



8° CIM

CONGRESSO INTERNACIONAL MULTIDISCIPLINAR

AS COMPETÊNCIAS SOCIOEMOCIONAIS NO MUNDO PROFISSIONAL

LEVANTAMENTO DE PROCESSOS DE MINERAÇÃO DE DADOS NO CONTEXTO DE BASES GOVERNAMENTAIS: UMA ABORDAGEM INTEGRATIVA PARA COMPILAÇÃO E ANÁLISE DE INFORMAÇÕES ESTRATÉGICAS

Rogério Sousa e Silva¹

Rodrigo Cândido Borges²

Henrique Valle de Lima³

Karla de Aleluia Batista⁴

Daniel Soares de Souza⁵

RESUMO

Diante da crescente abundância de dados gerados na era da tecnologia, é imperativo que a informação represente mais que uma simples análise de dados. Na esfera governamental, a análise cuidadosa de dados pode ampliar a compreensão de fenômenos complexos ao passo que se torna essencial para impulsionar políticas, alocar recursos eficientemente e gerar impactos positivos. Nesse contexto, o presente artigo propõe uma investigação sobre a aplicação de técnicas de mineração de dados em contextos governamentais e bases de proximidade. Apresenta-se uma revisão crítica de casos práticos da literatura relevante, proporcionando uma compreensão das complexidades associadas à mineração de dados em bases governamentais. Dessa maneira, a proposta pode contribuir atuando como uma referência para pesquisadores e profissionais envolvidos na interseção entre a ciência de dados e a administração pública.

Palavras chave: base de dados; governo; mineração.

ABSTRACT

In the face of the growing abundance of data generated in the technology era, information must represent more than mere data analysis. In the governmental sphere, careful data analysis can enhance

¹ Rogério Sousa e Silva, Mestre em Ciência da Computação, Instituto Federal de Goiás, rogerio.sousa@ifg.edu.br.

² Rodrigo Cândido Borges, Doutor em Ciências da Saúde, Instituto Federal de Goiás, rodrigo.borges@ifg.edu.br.

³ Henrique Valle de Lima, Doutor em Ciência da Computação, Universidade Evangélica de Goiás, henrique.lima@unievangelica.edu.br.

⁴ Karla de Aleluia Batista, Doutora em Ciências Biológicas, Instituto Federal de Goiás, karla.batista@ifg.edu.br.

⁵ Daniel Soares de Souza, Mestre em Gestão Pública, Instituto Federal de Brasília, daniel.souza@ifb.edu.br.



the understanding of complex phenomena while becoming essential for driving policies, efficiently allocating resources, and generating positive impacts. This article proposes an investigation into the application of data mining techniques in governmental contexts and proximity databases. A critical review of practical cases from relevant literature is presented, providing an understanding of the complexities associated with data mining in governmental databases. Thus, the proposal may serve as a reference for researchers and professionals engaged in the intersection of data science and public administration.

Keywords: data base; government; mining.

1. INTRODUÇÃO

A rápida evolução tecnológica característica da era da informação, tem gerado uma vertiginosa expansão do volume de dados gerados, atingindo picos até então não vivenciados. Registros heterogêneos, originados em diversos setores, agora se encontram facilmente acessíveis, configurando um panorama em que os dados assumem uma posição estratégica central na orientação de processos decisórios em diferentes domínios. Este fenômeno reflete a capacidade que os dados têm de não serem apenas como fontes de informação, mas atuarem como ativos estratégicos indispensáveis no atual contexto tecnológico (YOON, 2021).

A abundância de dados e informações estende-se não apenas aos domínios tradicionais, como saúde e educação, mas permeia áreas importantes, como transações financeiras governamentais, distribuição de renda entre governos estaduais e estratégias de cuidado populacional. Não obstante a aparente profusão desses dados, a compreensão abrangente da essência frequentemente se extravia na intrincada diversidade e complexidades agregadas (CATE, 2008). É nesse contexto que a mineração de dados se revela como um elemento catalisador essencial, transcendendo a mera compilação de informações e adentrando no âmbito estratégico, desvelando o potencial latente das vastas fontes governamentais em fornecer *insights* valiosos para análise e tomada de decisões estratégicas (SLOBOGIN, 2008).

No contexto nacional brasileiro, caracterizado por pronunciadas disparidades regionais, a habilidade de alocar recursos de maneira eficaz é vital no conjunto de políticas governamentais. A aplicação de técnicas avançadas de mineração de dados em bases governamentais não apenas revela lacunas e carências distintas em diversas regiões, mas também proporciona entendimentos substanciais capazes de orientar de forma prática e embasada as políticas de investimento (SILVA; DA ROCHA, 2010). Concebe-se, assim, um cenário no qual um governo habilitado pela mineração de dados adquire a destreza para



direcionar recursos com precisão, fomentando o desenvolvimento nas áreas mais necessitadas e mitigando desigualdades históricas arraigadas (VICTORINO et al., 2017). Essa capacidade estratégica de alocação de recursos não somente otimiza os resultados sociais e econômicos, mas também configura um instrumento eficaz na busca pela equidade regional e no combate a disparidades históricas persistes (PALMIERI; FRANKLIN, 2023).

