outra observação relevante é que, apesar de se ver um RIG extenso (270 páginas) nos textos da UFCE, apenas oito páginas para as IOFC apresentam similaridade com as demais.

A análise Word Cloud, disponível no *Text Mining* da *Orange*, também evidencia relevantes indicadores. A Figura 4, a seguir, apresenta a nuvem de palavras das IOFC contendo as 50 palavras ou números mais frequentes, configurados no *Widgets “Preprocess Text”*. Nesta análise, também é possível inserir um arquivo de texto “*Stop Word*” para excluir da nuvem elementos não relacionados com a análise, por exemplo, letras, números, sílabas que aparecem dispersos, sem significação na nuvem.

**Figura 4** – Nuvem de palavras das IOFC das universidades analisadas



Fonte: os autores (2023) a partir da *Orange*, versão 3.34.0.

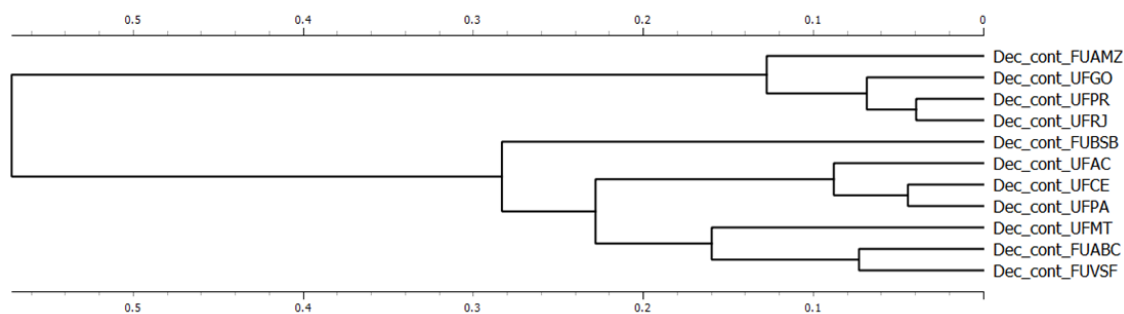


Como observado na Figura 4, as IOFC das universidades contemplam basicamente os números refletidos em várias tabelas, quadros, demonstrativos contábeis e ilustrações. Nesse sentido, reflexões a serem feitas pelos elaboradores das IOFC poderiam ser: Qual é a relevância do excesso de dados numéricos? Há qualidade no texto na apresentação desses dados? Como a sociedade ou a comunidade acadêmica, por exemplo, vai se interessar em ler tantos dados, números, textos?

Adicionalmente, não em relação à ferramenta em si, mas aos dados reportados nos RIGs, Castro e Costa (2022) argumentam, ao analisar o alinhamento das pesquisas nacionais às internacionais, que há prevalência de citações do modelo internacional do RI mencionando a necessidade de estudos que apresentem as dificuldades na consolidação do RI tanto na teoria quanto na prática.

Ainda nas análises textuais, o recorte da Declaração dos contadores não contempla informações da UFBA, UFMG, UFSC e UFRS, pois não foram localizadas no RIG de 2019. A Figura 5 apresenta os dados reportados da hierarquização, relevância, frequência e similaridade dos arquivos de onze universidades.

**Figura 5** – Frequência e relevância das palavras nos textos das IOFC



Fonte: os autores (2023) a partir da *Orange*, versão 3.34.0.

Na Figura 5, observam-se dois grandes grupos, o primeiro contempla três subgrupos de instituições de quatro regiões, nenhuma do NE, em que duas dessas instituições apresentaram declarações com o maior volume de informações (ver Tabela 1). O segundo grande grupo é composto por seis subgrupos de instituições alinhadas entre si.

Na sequência, a Figura 6 apresenta a nuvem com as 20 palavras mais presentes nos arquivos com as Declarações dos contadores.



**Figura 6** – Nuvem de palavras das declarações dos contadores



Fonte: os autores (2023) a partir da Orange, 3.34.0.

