Revista CADERNO PEDAGÓGICO

*REVISTA CADERNO PEDAGÓGICO* – Studies Publicações Ltda.
**ISSN: 1983-0882**

# Optimizing data mining processes in government databases: a comprehensive approach to strategic information analysis

# Otimização dos processos de mineração de dados em bancos de dados governamentais: uma abordagem abrangente para análise de informações estratégicas

# Optimización de los procesos de minería de datos en bases de datos gubernamentales: un enfoque integral para el análisis de información estratégica

DOI: 10.54033/cadpedv21n7-107

Originals received: 06/10/2024
Acceptance for publication: 06/28/2024

**Rodrigo Cândido Borges**
Doctor in Health Sciences
Institution: Instituto Federal de Goiás
Address: Inhumas, Goiás, Brasil
E-mail: rodrigo.borges@ifg.edu.br

**Rogério Sousa e Silva**
Master in Computer Science
Institution: Instituto Federal de Goiás
Address: Inhumas, Goiás, Brasil
E-mail: rogerio.sousa@ifg.edu.br

**Henrique Valle de Lima**
Doctor in Computer Science
Institution: Universidade Evangélica de Goiás
Address: Anápolis, Goiás, Brasil
E-mail: henrique.lima@unievangelica.edu.br

**Karla de Aleluia Batista**
Doctor in Biological Sciences
Institution: Instituto Federal de Goiás
Address: Goiânia, Goiás, Brasil
E-mail: karla.batista@ifg.edu.br

**Daniel Soares de Souza**
Master in Public Management
Institution: Instituto Federal de Brasília
Address: Brasília, Distrito Federal, Brasil
E-mail: daniel.souza@ifb.edu.br

## ABSTRACT

In the current technological era, marked by the massive and continuous generation of data, information must go beyond mere raw data analysis. In the governmental context, detailed data analysis is crucial to aid in the understanding and deep comprehension of complex phenomena within public administration. This type of analysis is essential for driving public policy formulation, allocating resources efficiently, and generating positive societal impacts. This article thus explores the application of data mining techniques in governmental contexts and related databases. Through a critical review of practical cases drawn from the literature, the study seeks to discuss the complexities and challenges inherent in mining data from governmental databases. The analyzed cases provide a comprehensive view of the methodologies used, highlighting best practices and common obstacles faced in the process of extracting valuable information from large volumes of governmental data. The investigation covers various aspects of data mining, from data collection and cleaning to the selection and application of specific algorithms for interpreting results. By addressing these elements, the article offers insights into the techniques and strategies for applying data mining in governmental contexts. This study is designed to serve as a reference for both researchers and professionals operating at the intersection of data science and public administration. Additionally, it aims to inform decision-making processes in the public sector, promoting more effective and efficient governance. By offering detailed insights into best practices, potential pitfalls, and emerging trends in the field of governmental data mining, the article contributes to enriching academic and professional discourse. Furthermore, it seeks to enhance public sector governance by providing the tools and understanding necessary for improved policy-making and resource management.

**Keywords:** Data Mining. Governmental Databases. Public Administration. Policy Decisions. Societal Impacts.

## RESUMO

Na era tecnológica atual, marcada pela geração massiva e contínua de dados, a informação precisa ir além da simples análise de dados brutos. No contexto governamental, uma análise de dados detalhada é importante para auxiliar o entendimento e a compreensão aprofundada de fenômenos complexos da administração pública. Essa análise tende a impulsionar a formulação de políticas públicas, alocar recursos de maneira eficiente e gerar impactos positivos na sociedade. Dessa maneira, este artigo explora a aplicação de técnicas de mineração de dados em contextos governamentais e em bases de dados correlatas. Através de uma revisão crítica de casos práticos extraídos da literatura, o estudo busca discutir as complexidades e desafios inerentes à mineração de dados em bases de dados governamentais. Os casos analisados fornecem uma visão abrangente das metodologias utilizadas, destacando as melhores práticas e os obstáculos comuns enfrentados no processo de extração de informações a partir de grandes volumes de dados governamentais. A investigação abrange diversos aspectos da mineração de dados, desde a coleta e limpeza de dados até a seleção e aplicação de algoritmos específicos para a interpretação dos resultados. Ao abordar esses elementos, o artigo oferece uma

compreensão das técnicas e estratégias para a aplicação da mineração de dados em contextos governamentais. Este estudo é projetado para servir como uma referência tanto para pesquisadores quanto para profissionais que operam na interseção entre a ciência de dados e a administração pública. Além disso, visa informar os processos de tomada de decisão no setor público, promovendo uma governança mais eficaz e eficiente.