Em uma ótica equidistante, a adoção da transparência no âmbito governamental representa um benefício de relevância incontestável decorrente da aplicação da mineração de dados. A metodologia integrativa delineada neste estudo, ao elencar o processo de convergência, organização e interpretação de dados aparentemente fragmentados, desvela padrões e relações previamente não discernidos (AMARAL, 2018). Este processo não apenas alarga a apreensão das intrincadas dinâmicas sociais e econômicas, mas também intensifica a responsabilidade governamental, facultando à sociedade uma visão mais perspicaz das decisões e ações empreendidas. A capacidade intrínseca da abordagem integrativa para revelar correlações complexas e dinâmicas não evidentes atua como um instrumento fundamental para a promoção da transparência no ambiente governamental, fomentando um entendimento mais robusto e informado por parte dos cidadãos sobre as operações e diretrizes administrativas. Essa transparência, por conseguinte, constitui um alicerce essencial para a fortificação da confiança entre a administração pública e a sociedade civil (BRAGA; DRUMMOND, 2018).

Neste contexto, o presente artigo propõe uma análise da aplicação de técnicas de mineração de dados, direcionando o foco para a destacada importância estratégica inerente a essa prática. A investigação a ser conduzida aborda casos práticos através da revisão crítica da literatura relevante associada. Essa abordagem, longe de se restringir a uma análise superficial dos dados, visa transformá-los em instrumentos robustos, consolidando-os como elementos fundamentais para o aprimoramento dos processos governamentais e para a promoção sustentável do desenvolvimento.

Este trabalho está organizado da seguinte forma. Na Seção 2, são apresentados os conceitos básicos e trabalhos relacionados à área de mineração de dados. Na Seção 3 são apresentados os métodos propostos para o trabalho e uma investigação acerca da mineração de dados em bases governamentais. Na Seção 4 é realizada a análise das questões discutidas no trabalho para fornecer uma melhor compreensão dos métodos empregados. Na Seção 5, as principais técnicas de mineração são confrontadas. Por fim, na Seção 6 encontram-se as considerações finais.



2. FUNDAMENTAÇÃO TEÓRICA E TRABALHOS RELACIONADOS

Nesta seção, estabelece-se uma base conceitual para entendimento do processo de mineração de dados, principalmente aplicados em bases governamentais e de proximidade. Em sintonia com o cenário contemporâneo, onde a gestão eficiente de dados governamentais é imperativa, são exploradas as nuances específicas inerentes à aplicação de técnicas de mineração de dados nesses contextos. A inter-relação entre bases governamentais e de proximidade é abordada para ampliar a compreensão das complexidades subjacentes, enfatizando a importância da extração de conhecimentos significativos dessas fontes heterogêneas. A literatura pertinente é examinada para identificar métodos e tópicos abordados, cronologias e estratégias de visualização presentes nos estudos.

A primeira fase na mineração de dados é a mineração de conteúdos disponibilizados na web. A mineração de conteúdo envolve o processo no qual estatísticas são verificadas e uma sondagem de manuscritos, imagens e gráficos dentro de sites é realizada, como indicado por (NUGROHO; MAHENDRA; BUDI, 2021; VARNAGAR et al., 2013). O processo de mineração de conteúdo é utilizado para avaliar a relevância das informações da página da web para uma busca específica. Quando o agrupamento de sites é concluído, realiza-se uma varredura por meio de um processo de mineração de estrutura. A varredura fornece resultados aplicáveis e a entrada da consulta de pesquisa é arquivada. Os resultados alcançados são fundamentados na aplicação de técnicas apropriadas de mineração de dados para identificar o conteúdo mais pertinente (CATE, 2008). Além disso, as estatísticas coletadas são usadas para auxiliar no desenvolvimento de páginas da web adaptativas, produtos e serviços de suporte (SILVA; DA ROCHA, 2010).

Em (BRAGA; DUMMOND, 2018) os autores abordam a extração de informações de bases de dados governamentais abertas, adotando uma abordagem de mineração descritiva. No contexto algorítmico, o trabalho demonstra uma aplicação eficaz de técnicas analíticas para revelar padrões e características relevantes presentes nos conjuntos de dados governamentais. A implementação dessas técnicas envolve a utilização de algoritmos específicos, que não são detalhados na descrição fornecida, mas presume-se que incluam métodos de mineração de dados descritivos, tais como estatísticas descritivas, agrupamento e visualização de dados. O artigo ressalta a escolha estratégica da ferramenta R, uma linguagem de programação e ambiente estatístico, indicando que ela foi utilizada como instrumento central para a análise exploratória dos dados governamentais. Por fim, os autores buscam extrair informações das bases de dados governamentais, que contribuem para uma



compreensão mais profunda e transparente das informações disponíveis. Para uma avaliação mais abrangente, seria importante incluir detalhes sobre os métodos específicos utilizados, os algoritmos implementados e os resultados obtidos.