Como se nota, a palavra mais frequente é o ano de exercício do RIG. As demais palavras se referem aos relatórios, demonstrativos contábeis, sistemas, entre outros documentos, aparentemente confirmando o caráter institucional normativo. A reflexão que se coloca na avaliação dos dados reportados na Figura 6 da nuvem é: qual a contribuição prática para o controle dessas informações no RI? Talvez, nesse sentido, Dumay *et al.* (2016) sugiram estudos sobre os determinantes da qualidade das informações, bem como sobre contribuições teóricas e práticas dos relatórios. Ao que parece, essas contribuições não estão refletidas na Figura 6.

#### 4.2 ANÁLISES DE INFORMAÇÕES FINANCEIRAS

Para a análise das informações financeiras do RIG das quinze universidades, foi criada uma planilha do Office com dados sobre número de servidores docentes e técnicos administrativos em educação (TAE), estudantes matriculados na graduação e pós-graduação, despesas com pessoal, custeio, investimento, receitas próprias e de emenda parlamentar, quando disponíveis nas informações sobre o orçamento das instituições.

A Figura 7 apresenta o conjunto das informações utilizadas para avaliação e plotagem gráfica da relação entre as variáveis categóricas, numéricas e textuais estabelecidas. Nessa análise, foram excluídas as informações da UFGO e da UFRJ, já que os dados sobre força de trabalho e estudantes matriculados estavam incompletos.

**Figura 7** – Conjunto de dados orçamentários, financeiros e não financeiros

	Universidade	Região	Docentes	TAE's	Estudantes Graduação	Estudantes Pós-graduação	Tesouro	Pessoal	ODC (Custeio)	Investimentos
1	FUABC	SE	796	758	12834.0	1481	301832.0	238730	52541	8911
2	FUAMZ	N	1670	1684	21993.0	4033	864945.0	682214	156712	26019
3	FUBSB	CO	2818	3171	39610.0	8435	1694992.0	1536773	146407	5183
4	FUVSF	NE	581	365	9639.0	1284	195234.0	147129	46432	1673
5	UFAC	N	717	719	8892.0	1276	373926.0	319189	46270	8467
6	UFCE	NE	2345	3402	29900.0	6700	1672517.0	1463696	200338	8484
7	UFMT	CO	2513	1259	19940.0	2360	1037352.0	881596	148561	7195
8	UFPA	N	2999	2539	38786.0	10024	1399489.0	1216715	194175	11287
9	UFPR	S	2536	3450	27408.0	5358	1892582.0	1605972	270911	15699
10	UFBA	NE	2551	3076	39646.0	7625	1822387.0	1365812	186339	13532
11	UFMG	SE	3334	4329	34487.0	11075	2475980.0	2043380	405690	26910
12	UFRS	S	2976	2566	31456.0	16562	1670925.0	1670925	262418	5506
13	UFSC	S	4538	1112	31075.0	8217	1871695.0	1632809	180266	14919

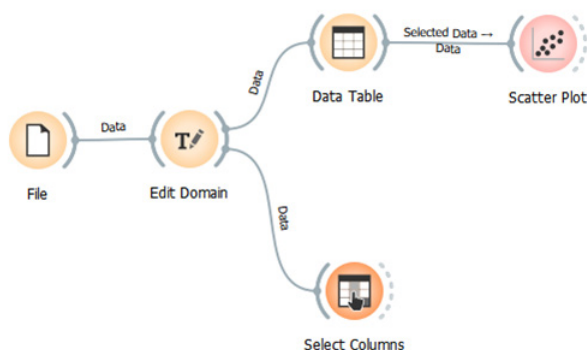
Fonte: os autores (2023) a partir da Orange, 3.34.0.



Demsar *et al.* (2004) e Demsar *et al.* (2013) argumentam que o design baseado em componentes, no nível de programação tanto processual quanto visual, para projetar novos métodos de mineração de dados e proporcionar um ambiente agradável ao usuário é o atributo mais significativo que a *Orange* pode oferecer aos interessados.

A Figura 8, na sequência, apresenta os *Widgets* para plotagem gráfica das análises da relação entre os dados orçamentários, financeiros e não financeiros constantes do RIG das universidades analisadas.