**Palavras-chave:** Mineração de Dados. Bases de Dados Governamentais. Administração Pública. Decisões Políticas. Impactos Societais.

**RESUMEN**

En la era tecnológica actual, marcada por la generación masiva y continua de datos, la información debe ir más allá del simple análisis de datos en bruto. En el contexto gubernamental, un análisis de datos detallado es importante para ayudar en la comprensión y el entendimiento profundo de fenómenos complejos dentro de la administración pública. Este análisis tiende a impulsar la formulación de políticas públicas, a asignar recursos de manera eficiente y a generar impactos positivos en la sociedad. Por lo tanto, este artículo explora la aplicación de técnicas de minería de datos en contextos gubernamentales y bases de datos relacionadas. A través de una revisión crítica de casos prácticos extraídos de la literatura, el estudio busca discutir las complejidades y desafíos inherentes a la minería de datos en bases de datos gubernamentales. Los casos analizados proporcionan una visión integral de las metodologías utilizadas, destacando las mejores prácticas y los obstáculos comunes enfrentados en el proceso de extracción de información a partir de grandes volúmenes de datos gubernamentales. La investigación abarca varios aspectos de la minería de datos, desde la recopilación y limpieza de datos hasta la selección y aplicación de algoritmos específicos para la interpretación de los resultados. Al abordar estos elementos, el artículo ofrece una comprensión de las técnicas y estrategias para aplicar la minería de datos en contextos gubernamentales. Este estudio está diseñado para servir como una referencia tanto para investigadores como para profesionales que operan en la intersección entre la ciencia de datos y la administración pública. Además, tiene como objetivo informar los procesos de toma de decisiones en el sector público, promoviendo una gobernanza más eficaz y eficiente.

**Palabras clave:** Minería de Datos. Bases de Datos Gubernamentales. Administración Pública. Decisiones Políticas. Impactos Sociales.

## 1 INTRODUCTION

The rapid technological evolution characteristic of the information age has led to an expansion in the volume of data generated, reaching unprecedented peaks. Heterogeneous records from various sectors are now readily accessible,

shaping a landscape where data occupies a central strategic position in guiding decision-making processes across diverse domains. This phenomenon highlights the pivotal role of data as both a source of information and an indispensable strategic asset in today's technological context (Yoon, 2021).

The abundance of data and information spans across traditional domains like health and education, as well as critical areas such as financial transactions, governmental authorities, income distribution among state governments, and population care strategies. However, despite the apparent wealth of data, achieving a comprehensive understanding of its essence is often obscured by the intricate diversity and complexity of the aggregates (Cate, 2008).

In this context, data mining emerges as an essential catalytic element, transcending mere information compilation and entering the strategic realm. It unveils the latent potential inherent in vast government sources, offering valuable insights for analysis and strategic decision-making (Slobogin, 2008).

Effective resource allocation for public policies is crucial in Brazil, a country characterized by pronounced regional disparities. The application of advanced data mining techniques in government databases unveils gaps and distinct needs across different regions, offering substantial insights capable of guiding investment policies in a pragmatic and evidence-based manner (Silva; Rocha, 2010). Consequently, a scenario emerges where a data-empowered government gains the agility to precisely direct resources, fostering development in the most deprived areas and ameliorating entrenched historical inequalities (Victorino *et al.*, 2017). This strategic capacity for resource allocation optimizes social and economic outcomes, being a potent tool in pursuing regional equity and mitigating persistent historical disparities (Palmieri; Franklin, 2023).

From an equitable standpoint, embracing transparency at the governmental level stands as an indisputable and relevant benefit derived from data mining. The integrative methodology delineated in this study, by orchestrating the convergence, organization, and interpretation of ostensibly fragmented data, unveils patterns and relationships previously imperceptible (Amaral, 2017). This process enhances understanding of intricate social and economic dynamics. It heightens governmental accountability, giving society a

more insightful perspective on decisions and actions. The intrinsic capacity of the integrative approach to unveil complex and dynamic correlations that remain obscure is a pivotal instrument in promoting transparency within the government realm, fostering a more robust and informed understanding by citizens of administrative operations and directions. Consequently, this transparency forms an essential foundation for bolstering trust between public administration and civil society (Braga; Drummond, 2018).