Em uma ótica de proximidade, o autor em (AMARAL, 2018) propõem uma investigação que emprega mineração de dados em conjuntos governamentais para analisar o perfil dos eleitores no Brasil. O trabalho utiliza técnicas de mineração de dados, indicando uma abordagem analítica para explorar padrões, correlações e características relevantes nos conjuntos de dados eleitorais. Quanto aos resultados, o artigo fornece uma visão específica acerca da análise dos dados eleitorais.

Já em (SILVA; RALHA, 2010), os autores apresentam um estudo que utiliza técnicas de mineração de dados para detectar cartéis em licitações, empregando algoritmos de análise de padrões, agrupamento e detecção de anomalias. O K-Means é adotado para identificar padrões de comportamento entre os participantes das licitações. Além disso, o *Isolation Forests*, é aplicado para identificar comportamentos fora do padrão que podem indicar práticas cartelizadas. A análise estatística e visualização de dados é incorporada para compreender a distribuição e relacionamentos entre variáveis. Os resultados incluem a identificação de padrões suspeitos de comportamento cartelizado, a indicação de participantes frequentes em licitações e a detecção de anomalias que mereçam investigação adicional.

Por fim, os autores em (GOMIDE et al., 2023) apresentam uma perspectiva inovadora ao explorar técnicas de mineração de dados para analisar as despesas públicas em municípios mineiros, visando identificar potenciais alertas de fraudes. O trabalho emprega o *One-Class SVM* como o algoritmo de detecção de anomalias. Para lidar com a complexidade dos dados financeiros municipais, são aplicadas técnicas de pré-processamento, incluindo normalização, tratamento de *outliers* e manipulação de valores ausentes. A integração de dados provenientes de diversas fontes, como orçamentos, licitações e folhas de pagamento, é importante para identificação do conjunto de respostas esperado. Os resultados esperados incluem a identificação de transações anômalas, indicando áreas de risco que merecem uma investigação mais aprofundada.



3. METODOLOGIA

Na busca por uma compreensão aprofundada da mineração de dados em bases governamentais, a coleta de dados, combinada às práticas mais relevantes encontradas na literatura atual, constituiu a base para o desenvolvimento do presente trabalho. Dessa maneira, este artigo se orienta por uma abordagem de mapeamento sistemático, que visa caracterizar e resumir o processo geral de revisão de diversos trabalhos.

3.1 Questões de Pesquisa

Com base nas estratégias detalhadas em (CURCIO et al., 2018), o principal objetivo ao realizar este estudo de mapeamento sistemático foi definir como a mineração de dados na web é abordada em contextos governamentais. Este objetivo foi subdividido em quatro questões de pesquisa, conforme apresentadas a seguir:

- Q-1: Quais são os principais métodos para conduzir a mineração de dados na web?
- Q-2: Quais tópicos são abordados na mineração de dados na web em contextos governamentais?
- Q-3: Quando e onde os estudos foram publicados?
- Q-4: Como os estudos foram conduzidos em termos de visualização dos resultados?

3.2 Busca Bibliográfica

Para identificação e seleção de estudos primários de pesquisa, utilizaram-se as bases de dados principais ABI/INFORM, DSL e Google Scholar. Estas bases abrigam uma vasta coleção de artigos de pesquisa revisados por pares no campo da mineração de dados em bases governamentais e de proximidade, proporcionando recursos eficazes, como filtros de data, filtros de tópicos similares, a capacidade de filtrar por citação de um artigo específico e a habilidade de incluir ou excluir patentes.

As buscas nessas bases foram realizadas utilizando frases-chave, tais como “mineração de dados em bases governamentais”, “técnicas de mineração de dados” e “métodos de mineração de dados governamentais”. Para a base Google Scholar, por exemplo, ao se buscar por “técnicas de mineração de dados” são disponibilizados aproximadamente 155.000 estudos correlacionados.



3.3 Seleção de Bibliografia

Os estudos de pesquisa foram selecionados com base em diversos critérios relacionados às características da base de dados. Os critérios são os seguintes: aplicação de busca na base de dados, remoção de artigos anteriores ao ano de 2010, aplicação de busca por texto completo, busca por artigos submetidos ao processo de revisão por pares, busca por trabalhos em idioma português ou inglês, exclusão de trabalhos com menos de 10 citações, exclusão de trabalhos que atuem fora contexto governamental ou áreas correlatas. A avaliação da qualidade baseou-se na robustez do material citado, sendo que artigos sem citações foram excluídos.