**Figura 8** – *Widgets* para análise da relação de dados financeiros e não financeiros

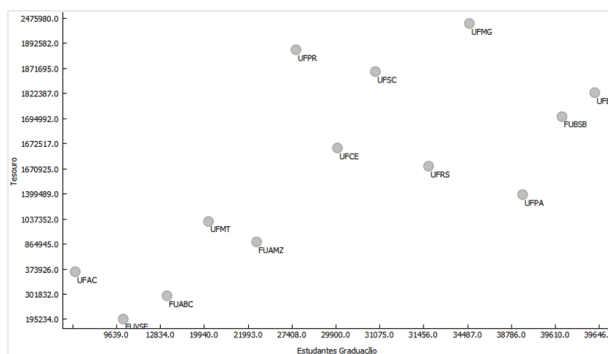


Fonte: os autores (2023) a partir da *Orange*, 3.34.0.

As referências aos dados financeiros correspondem aos valores dos orçamentos, das despesas com pessoal, do custeio, entre outros. Já os dados não financeiros são representados pelo número de estudantes matriculados na graduação e na pós-graduação, pelo número de docentes, de TAE, entre outros. Esses foram os dados escolhidos para as análises exploratórias utilizadas neste estudo, podendo os interessados não só explorar outras variáveis – como o valor de recursos próprios, que representa uma variável significativa e discrepante entre as universidades –, mas também obter, em informações não financeiras, número de cursos, de projetos, entre outros constantes no RIG.

A Figura 9 apresenta a plotagem da análise da relação entre o valor da dotação orçamentária das universidades e o número de estudantes matriculados.

**Figura 9** – Relação “valor do orçamento” x “número de estudantes matriculados na graduação”



Fonte: os autores (2023) a partir da *Orange*, 3.34.0.



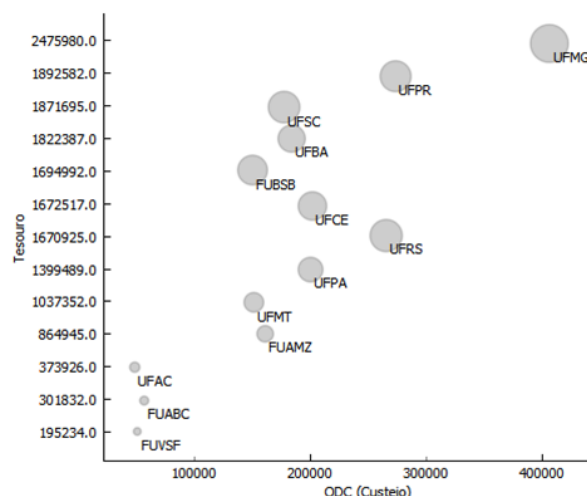


O gráfico da Figura 9 apresenta uma distribuição não normal, visto que instituições com o quantitativo inferior no “número de alunos matriculados na graduação” apresentam valor do orçamento superior, como UFMG, UFPR e UFSC em relação à UFBA, FUBSB e UFPA. Neste sentido, pode haver inconsistências nos dados apresentados nos RIGs ou então nenhuma relação entre o “valor do orçamento” e o “número de estudantes matriculados”. Castro e Costa (2022) alertam que a prestação de contas por meio do RI é como o gestor público deve demonstrar o alcance dos resultados nas instituições, portanto, ele é o responsável pelas informações e pelo cumprimento da legislação no relatório.

Em razão da perspectiva de que os Tribunais de Contas estaduais e municipais irão se engajar na institucionalização do RI nos próximos anos, as pesquisas sobre governança e *accountability* em governos locais são incentivadas (Castro; Costa, 2022). Assim, o uso da ODM para análises das informações dos RIG também é estimulado a partir deste estudo.

A Figura 10, a seguir, apresenta a relação entre “valor do orçamento”, “despesas de custeio” e “despesas com pessoal” nas universidades.

**Figura 10** – Relação valor do orçamento x custeio x despesas com pessoal



Fonte: os autores (2023) a partir da Orange, 3.34.0.

A circunferência dos círculos da Figura 10 representa os valores das “despesas com pessoal” nas universidades. A UFMG apresenta o maior orçamento, recursos de custeio e despesas com pessoal. Porém, nas demais instituições, os valores não seguem uma distribuição normal; por exemplo, a UFRS apresenta o terceiro maior valor de recursos de “custeio” e o segundo maior valor de “despesas com pessoal”, mas aparece na sétima posição em “valor de orçamento”.