This article proposes an analysis of the application of data mining techniques, focusing on elucidating the strategic importance inherent in this practice. The investigation delves into practical cases through a critical literature review. Far from merely skimming the surface of data analysis, this approach aims to transform raw data into robust instruments, solidifying them as fundamental elements for enhancing government processes and fostering sustainable development.

The structure of this work is as follows: Section 2 presents basic concepts and literature relevant to the field of data mining. Section 3 outlines the methodology employed in this study. Section 4 compares the main models, while Section 5 delves into the discussion. Finally, Section 6 presents the concluding remarks, with the primary objective of this article being to provide a comprehensive understanding of how data mining can be utilized to improve government processes and promote sustainable development.

## 2 THEORETICAL FRAMEWORK AND RELATED RESEARCH WORKS

This section lays the conceptual groundwork for comprehending the data mining process, primarily applied to governmental and proximity databases. Aligned with the contemporary landscape where effective government data management is paramount, this section delves into the intricacies inherent in employing data mining techniques within these contexts. It addresses the interplay between governmental and proximity databases to deepen understanding of the underlying complexities, stressing the significance of extracting meaningful insights from these heterogeneous sources. The relevant

literature is scrutinized to discern the methodologies, themes, chronologies, and visualization strategies employed in previous studies.

The initial data mining phase involves mining content that is available on the web. Content mining encompasses scrutinizing statistics and searching manuscripts, images, and graphics within websites, as highlighted by Nugroho, Mahendra and Indra (2021) and Varnagar *et al.* (2013). This process assesses the relevance of information on web pages for specific searches. Once the clustering of sites is complete, a structured mining process is employed for scanning. This scan yields actionable results, and the input from search queries is archived. The results are predicated on the application of appropriate data mining techniques to identify the most pertinent content (Cate, 2008). Furthermore, the gathered statistics are utilized to aid in the development of adaptive web pages, products, and support services (Silva; Rocha, 2010).

In Braga and Drummond (2018), the authors employ a descriptive mining approach to extract information from open government databases. Within the algorithmic context, the study effectively applies analytical techniques to unveil relevant patterns and characteristics within government datasets. Although the specific algorithms adopted are not explicitly detailed, they are presumed to encompass mining methods such as descriptive statistics, clustering, and data visualization. Notably, the article highlights the strategic selection of the R tool, a programming language and statistical environment, which is a central instrument for the exploratory analysis of government data. Lastly, the authors aim to extract information from government databases, fostering a deeper and more transparent understanding of available information. However, it would be beneficial to include details regarding the specific methods employed, the algorithms implemented, and the results obtained for a more comprehensive evaluation.

From a proximity perspective, Amaral (2017) proposes an investigation utilizing data mining techniques within governmental datasets to analyze the profile of Brazil's voters. The study uses an analytical approach to explore patterns, correlations, and relevant characteristics within electoral datasets. Although the article offers a specific view on the analysis of electoral data, it would

enhance clarity to provide more comprehensive insights into the methodologies used and the outcomes achieved.

Silva and Ralha (2010) utilize data mining techniques to detect cartels in bidding, employing standard analysis algorithms such as clustering, grouping, and anomaly detection. K-Means is adopted to identify behavioral patterns among bid participants, while Isolation Forests are applied to pinpoint non-standard behaviors indicative of cartelized practices. Statistical analysis and data visualization aid in comprehending the distribution and relationships between variables. The results include the identification of suspicious patterns of cartelized behavior, the indication of frequent participants in bids, and the detection of anomalies warranting further investigation.

Lastly, Gomide *et al.* (2023) offer an innovative perspective by employing data mining techniques to analyze public expenditure in municipalities in Minas Gerais, aiming to identify potential fraud alerts. The study utilizes One-Class SVM as the anomaly detection algorithm. To address the complexity of municipal financial data, pre-processing techniques such as normalization, outlier treatment, and handling of missing values are applied. Integration of data from diverse sources, including budgets, bids, and payrolls, proves crucial for identifying expected responses. Expected results encompass identifying anomalous transactions and indicating areas of risk that deserve further investigation.