Os critérios de inclusão aplicados a cada base de dados são os seguintes:

- Estudos que apresentam os métodos e resultados da mineração de dados em bases governamentais;
- Estudos publicados entre 2010 e 2024;
- Estudos conduzidos especificamente no campo da mineração de dados.

Os critérios de exclusão aplicados às bases de dados são os seguintes:

- Estudos que apresentam texto incompleto ou abreviado;
- Estudos que não passaram por revisão por pares;
- Estudos não apresentados em língua portuguesa ou inglesa.

4. RESULTADOS

A seção de resultados deste artigo é dedicada à análise e resposta das quatro questões levantadas na seção 3.1 voltadas para aspectos fundamentais no contexto da mineração de dados na web aplicada em contextos governamentais e bases de proximidade. Cada tópico foi investigado de maneira a fornecer uma compreensão aprofundada dos principais métodos empregados na condução da mineração de dados na web.

4.1. Q-1: Quais são os principais métodos para conduzir a mineração de dados na web?

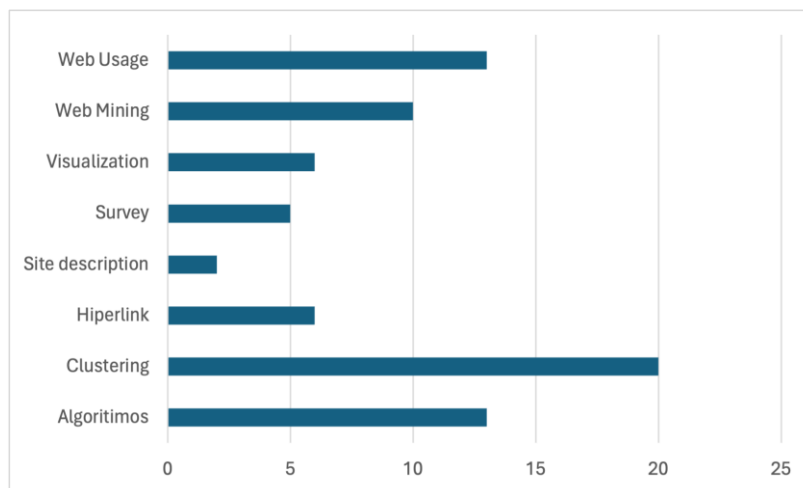
O campo da mineração de dados na web apresenta uma ampla gama de métodos e técnicas, evidenciando uma diversidade de abordagens identificadas em estudos correlatos. Na análise do Google Scholar, emergem como principais métodos aqueles relacionados a



algoritmos e métodos de agrupamento, estatísticas, mineração web, mineração de uso da web e técnicas de visualização.

A abordagem de coleta aleatória de dados revela a presença de métodos diversos, como algoritmos, agrupamento, mineração de dados, pré-processamento de dados, análise de redes de hiperlinks, estatísticas multivariadas, análise quantitativa da estrutura de hiperlinks, métodos estatísticos de visualização de dados, mineração de conteúdo web e mineração de estrutura da web. Mais detalhes são encontrados na Figura 1.

Figura 1: Principais métodos utilizados para mineração de dados em bases governamentais



Fonte: De autoria própria.

Essa notável diversidade de métodos evidencia a intrincada complexidade e a amplitude de abordagens adotadas no cenário da mineração de dados na web. Este panorama reflete a contínua busca por estratégias inovadoras e adaptáveis diante da crescente complexidade das informações disponíveis na web.

4.2. Q-2: Quais tópicos são abordados na mineração de dados na web em contextos governamentais?

A abordagem adotada para classificar os tópicos nesta pesquisa foi embasada no corpus de conhecimento existente sobre mineração de dados na web, buscando alcançar uma representação direcionada das atividades desenvolvidas nesse domínio, especialmente em contextos governamentais. A pesquisa, embora tenha focalizado predominantemente a mineração web, abordou de maneira abrangente todas as atividades correlatas pertinentes



ao escopo do estudo. O mapeamento detalhado proporcionou uma identificação precisa de lacunas significativas na pesquisa, destacando áreas que necessitam de investigação mais aprofundada para um entendimento claro do panorama da mineração de dados governamentais.

Essas categorias incluem técnicas fundamentais, como *data mining*, que envolve a extração de informações úteis e padrões a partir de conjuntos de dados relacionados à informações governamentais. *Web usage* referindo-se à análise da web, ao passo que *web log* diz respeito à exploração e análise de registros encontrados diretamente na web. *Clustering* envolvendo a categorização de dados governamentais em grupos, enquanto *hyperlink* busca tratar da investigação das ligações entre informações posicionadas em diferentes páginas web. Pode-se ainda encontrar *Fuzzy* referindo-se à aplicação de lógica fuzzy na mineração de dados, e *Government search* envolvendo técnicas específicas de busca voltadas especificamente para contextos governamentais. Esses tópicos abrangem desde abordagens fundamentais até técnicas mais avançadas, oferecendo uma visão abrangente e estratificada da pesquisa na área.