As análises deste estudo foram limitadas à análise textual e de relação entre variáveis financeiras e não financeiras das informações dos RIGs de um grupo de universidades federais brasileiras com o uso da ODM e referentes ao exercício de 2019. Contudo, a ferramenta oferece outras variadas tarefas, do pré-processamento de dados à modelagem e avaliação, como o gerenciamento de dados, amostragem, filtragem, indução de modelos de classificação e



regressão, incluindo árvores, regressão linear e logística, entre outras possibilidades (Demisar *et al.*, 2013). *et al.*, 2013).

Portanto, este estudo é apenas uma introdução às possibilidades de análises com a ODM, sendo direcionado a estudantes de contabilidade, pesquisadores da área pública, do controle, entre outras, em que se utilizou como exemplo uma abordagem prática e analítica das informações do RIG. Como se pôde perceber, é possível avançar mais, tanto nas análises aqui reportadas quanto no uso da ferramenta, considerando a sua característica intuitiva. Neste sentido, o que foi apresentado poderá ser potencializado, considerando-se que a adoção do RIG deverá ser estendida às Unidades Prestadoras de Contas nos Estados e Municípios.

## 5. CONSIDERAÇÕES FINAIS

A mineração de dados tem sido cada vez mais utilizada nas pesquisas nas mais diversas áreas do conhecimento. E, acompanhando esse avanço, as ferramentas de análises de mineração de dados e estatísticas têm sido desenvolvidas e disponibilizadas para diferentes perfis de pesquisadores, do iniciante ao mais experiente.

Este estudo objetivou analisar, sob o enfoque da mineração de dados, as informações do RIG nas UPCs brasileiras com a ferramenta *Orange Data Mining* (ODM).

A primeira e principal inferência da abordagem prática foi de que a ODM se apresenta como oportunidade em diferentes tipos de investigações das informações financeiras e não financeiras do RIG, exemplos das análises textual, comparativa (*benchmarking*), qualitativa, quantitativa, entre outras. A segunda conclusão foi de que, dadas as características intuitivas e a possibilidade de acesso sem custos, a ferramenta permite aos interessados, iniciantes ou experientes avançarem nas investigações dos RIGs do setor público.

Nesse sentido, o exemplo prático na análise das informações do RIG de universidades federais brasileiras contribui para a literatura sobre o uso da mineração de dados e da ODM nas investigações do setor público, o aprimoramento da administração pública, o controle e a transparência das informações do RIG das UPCs brasileiras. Adicionalmente e intrinsecamente, o estudo apresenta contributos para a sociedade e para ciência.

As análises, neste estudo, circunscreveram-se a duas: introdutórias e básicas – textual; financeira e não financeira, relativas ao exercício de 2019, único disponível na página do TCU na internet. A pesquisa poderá ser expandida, com relação aos exercícios seguintes, por meio de consultas às páginas de cada universidade.

Foram, portanto, observadas na literatura, diversas possibilidades de investigação com a utilização da ODM: análise de regressão linear logística, análises de sentimentos, imagens, modelos de previsão, entre outras.

Assim, sugerem-se, para estudos futuros, investigações comparativas (*benchmarking*) nas UPCs brasileiras; ampliação das análises (textos, imagens, sentimentos etc.), ampliação de



amostras e exercícios; e análises das informações nos Estados e Municípios com a utilização da ODM, considerando-se as expectativas de expansão da adoção da EIRI para os relatórios de prestação de contas além do âmbito federal.

## REFERÊNCIAS

ALEXANDRE, Cláudio Reginaldo; Balsa, João. Incorporating machine learning and a risk-based strategy in an anti-money laundering multiagent system. **Expert Systems With Applications**, [s. l.], v. 217, 119500, maio 2023. DOI: <https://doi.org/10.1016/j.eswa.2023.119500>. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0957417423000015?via%3Dihub>. Acesso em: 12 abr. 2024.

ARAÚJO, Ricardo Henrique Miranda de; CALLADO, Antônio André Cunha. Características qualitativas da informação contábil: a percepção de profissionais de controle interno do setor público. **Management Control Review**, Universidade Federal Rural de Pernambuco, v. 5, n. 1, p. 47-65, jan./jun. 2020. DOI: <https://doi.org/10.51720/mcr.v5i1.3696>. Disponível em: <https://www.journals.ufrpe.br/index.php/managementcontrolreview/article/view/3696>. Acesso em: 12 jan. 2024.