## 3 METHODOLOGY

In the quest for a comprehensive understanding of data mining within government databases, data collection and an exploration of the most pertinent practices identified in current literature formed the cornerstone of this endeavor. As such, this article adheres to a systematic mapping approach designed to delineate and synthesize the overarching systematic literature review processes employed in various studies.

## 3.1 RESEARCH QUESTIONS

Building upon the strategies delineated by Curcio *et al.* (2018), the primary aim of this systematic mapping study was to elucidate the approaches to web data mining within governmental contexts. This objective was subdivided into four research questions, as outlined below:

Q-1: What are the main methods for conducting web data mining?

Q-2: What topics are covered in web data mining in government contexts?

Q-3: When and where were the studies published?

Q-4: How were the studies conducted in terms of visualizing the results?

## 3.2 LITERATURE REVIEW

The databases ABI/INFORM, DSL, and Google Scholar were used to identify and select primary research studies. These databases harbor an extensive repository of peer-reviewed research articles within the realm of data mining in government databases, offering practical functionalities such as date filters, similar topic filters, citation filters for specific articles, and the option to include or exclude patents. Searches within these databases were conducted using key phrases such as “data mining in government databases,” “data mining techniques,” and “mining methods for government data.” For instance, when searching for “data mining techniques” in the Google Scholar database, approximately 155,000 correlated studies were available.

## 3.3 SELECTION OF STUDIES

The research studies were selected based on various criteria pertaining to the database’s characteristics. These criteria include the following: application of database search filters, exclusion of articles published before the year 2010, application of full-text search filters, inclusion of articles submitted to the peer review process, inclusion of works available in Portuguese or English, exclusion of works with fewer than 5 citations, exclusion of works unrelated to the

governmental context or associated fields. The quality assessment was predicated on the robustness of the cited material, with articles lacking citations being excluded.

The inclusion criteria applied to each database were as follows:

- studies that present the methods and results of data mining in government databases;
- studies published between 2010 and 2024;
- studies conducted specifically in the field of data mining.

The exclusion criteria applied to databases are:

- studies that present incomplete or abbreviated text;
- studies that have not undergone peer review;
- studies not presented in Portuguese or English.

## 4 RESULTS

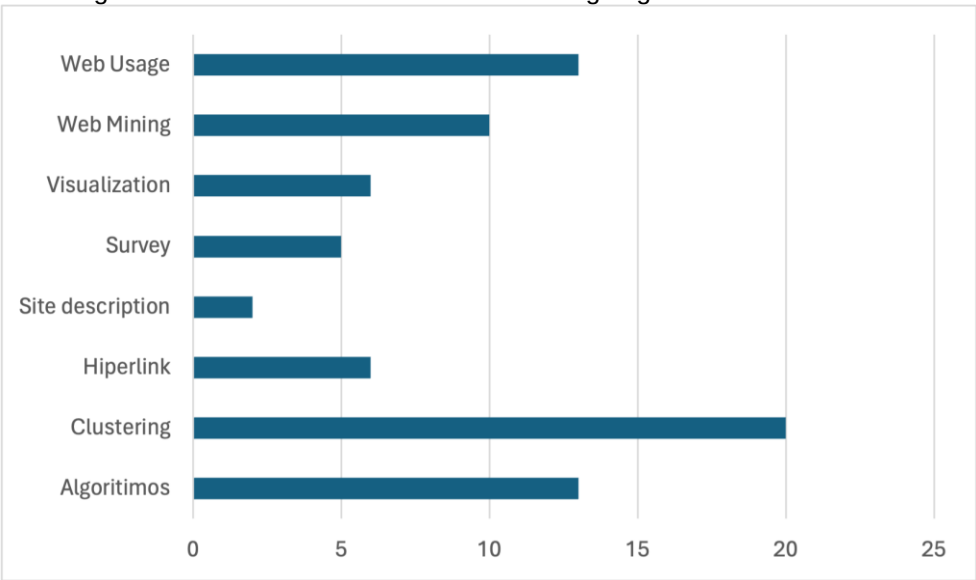
The results section of this article is dedicated to the analysis and response to the four questions raised in section 3.1, which are focused on fundamental aspects of web data mining in government contexts and on proximity bases. After applying the filters, we have 76 papers for analysis. Each topic has been investigated to provide an in-depth understanding of the main methods employed in web data mining.