Vale ressaltar que a pesquisa não se restringiu unicamente à mineração web, mas abordou de maneira abrangente todas as atividades correlatas, incluindo aspectos como a mineração de estrutura de web, mineração de conteúdo web, análise de redes de hiperlinks, entre outros. Essa abordagem integral permite uma compreensão mais rica e contextualizada dos desafios e oportunidades presentes na mineração de dados em âmbito governamental, ressaltando a importância de uma abordagem multidimensional para explorar a complexidade inerente a esse domínio de pesquisa. O campo da mineração de dados na web apresenta uma ampla gama de métodos e técnicas, evidenciando uma diversidade de abordagens identificadas em estudos correlatos. Na análise do Google Scholar, emergem como principais métodos aqueles relacionados a algoritmos e métodos de agrupamento, estatísticas, mineração web, mineração de uso da web e técnicas de visualização.

4.3. Q-3: Quando e onde os estudos foram publicados?

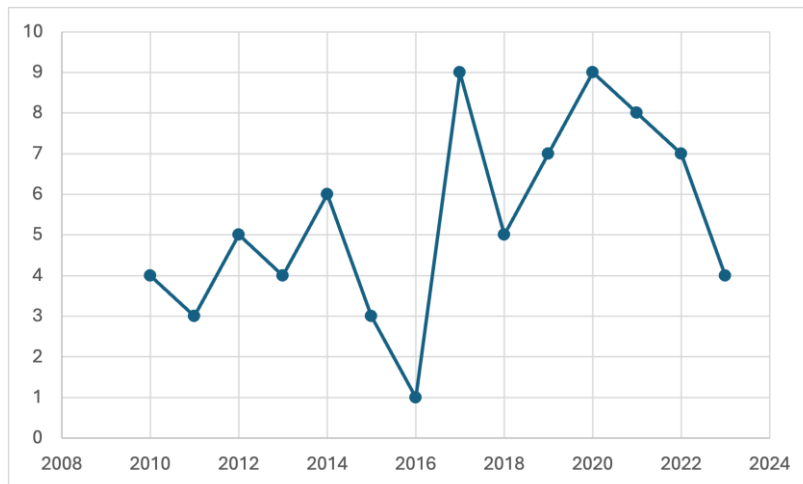
Ao realizar uma análise longitudinal da frequência de publicações no período compreendido entre 2010 e 2024 no âmbito da mineração de dados na web, constatou-se um crescimento consistente no interesse por esse campo de pesquisa. O exame detalhado revelou um severo declínio pontual em 2016, seguido reduções suaves em 2018 e 2023,



indicando possíveis mudanças nas agendas de pesquisa ou na dinâmica intrínseca do campo de pesquisa.

A Figura 2 proporciona uma visualização elucidativa desse padrão temporal nas publicações. No tocante aos locais de publicação, a análise abarcou distintos formatos, abrangendo desde trabalhos apresentados em conferências até artigos em periódicos e participações em *workshops*. A diversidade de plataformas evidencia uma disseminação ampla e colaborativa do conhecimento nesse domínio específico, sugerindo uma comunidade científica engajada e aberta ao intercâmbio de informações.

Figura 2: Levantamento do número de publicações por ano de 2010 a 2023 relacionadas a mineração de dados em bases governamentais



Fonte: De autoria própria.

4.4. Q-4: Como os estudos foram conduzidos em termos de visualização dos resultados?

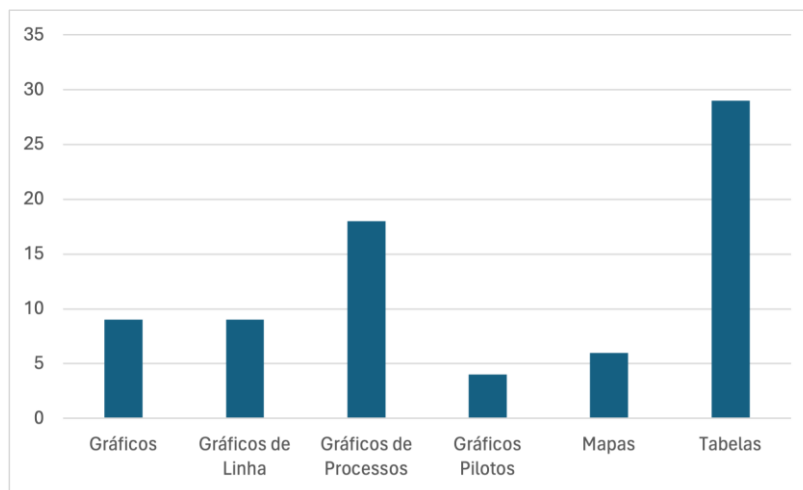
A investigação dos métodos de visualização revelou uma variedade de abordagens, sendo preponderante a utilização de gráficos e tabelas para representar de maneira eficaz a quantidade relativa de dados em cada categoria, conforme evidenciado na Figura 3. Essa abordagem tradicional é complementada por outras formas de representação visual, como mapas, gráficos de processos, gráficos de linha, gráficos pilotos e gráficos de barras. Cada uma dessas técnicas tem suas particularidades:

- Maps (Mapas): Permitem a representação geoespacial dos dados, facilitando a compreensão de padrões geográficos e distribuição regional;



- Tables (Tabelas): Estrutura tabular que apresenta dados de maneira organizada e sistemática, tornando a análise mais acessível e compreensível;
- Charts (Gráficos): Incluindo gráficos de barras, de linhas e outros formatos visuais para ilustrar tendências, padrões e relações nos dados;
- Process (Gráficos de Processos): Visualizações que descrevem etapas sequenciais ou processos, fornecendo uma perspectiva cronológica e hierárquica dos dados;
- Pilots (Gráficos Pilotos): Representações visuais que proporcionam uma visão geral ou panorâmica dos dados, permitindo identificar rapidamente aspectos relevantes;
- Line Graph (Gráficos de Linha): Indicam a variação de uma ou mais variáveis ao longo do tempo, proporcionando visões sobre tendências e flutuações.

Figura 3: Levantamento das estratégias de apresentação de resultados em trabalhos relacionadas a mineração de dados em bases governamentais



Fonte: De autoria própria.

5. DISCUSSÃO

Esta pesquisa apresenta uma discussão acerca da aplicação de técnicas de otimização de mineração de dados aplicado no contexto de bases heterogêneas. O objetivo é mitigar redundâncias de dados originárias de técnicas de agrupamento provenientes da mineração na web, com especial ênfase no contexto governamental (DA SILVA; SIRQUEIRA, 2022). O estudo engloba uma análise da estratégia de otimização de *clusters*, destacando a importância do acesso a representações gráficas contidas nos registros web para a



identificação de comportamentos simétricos e assimétricos em gestões públicas. A análise de padrões, componente crítico dessa abordagem, é conduzida por meio de registros web previamente depurados (SILVA; RALHA, 2010).

A avaliação de algoritmos adotados, como PageRank, PageRank Ponderado e HITS, foi realizada (JAIN; PUROHIT, 2011). Utilizando uma ferramenta para de mineração na web, explorou-se informações derivadas de padrões de movimentações convencionais. Ressalta-se a importância da mineração da estrutura web no âmbito governamental, onde a classificação de páginas compatíveis e apropriadas é efetivada por meio de diversos algoritmos especializados. PageRank, PageRank Ponderado e HITS são notáveis por conferirem tratamento uniforme a todos os links, compartilhando, simultaneamente, a pontuação de classificação (SHARMA et al., 2020).

Visando aprimorar os índices de assertividade, a discussão abarca ambos os níveis do modelo preditivo. O primeiro nível tem como meta filtrar categorias com alta probabilidade de atrair público de interesse, pessoas que se relacionam com o governo de alguma forma, que possam fornecer informações que enriqueçam o conjunto da análise, enquanto o segundo nível se concentra especificamente nas páginas web de maior probabilidade e conteúdo disponíveis. Todavia, é crucial reconhecer o desafio intrínseco ao comportamento heterogêneo do usuário no âmbito governamental (COLAK et al., 2012).

Na busca pela intensificação da personalização do conteúdo web, a revisão de registros web governamentais é conduzida por meio de ferramentas avançadas de mineração de dados (MASSEGLIA et al., 2000). Essa abordagem engloba uma análise detalhada da utilidade dos documentos, sob a perspectiva do usuário, atribuindo pontuações mais elevadas a *clusters* que apresentam correspondência ou semelhança com palavras-chave especificadas. Contrastando com sistemas predominantes que se limitam a buscas baseadas em conteúdo, essa abordagem é considerada mais eficaz em termos de especificidade no contexto governamental (VARNAGAR et al, 2013). Adicionalmente, destaca-se a proposta de identificar automaticamente pontos de referência e eventos em imagens marcadas, enfatizando a relevância dessas detecções na análise e alinhamento de extensos conjuntos de imagens compartilhadas pelos usuários que tenham relação com medidas e implementações oriundas de recursos advindos do governo (ALLAN, 2002).

Em uma ótica semelhante, os hiperlinks são utilizados como meios de interconectar diversas páginas da web, constituindo uma fonte inestimável de informações e



desempenhando um papel significativo na mineração de dados governamentais. Dessa maneira, esta pesquisa empreende uma análise da estrutura de sites governamentais, incorporando processos de *webometria* para identificar informações latentes nos hiperlinks (VANTI, 2002).