AUDIBERT, Julien; MICHIARDI, Pietro; GUYARD, Frédéric; MARTI, Sébastien; ZULUAGA, Maria A. USAD: UnSupervised anomaly detection on multivariate time series. In: **Proceedings of the 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '20)**, ago. 2020, Virtual. p. 3395-3404. DOI: <https://doi.org/10.1145/3394486.3403392>. Disponível em: <https://dl.acm.org/doi/10.1145/3394486.3403392>. Acesso em: 15 abr. 2023.

AUDIBERT, Julien; MICHIARDI, Pietro; GUYARD, Frédéric; MARTI, Sébastien; ZULUAGA, Maria A. Do deep neural networks contribute to multivariate time series anomaly detection? **Pattern Recognition**, [s. l.], v. 132, 108945, dez. 2022. DOI: <https://doi.org/10.1016/j.patcog.2022.108945>. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0031320322004253?via%3Dihub>. Acesso em: 14 abr. 2024.

BAGAGI, Leone C. **Dinâmica de inovação em diferentes modelos de gestão hospitalar**. 2022. Tese (Doutorado em Administração) – Universidade Federal da Bahia, Escola de Administração, Salvador, 2022. Disponível em: <https://repositorio.ufba.br/handle/ri/36038>. Acesso em: 15 abr. 2023.

BRASIL. Tribunal de Contas da União. **Decisão Normativa nº 170, de 19 de setembro de 2018**. Dispõe acerca das unidades cujos dirigentes máximos devem prestar contas de suas gestões ocorridas no exercício de 2018, especificando a forma, os conteúdos e os prazos de apresentação, nos termos do art. 3º da Instrução Normativa TCU 63, de 1º de setembro de 2010. Brasília: TCU, 2018. Disponível em: <https://portal.tcu.gov.br/contas/contas-e-relatorios-de-gestao/contas-do-exercicio-de-2018.htm>. Acesso em: 16 abr. 2023.



BRASIL. Tribunal de Contas da União. **Instrução Normativa nº 84, de 22 de abril de 2020**. Estabelece normas para a tomada e prestação de contas dos administradores e responsáveis da administração pública federal, para fins de julgamento pelo Tribunal de Contas da União, nos termos do art. 7º da Lei 8.443, de 1992, e revoga as Instruções Normativas TCU 63 e 72, de 1º de setembro de 2010 e de 15 de maio de 2013, respectivamente. Brasília: TCU, 2020. [https://www.gov.br/defesa/pt-br/arquivos/lai/transparencia\\_e\\_prestacao\\_de\\_contas/Norma\\_InTcu84.2020.pdf](https://www.gov.br/defesa/pt-br/arquivos/lai/transparencia_e_prestacao_de_contas/Norma_InTcu84.2020.pdf). Acesso em: 15 abr. 2023.

CALADO, João Eudes de Souza; MATIAS-PEREIRA, José; RODRIGUES, Evaldo César Cavalcante; COSTA, Abimael de Jesus Castro. Tópicos contemporâneos da pesquisa contábil no setor público sob a ótica interdisciplinar: um estudo exploratório e descritivo. **Revista de Gestão e Contabilidade da UFPI**, no prelo.

CAMILO, Cássio Oliveira; SILVA, João Carlos da. **Mineração de dados**: conceitos, tarefas, métodos e ferramentas. Relatório técnico, Instituto de Informática, Universidade Federal de Goiás (UFG), p. 1-29, 2009. Disponível em: <https://ww2.inf.ufg.br/node/355>. Acesso em: 15 abr. 2023.

CASTRO, Sarah; COSTA, Abimael. O status quo do relato integrado no setor público brasileiro. **Revista Gestão Organizacional**, Chapecó, v. 15, n. 3, p. 254-270, set./dez, 2022. DOI: <https://doi.org/10.22277/rgo.v15i3.6542>. Disponível em: <https://bell.unochapeco.edu.br/revistas/index.php/rgo/article/view/6542>. Acesso em: 15 jan. 2024.