### 4.1 Q-1: WHAT ARE THE MAIN METHODS FOR CONDUCTING WEB DATA MINING?

The field of web data mining encompasses a wide array of methods and techniques, highlighting a diversity of approaches as identified in related studies. Through analysis on Google Scholar, the primary techniques encompass algorithms and clustering methods, statistics, web mining, web usage mining, and visualization technology techniques. Meanwhile, the random data collection approach unveils a plethora of methods, including algorithms, clustering, data mining, data pre-processing, analysis of hyperlink networks, multivariate

statistics, quantitative analysis of hyperlink structure, statistical data visualization methods, web content mining, and web structure mining. Further details can be found in Figure 1. This notable diversity of methods underscores the intricate complexity and breadth of approaches employed in the web data mining landscape. It reflects the ongoing quest for innovative and adaptable strategies in response to the escalating complexity of information available on the web.

Figure 1. Main methods used for data mining in government databases




Source: prepared by the authors.

## 4.2 Q-2: WHAT TOPICS ARE COVERED IN WEB DATA MINING IN GOVERNMENT CONTEXTS?

The approach taken to classify topics in this research was rooted in the existing body of knowledge on web data mining, with a specific aim to provide a targeted representation of activities within this domain, particularly in governmental contexts. While the research primarily focused on web mining, it comprehensively addressed all related activities pertinent to the study's scope. The detailed mapping effectively pinpointed significant research gaps, shedding light on areas necessitating further investigation to comprehensively understand the governmental data mining landscape.

These categories encompass fundamental techniques such as “data mining,” which involves extracting useful information and patterns from government-related data sets. “Web usage” pertains to web analysis, while “weblog” entails exploring and analyzing records directly sourced from the web. “Clustering” involves categorizing government data into groups, and “hyperlinking” investigates links between information across various web pages. Additionally, “fuzzy” refers to applying fuzzy logic in data mining, while “government search” involves specific search techniques tailored to government contexts. These topics span from foundational approaches to more advanced techniques, offering a comprehensive and stratified perspective on research within the field.

It is important to note that the research was not limited solely to web mining but inclusively addressed all related activities, encompassing aspects such as web structure mining, web content, and analysis of hyperlink networks, among others. This holistic approach enables a deeper and more contextualized understanding of the challenges and opportunities inherent in governmental data mining, underscoring the significance of adopting a multidimensional approach to explore the complexity of this research domain.

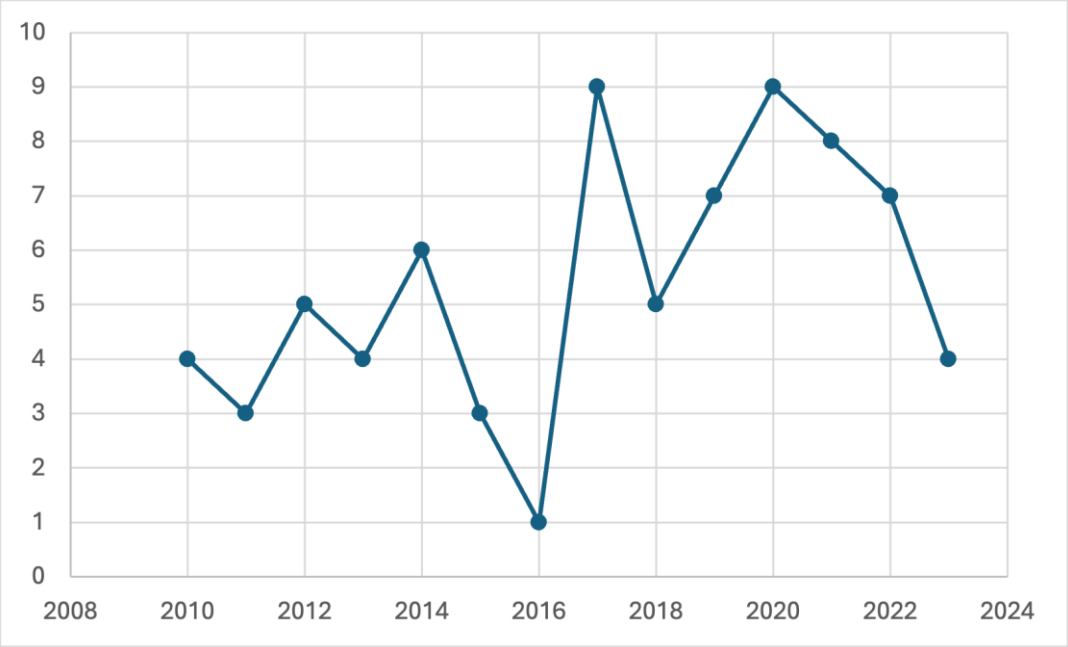
## 4.3 Q-3: WHEN AND WHERE WERE THE STUDIES PUBLISHED?