A mineração de gráficos, elemento vital na prática de mineração de dados, é explorada no contexto governamental para a identificação de subgráficos repetitivos, destacando a aplicação na derivação eficiente de subgráficos candidatos. Este estudo proporciona uma análise dos processos periodicamente adotados nessa técnica, abordando desafios específicos relacionados à complexidade dos dados governamentais. Adicionalmente, são avaliados algoritmos de máquinas de estados finitos (FSM) em termos de estratégias de busca e abordagens para a contagem de frequências durante processos de categorização aplicados a dados governamentais (SZOSTAK et al., 2021).

No âmbito da comunicação governamental, blogs, enquanto apresentações sociais baseadas na web, são considerados neste tratado científico. Apresenta-se um modelo abrangente para avaliação de inovações no design e conteúdo de blogs governamentais, contribuindo para uma compreensão aprofundada do papel dessas ferramentas na divulgação e comunicação governamental. Essa análise minuciosa é essencial para a compreensão das dinâmicas de informação governamental na era digital (AGERDAL-HGERMIND; VALENTINI, 2015).

A mineração na web, como uma aplicação essencial para a compreensão aprofundada das necessidades de suporte em programas e serviços online, constitui-se como um campo de estudo em constante evolução. Esta abordagem, direcionada à identificação de padrões de uso nos dados web, se desdobra em três fases distintas e cruciais: pré-processamento, descoberta de padrões e análise de padrões. Este panorama científico abarca ainda uma revisão de estudos vinculados a essa técnica de rápida expansão (ZHANG et al., 2024).

No contexto contemporâneo, a aplicação da mineração na web em bases governamentais destaca-se como um aspecto crítico. Os dados resultantes dessa prática tornam-se fundamentais para uma compreensão mais profunda dos padrões gerais de utilização da web no âmbito governamental. Além de otimizar a compreensão das informações disponíveis, esta técnica destaca-se pela aplicabilidade mais eficaz aos usuários finais do setor público. O crescente interesse governamental nessa forma de análise ressalta o papel



preponderante no panorama tecnológico. Contudo, os desafios inerentes à mineração na web em contextos governamentais suscitam questões que exigem respostas substanciais antes do desenvolvimento de ferramentas robustas (NUGROHO; MAHENDRA; BUDI, 2021).

A implementação da mineração de dados em sistemas educacionais, por exemplo, com foco específico em cursos online, sistemas de gerenciamento de conteúdo de aprendizagem e sistemas educacionais adaptativos e intelectuais baseados na web, constitui uma área de investigação de grande relevância para direcionamento de decisões estratégicas governamentais. Esse domínio específico demanda requisitos singulares para uma aplicação eficaz da mineração de dados, apresentando complexidades distintas de outros setores. A inserção de agentes recomendadores em sistemas de *e-learning*, voltados para a busca e compilação de recursos educacionais disponíveis na web, figura como um aspecto crucial desse processo. Para alcançar um nível de sucesso equiparável a outras esferas, a mineração de dados educacionais surge como uma análise que demanda de especialização mais aprofundada e um enfoque mais direcionado dentro do contexto educacional (YANRONG, 2021).

Neste contexto, a integração sinérgica da mineração na web governamental e a análise de dados educacionais podem proporcionar uma compreensão holística, possibilitando não apenas a eficácia na governança, mas também a otimização dos processos educacionais, promovendo um desenvolvimento mais equitativo e sustentável.

6. CONCLUSÃO

Ao concluir esta pesquisa sobre o processo de mineração de dados em meios eletrônicos, destaca-se como fator ponderante a análise aprofundada da crescente abundância de informações na era da tecnologia. A compilação de dados para análise de informações vai além da análise, especialmente no contexto governamental, onde a organização orientada ao caso é essencial para impulsionar políticas, alocar recursos eficientemente e gerar impactos positivos.

Em alinhamento com a proposta inicial, recomenda-se uma abordagem integrativa que considera tanto fatores intermediários quanto governamentais na definição da posição dos atores nos setores de serviços governamentais. Essa recomendação se torna ainda mais importante ao incorporar a mineração de dados como ferramenta essencial para analisar custos, comportamentos de usuários e eficiência operacional em bases governamentais.



A pesquisa destaca a necessidade da tomada de decisões de maneira mais robusta e do aprimoramento contínuo dos serviços oferecidos pelo governo. Assim, este estudo não apenas serve como referência para pesquisadores e profissionais na interseção entre ciência de dados e administração pública, mas também reforça a importância da análise profunda dos dados para impulsionar inovações significativas no âmbito governamental.