CURK, Tomaz.; DEMSAR, Janez; XU, Qikai; LEBAN, Gregor; PETROVIC, Uros; BRATKO, Ivan; SAULSKY, Gad; ZUPAN, Blaz. Microarray data mining with visual programming. **Bioinformatics**, Oxford University, v. 21, n. 3, p. 396-398, fev. 2005. DOI: <https://doi.org/10.1093/bioinformatics/bth474>. Disponível em: <https://academic.oup.com/bioinformatics/article/21/3/396/237562>. Acesso em: 15 abr. 2023.

DEMSAR, Janez; CURK, Tomaz.; ERJAVEC, Ales.; GORUP, Crt.; HOCEVAR, Tomaz; MILUTINOVIC, Mitar; MOZINA, Martin; POLAJNAR, Matija; TOPLAK, Marko; STARIC, Anze; STAJDOHAR, Miha; UMEK, Lan; ZAGAR, Lan; ZBONTAR, Jure; ZITNIK, Marinka; ZUPAN, Blaz. Orange: Data Mining Toolbox in Python, **Journal of Machine Learning Research**, [s. l.], v. 14, p. 2349-2353, ago. 2013. Disponível em: <https://jmlr.org/papers/volume14/demsar13a/demsar13a.pdf>. Acesso em: 15 abr. 2023.

DEMSAR, Janez; ZUPAN, Blaz; LEBAN, Gregor; CURK, Tomaz. Orange: from experimental machine learning to interactive Data Mining. In: BOULICAUT, Jean-François; ESPOSITO, Floriana; GIANNOTTI, Fosca; PEDRESCHI, Dino (eds.). **Knowledge discovery in databases: PKDD 2004**. Springer-Verlag Berlin Heidelberg, 2004. p. 537-539. Disponível em: [https://link.springer.com/chapter/10.1007/978-3-540-30116-5\\_58](https://link.springer.com/chapter/10.1007/978-3-540-30116-5_58). Acesso em: 16 abr. 2023.

DUMAY, John; BERNARDI, Cristiana; GUTHRIE, James; DEMARTINI, Paola. Integrated Reporting: A Structured Literature Review. **Accounting Forum**, Londres, v. 40, n. 3, p. 166-185, 2016. DOI: <http://dx.doi.org/doi:10.1016/j.accfor.2016.06.001>. Disponível em: <https://www.tandfonline.com/doi/full/10.1016/j.accfor.2016.06.001>. Acesso em: 21 abr. 2023.



FATIMA, Samar; SOUZA, Kevin C. de; DAWSON, Gregory S. National strategic artificial intelligence plans: a multi-dimensional analysis. **Economic Analysis and Policy**, Austrália, v. 67, p. 178-194, set. 2020. DOI: <https://doi.org/10.1016/j.eap.2020.07.008>. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0313592620304021?via%3Dihub>. Acesso em: 18 jan. 2024.

GUTHRIE, James; PARKER, Lee D.; DUMAY, John; MILNE, Markus J. What counts for quality in interdisciplinary accounting research in the next decade: a critical review and reflection. **Accounting, Auditing & Accountability Journal**, [s. l.], v. 32, n. 1, p. 2-25, jan. 2019. DOI: <https://doi.org/10.1108/AAAJ-01-2019-036>. Disponível em: <https://www.emerald.com/insight/content/doi/10.1108/AAAJ-01-2019-036/full/html>. Acesso em: 21 abr. 2023.

HOZAIRI, Hozairi; ANWARI, Anwari; ALIM, Syariful. Orange Data mining implementation for student graduation classification using K-Nearest Neighbor, Decision Tree and Naive Bayes Models. **Network Engineering Research Operation**, [s. l.], v. 6, n. 2, 2021. Disponível em: <https://www.mendeley.com/catalogue/e4e2d21c-9a8a-3ba8-9a3a-252e01bb9dfc/>. Acesso em: 2 jun. 2023.

INTERNATIONAL INTEGRATED REPORTING COUNCIL. **Integrated Reporting Framework**. Londres: IIRC, 2013. Disponível em: <https://www.integratedreporting.org/resource/international-ir-framework/>. Acesso em: 22 abr. 2023.

JOVIC, Alan; BRKIC, Karla; BOGUNOVIC, Nikola. An overview of free software tools for general data mining. *In: International Convention on Information and Communication Technology, Electronics and Microelectronics*, 37., 2014. Opatija: **IEEE Xplore**, 2014. p. 1112-1117. Disponível em: <https://ieeexplore.ieee.org/document/6859735>. Acesso em: 15 abr. 2024.