A consistent growth trend in research interest was observed when conducting a longitudinal analysis of publication frequency spanning the period from 2010 to 2024 within the realm of web data mining. However, a detailed examination unveiled a significant one-time decline in 2016, followed by gradual reductions in 2018 and 2023, hinting at potential shifts in research agendas or the intrinsic dynamics of the field. Figure 2 offers a compelling visualization of this temporal pattern in publications.

As for publication venues, the analysis encompassed various formats, ranging from conference presentations to journal articles and workshop participation. The diversity of platforms underscores a wide and collaborative

dissemination of knowledge within this specific domain, indicative of an engaged scientific community receptive to information exchange.

Figure 2. Number of publications per year from 2010 to 2023 related to data mining in government databases




Source: prepared by the authors.

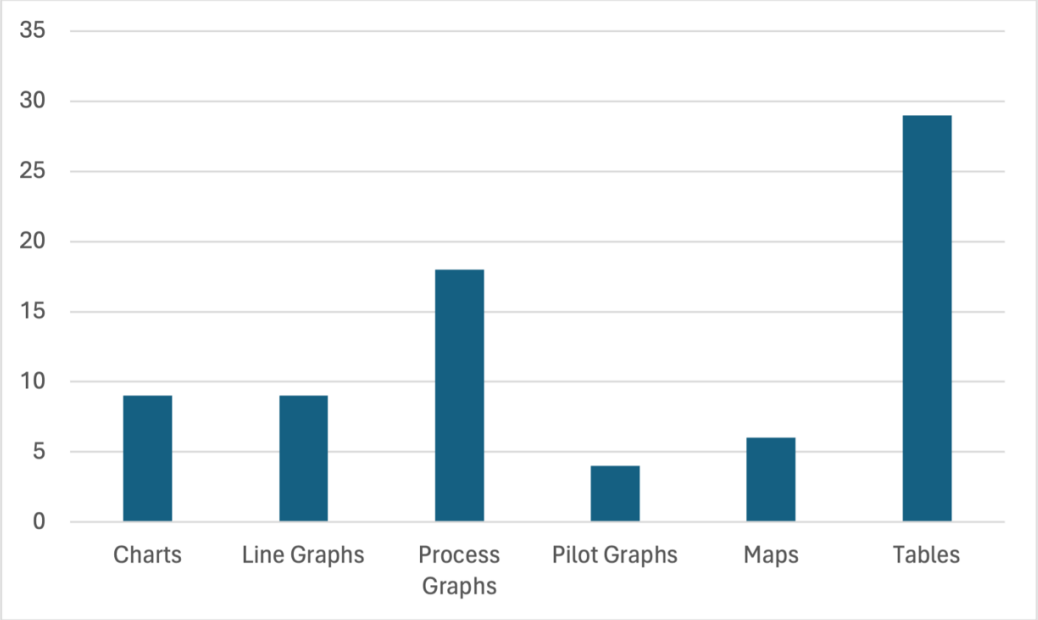
## 4.4 Q-4: HOW WERE THE STUDIES CONDUCTED IN TERMS OF VISUALIZING THE RESULTS?