7. REFERÊNCIAS BIBLIOGRÁFICAS

AGERDAL-HJERMIND, A.; VALENTINI, C. Blogging as a communication strategy for government agencies: A Danish case study. **International Journal of Strategic Communication**, v. 9, n. 4, p. 293-315, 2015.

ALLAN, J. Topic detection and tracking: event-based information organization. **Springer Science & Business Media**, 2002.

AMARAL, B. C. **Análise do perfil de eleitores brasileiros com mineração de dados abertos governamentais**. Curso de Graduação em Ciência da Computação. Niterói: UFF, 2018.

BRAGA, L. C. C.; DRUMMOND, I. N. Extração de informação em bases de dados abertas governamentais através de uma abordagem de mineração descritiva empregando a ferramenta R. **Revista de Informática Aplicada**, v. 14, n. 1, 2018.

CATE, Fred H. Government data mining: The need for a legal framework. **Harv. CR-CLL Rev.**, v. 43, p. 435, 2008.

COLAK, I.; SAGIROGLU, S.; YESILBUDAK, M. Data mining and wind power prediction: A literature review. **Renewable energy**, v. 46, p. 241-247, 2012.

CURCIO, Karina et al. Requirements engineering: A systematic mapping study in agile software development. **Journal of Systems and Software**, v. 139, p. 32-50, 2018.

DA SILVA, L. G.; SIRQUEIRA, T. F. M. Uma Ferramenta para Mineração de Dados de Programas Sociais do Governo Federal. **Caderno de Estudos em Sistemas de Informação**, v. 7, n. 1, 2022.

GOMIDE, L. D. et al. **Mineração de Dados sobre Despesas Públicas de Municípios Mineiros para Gerar Alertas de Fraudes**. In: Anais do XXXVIII Simpósio Brasileiro de Bancos de Dados. SBC, 2023. p. 378-383.

JAIN, Rekha; PUROHIT, G. N. Page ranking algorithms for web mining. **International journal of computer applications**, v. 13, n. 5, p. 22-25, 2011.

MASSEGLIA, F.; PONCELET, P.; CICHETTI, R. An efficient algorithm for web usage mining. **Networking and Information Systems Journal**, v. 2, n. 5/6, p. 571-604, 2000.

NUGROHO, J. D. T.; MAHENDRA, R.; BUDI, I. **Web Mining in e-Procurement: A Case Study in Indonesia**. In: Proceedings of the 2021 3rd Asia Pacific Information Technology Conference. 2021. p. 101-107.

PALMIERI, M. A.; FRANKLIN, B. L. **Práticas de mineração de dados: considerações sobre os dados governamentais abertos em Londrina**. In: Anais do Workshop de Informação, Dados e



Tecnologia-WIDaT. 2023.

SHARMA, P. S.; YADAV, D.; GARG, P. A systematic review on page ranking algorithms. **International Journal of Information Technology**, v. 12, n. 2, p. 329-337, 2020.

SILVA, C. V. S.; DA ROCHA, H. A. Técnicas de mineração de dados como apoio às auditorias governamentais. **Revista da CGU**, v. 5, n. 8, p. 29-40, 2010.

SILVA, C. V. S.; RALHA, C. G. **Utilização de técnicas de mineração de dados como auxílio na detecção de cartéis em licitações**. In: XXX Congresso Da Sociedade Brasileira de Computação. 2010. p. 1-14.

SLOBOGIN, Christopher. Government data mining and the fourth amendment. **The University of Chicago Law Review**, v. 75, n. 1, p. 317-341, 2008.

SZOSTAK, D.; WŁODARCZYK, A.; WALKOWIAK, K. Machine learning classification and regression approaches for optical network traffic prediction. **Electronics**, v. 10, n. 13, p. 1578, 2021.

VANTI, N. A. P. Da bibliometria à webometria: uma exploração conceitual dos mecanismos utilizados para medir o registro da informação e a difusão do conhecimento. **Ciência da informação**, v. 31, p. 369-379, 2002.

VARNAGAR, Chintan R. et al. **Web usage mining: a review on process, methods and techniques**. In: 2013 International Conference on Information Communication and Embedded Systems (ICICES). IEEE, 2013. p. 40-46.

VICTORINO, M. C. et al. Uma proposta de ecossistema de big data para a análise de dados abertos governamentais concetados. **Informação & sociedade**, v. 27, n. 1, p. 225-242, 2017.

YANRONG, P. **Research on Intelligent Education Method Based on Data Mining of Distance Education System**. In: 2021 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS). IEEE, 2021. p. 586-589.

YOON, Y. **Discovering knowledge in corporate databases**. In: Handbook of Data Management 1999 Edition. Auerbach Publications, 2021. p. 783-794.

ZHANG, Jing; QIANG, Jipeng; ZHOU, Cangqi. New Horizons in Web Search, Web Data Mining, and Web-Based Applications. **Applied Sciences**, v. 14, n. 2, p. 530, 2024.