KODATI, Sarangam; VIVEKANANDAM, Reethu. Analysis of heart disease using in data mining tools Orange and Weka. **Global Journal of Computer Science and Technology**, [s. l.], v. 18, n. 1, versão 1.0, 2018. Disponível em: <https://api.semanticscholar.org/CorpusID:145873002>. Acesso em: 15 abr. 2023.

MARWA, Emelia Aprodaid; KRISTANTO, Ari Budi. Analisis sentimen pengungkapan informasi manajemen: text mining berbasis metode VADER. **Owner: Riset Dan Jurnal Akuntansi**, [s. l.], v. 6, n. 3, p. 2973-2984, jul. 2022. DOI: <https://doi.org/10.33395/owner.v6i3.895>. Disponível em: <https://owner.polgan.ac.id/index.php/owner/article/view/895>. Acesso em: 12 jun. 2023.

OTIA, Javis Ebu; BRACCI, Enrico. Digital transformation and the public sector auditing: the SAI's perspective. **Financial Accountability & Management**, [s. l.], v. 38, n. 2, p. 252-280, maio 2022. DOI: <https://doi.org/10.1111/faam.12317>. Disponível em: <https://onlinelibrary.wiley.com/doi/10.1111/faam.12317>. Acesso em: 22 jan. 2024.

PEKER, Musa; ÖZKARACA, Osman; Sasar, Ali. Use of Orange Data Mining Toolbox for data analysis in clinical decision making: the diagnosis of diabetes disease. **IGI Global**, Harshey, cap. 7, p. 143-167, 2018. Disponível em: <https://www.igi-global.com/gateway/chapter/205478>. Acesso em: 2 jun. 2023.



VAISHNAV, Devashree; RAO, B. Rama. Comparison of machine learning algorithms and fruit classification using Orange Data Mining Tool. *In: International Conference on Inventive Computation Technologies*, 3, 2020. Coimbatore: **IEEE Xplore**, 2020. Disponível em: <https://ieeexplore.ieee.org/document/9034442>. Acesso em: 15 abr. 2024.

WAHBERH, Abdullah H.; AL-RADAIDEH, Qasem A.; AL-KABI, Mohammed N.; AL-SHAWAKFA, Emad M. A comparison study between Data Mining Tools over some classification methods. **International Journal of Advanced Computer Science and Applications**, [s. l.], v. 8, n. 2, p. 18-26, 2011. Disponível em: <https://thesai.org/Publications/ViewPaper?Volume=1&Issue=3&Code=SpecialIssue&SerialNo=4>. Acesso em: 5 jun. 2023.

WIGUNA, Ratu Aghnia Raffaidy; RIFAI, Andri Irfan. analisis text clustering masyarakat di twitter mengenai omnibus law menggunakan Orange Data Mining. **Journal of Information Systems and Informatics**, [s. l.], v. 3, n. 1, p. 1-12, mar. 2021. DOI: <https://doi.org/10.33557/journalisi.v3i1.78>. Disponível em: <https://journal-isi.org/index.php/isi/article/view/78>. Acesso em: 19 abr. 2024.

WIRTZ, Bernd. W.; WEYERER, Jan C.; GEYER, Carolin. Artificial Intelligence and the public sector: applications and challenges. **International Journal of Public Administration**, Londres, v. 42, n. 7, p. 596-615, jul. 2018. DOI: <https://doi.org/10.1080/01900692.2018.1498103>. Disponível em: <https://www.tandfonline.com/doi/full/10.1080/01900692.2018.1498103>. Acesso em: 10 jan. 2024.

ZHANG, Shichao; ZHANG, Chengqi; YANG, Qiang. Data preparation for Data Mining. **Applied Artificial Intelligence**, Taylor & Francis, v. 17, n. 5-6, p. 375-381, 2003. DOI: <https://doi.org/10.1080/713827180>. Disponível em: <https://www.tandfonline.com/doi/abs/10.1080/713827180>. Acesso em: 15 abr. 2023.

Os conceitos e interpretações emitidos nos trabalhos assinados são de exclusiva responsabilidade de seus autores.