The exploration of visualization methods uncovered diverse approaches, with graphs and tables emerging as the predominant tools for effectively representing the relative amounts of data in each category, as depicted in Figure 3. This conventional approach is complemented by alternative forms of visual representation, including maps, charts, process graphs, line graphs, and pilot graphs. Each of these techniques possesses unique characteristics:

1. maps: enable the geospatial representation of data, facilitating the comprehension of geographic patterns and regional distribution;
2. tables: present data in a tabular structure, organized systematically to enhance accessibility and understanding during analysis;

3. charts: include various visual formats, such as bar and line graphs, to illustrate trends, patterns, and relationships within the data;
4. process graphs: visualization describing sequential steps or processes, offering a chronological and hierarchical data perspective;
5. pilot graphs: provide a general or panoramic view of the data, enabling swift identification of relevant aspects;
6. line graphs: demonstrate the variation of one or more variables over time, offering insights into trends and fluctuations.

Figure 3. Strategies for presenting results in studies related to data mining in government databases




Source: prepared by the authors.

## 5 DISCUSSION

This research delves into the application of data mining optimization techniques within the context of heterogeneous databases, aiming to alleviate data redundancies through grouping techniques in web mining, with a particular focus on the governmental context (Silva; Sirqueira, 2022). The study analyzes cluster optimization strategies, stressing the importance of accessing graphical representations within web records for identifying symmetrical and asymmetrical

behaviors in public management. Pattern analysis, a crucial component of this approach, is conducted through previously refined web records (Silva; Ralha, 2010).

The algorithms adopted and evaluated included PageRank, Weighted PageRank, and HITS (Jain; Purohit, 2011). Information derived from conventional movement patterns was explored using a web mining tool. The significance of web mining in the governmental sphere is highlighted, where the classification of compatible and appropriate pages is conducted through several specialized algorithms. PageRank, Weighted PageRank, and HITS stand out for offering uniform treatment to all links, thereby concurrently sharing the classification score (Sharma *et al*., 2020).

The discussion encompasses both levels of the predictive model aimed at enhancing assertiveness indices. The first level endeavors to filter categories with a high probability of attracting an audience of interest – individuals related to the government capacity to provide enriching information for the analysis. The second level focuses specifically on web pages with the highest probability and available content. However, the acknowledge is an intrinsic challenge posed by heterogeneous user behavior within the government (Colak; Sagiroglu; Yesilbudak, 2012).

To bolster the personalization of web content, the review of government web records is conducted through advanced data mining tools (Masseglia *et al.,* 2000). This approach entails a meticulous analysis of document utility from the user's perspective, assigning higher scores to clusters that match or resemble specified keywords. Unlike prevailing systems limited to content-based searches, this approach is deemed more effective in terms of specificity within the government context (Varnagar *et al.,* 2013). Additionally, the proposition to automatically identify reference points and events in marked images is noteworthy, underscoring the relevance of these detections in analyzing and aligning extensive sets of images shared by users connected to measures and implementations originating from government resources (Allan, 2002).

From a similar perspective, hyperlinks interconnect different web pages, constituting an invaluable source of information and playing a significant role in government data mining. Consequently, this research scrutinizes the structure of

government websites, incorporating webometrics processes to uncover latent information in hyperlinks (Vanti, 2002).

Graph mining, a vital element in data mining, is explored in the government context for identifying repetitive subgraphs, highlighting the application in the efficient derivation of candidate subgraphs. This study analyzes the processes periodically adopted in this technique, addressing specific challenges related to the complexity of government data. Additionally, Finite State Machine (FSM) algorithms are evaluated regarding search strategies and approaches for counting frequencies during categorization processes applied to government data (Szostak; Wlodarszyk; Walkowiak, 2021).

This research considers blogs as web-based social presentations in government communication. A comprehensive model is presented for evaluating innovations in the design and content of government blogs, contributing to an in-depth understanding of the role of these tools in the dissemination and communication of government action. This thorough analysis is essential for understanding the dynamics of government information in the digital era (Agerdal-Hjermind; Valentini, 2015).

As an essential application for the in-depth understanding of support needs in online programs and services, web mining constitutes a field of study in constant evolution. This approach, aimed at identifying usage patterns in web data, unfolds into three distinct and crucial phases: pre-processing, pattern discovery, and pattern analysis. This scientific overview also includes reviewing studies linked to this rapidly expanding technique (Zhang *et al*., 2024).

In the contemporary landscape, web mining of government databases emerges as a critical facet. The data from this practice becomes foundational for gaining deeper insights into general web usage patterns within the governmental sphere. Beyond enhancing the comprehension of available information, this technique stands out for its enhanced applicability to end users in the public sector. The escalating governmental interest in this analysis underscores its pivotal role in the technological panorama. However, the inherent challenges of web mining in government contexts pose questions requiring substantive answers before developing robust tools (Nugroho; Mahendra; Indra, 2021).

The implementation of data mining in educational systems, for example, with a specific focus on online courses, learning content management systems, and web-based adaptive and intellectual educational systems, constitutes an area of investigation of great relevance for directing strategic government decisions. This specific domain demands unique requirements for effectively applying data mining, presenting different complexities from other sectors. The insertion of recommender agents in e-learning systems, aimed at searching and compiling educational resources on the web, is a crucial aspect of this process. To achieve a level of success comparable to other spheres, educational data mining emerges as an analysis that demands deeper specialization and a more targeted focus within the educational context (Yanrong, 2021).

In this context, the synergistic integration of government web mining and the analysis of educational data can provide a holistic understanding, enabling both efficiency in governance and the optimization of educational processes, promoting more equitable and sustainable development.

## 6 CONCLUSION

This research on data mining in electronic media underscores the critical importance of conducting in-depth analysis in light of today's abundant available information. It reaffirms that compiling data for information analysis extends beyond mere data scrutiny, particularly within the governmental context, where structured, case-oriented approaches are essential for guiding policies, efficiently allocating resources, and fostering positive impacts. Reflecting on the implications, the findings of this study highlight how data mining can serve as a powerful tool to enhance societal welfare by enabling more informed and effective decision-making processes. The insights gained can help policymakers better understand user behaviors, optimize resource distribution, and improve operational efficiencies, thereby contributing to more responsive and transparent governance.

This research also offers substantial benefits to the academic community by providing a comprehensive framework for future studies that aim to explore

the intersection of data science and public administration. The emphasis on an integrative approach that considers both intermediary and governmental factors in defining the roles of individuals within government service sectors is particularly noteworthy. Such an approach leveraging data mining techniques to analyze costs, user behavior, and operational efficiency within government databases.

However, this study is not without its limitations. The primary constraints include the scope of data sources examined and the potential biases inherent in the selected case studies. Additionally, the rapidly evolving nature of data mining technologies means that continual updates and refinements to methodologies are necessary to stay current with best practices.

For future research, it is recommended to expand the range of data sources and to include longitudinal studies that track changes over time. Additionally, exploring the ethical implications of data mining in governmental contexts and developing frameworks to mitigate potential privacy concerns would be valuable. By addressing these areas, subsequent studies can build on the foundation laid by this research, further enhancing the effectiveness and efficiency of government services through advanced data analysis techniques.

## REFERENCES

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. (Ed.). **Topic detection and tracking: event-based information organization**. New York: Springer Science & Business Media, 2002.

AMARAL, B. C. **Análise do perfil de eleitores brasileiros com mineração de dados abertos governamentais.** Universidade Federal Fluminense, Instituto de Computação, 2017.

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

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

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

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.

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

MASSEGLIA, F.; PONCELET, P.; CICCHETTI, R. An efficient algorithm for web usage mining. **Networking and Information Systems Journal**, v. 2, n. 5/6, 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.

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**. 2023.

SHARMA, P. S. *et al.* A systematic review on page ranking algorithms. **International Journal of Information Technology**, v. 12, n. 2, 2020.

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.

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

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.

SLOBOGIN, C. Government data mining and the fourth amendment. **The University of Chicago Law Review**, v. 75, n. 1, 2008.

SZOSTAK, D.; WŁODARCZYK, A.; WALKOWIAK, K. Machine learning classification and regression approaches for optical network traffic prediction. **Electronics**, v. 10, n. 13, 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, 2002.

VARNAGAR, C. R. *et al*. Web usage mining: a review on process, methods and techniques. *In:* **2013 International Conference on Information Communication and Embedded Systems**. IEEE, 2013.

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**. IEEE, 2021.

YOON, Y. Discovering knowledge in corporate databases. *In:* **Handbook of Data Management**. Boca Raton: Auerbach Publications, 2021.

ZHANG, J. *et al.* New Horizons in Web Search, Web Data Mining, and Web-Based Applications. **Applied Sciences**, v. 14, n. 2, 2024.