IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 20 November 2024, accepted 9 January 2025, date of publication 20 January 2025, date of current version 24 January 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3531834

SURVEY

The Use of Data Mining in Public Budgeting: A Systematic Literature Mapping

JOSÉ CLÁUDIO GUEDES DAS NEVES [1,2] AND VERONICA OLIVEIRA DE CARVALHO [1]
[1]Departamento de Estatística, Matemática Aplicada e Computação, Universidade Estadual Paulista (UNESP), São Paulo 01049-010, Brazil
[2]Secretaria de Orçamento Federal (SOF), São Paulo 01049-010, Brazil

Corresponding author: Veronica Oliveira de Carvalho (veronica.carvalho@unesp.br)

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) – Finance Code 001.

ABSTRACT Planning and allocating public resources is essential because resources are always limited and must be sufficient to meet a country's needs. Therefore, it is necessary to define how resources are distributed based on the amount collected, directly affecting society in the most diverse areas, such as education and health. Owing to advances in artificial intelligence in recent years, studies have been conducted to explore and propose intelligent solutions that enable the most diverse analyses in this critical area. Among these, data mining has emerged as a viable solution. Generally, data mining consists of three major steps: pre-processing, pattern extraction, and post-processing. Thus, to understand how data mining has been used in the most diverse subjects related to public planning and budgeting, this study presents systematic literature mapping. The aims were (i) to provide an overview of the aspects related to the data mining steps in the presented context and, (ii) to identify gaps that can be addressed and/or explored. The results are presented and discussed throughout this paper based on 30 papers selected over 10 years (from 2014 to 2023), with the potential to significantly impact future research and practice in public planning and data mining.

INDEX TERMS Public budget, data mining, artificial intelligence, systematic literature mapping.

I. INTRODUCTION

THE budget is a fundamental means of exerting political influence through efficient and optimized planning, management, and control of public services or activities. Therefore, it requires a complex and dynamic set of tasks to be performed by the public administration [1]; thus, the financing decisions taken during the budget preparation process are of great importance and, have a substantial impact on society [2], [3].

Given that this is one of the most critical problems of the government, since public resources are always limited and must be sufficient to meet the needs of a country, this process is usually reflected in the structure of the budget law initiative based on the following questions: "How much money should be raised?" and, "How should these resources be distributed?" [4]. Considering the main instrument of fiscal policy, the information contained in budget laws is of fundamental importance for fiscal transparency [5], given that governments need to be responsible and transparent in their public spending decisions in order to avoid losses due to fraud, corruption and money laundering [6] - around 15% of global GDP [7] - as well as to build healthy and sustainable economies [8].

As a critical component of fiscal policies, financing decisions are essential to an economy's ability to allocate resources among various economic sectors [9]. These decisions are not only about the allocation of funds but also about the perspective of a policymaker who aims to maximize social well-being, often using available data to make decisions that will lead to the best possible outcomes. Thus, social welfare assumes a certain similarity with other public finance problems, in which the main trade-off lies between the weighted sum of private services and public revenues [10].

The goal of a public budget is to promote the quality of public spending, making it more productive and effective.

The associate editor coordinating the review of this manuscript and approving it for publication was Senthil Kumar.

Moreover, it is essential to improve the cost-benefit ratio to support priority programs, such as education and health. To achieve this, governments conduct a careful analysis of resources and strategic planning for future spending [11]. The public budget can be conceptualized as a planning instrument adopted by the public administration, which estimates the revenues to be collected and sets the expenses incurred in the following financial year. The aim is to ensure continuity, effectiveness, efficiency, and economy of the quality of services provided to society. It is also characterized by a continuous, dynamic, and flexible process that translates the government's work plans and programs into financial terms for a given period [12].

Given the relevance of public budgeting to society, it is essential to use solutions that support public administrators in making better decisions. With the advancement of artificial intelligence (AI), recent studies have explored and proposed intelligent solutions that enable diverse analyses in this critical economic area and, consequently, for society. Among these possibilities, data mining has emerged as a viable solution for modeling and understanding several existing budgetary subjects. Since strategic anticipation is one of the practices recommended by the OECD (Organisation for Economic Co-operation and Development), creating working groups focused on using AI techniques in the public sector has become imperative. In fact, in this year 2024, at the "46th Annual Meeting of the Committee of Senior Budget Officials", the subject "Using Artificial Intelligence in Public Financial Management" was addressed.[1] The OECD itself makes available in its digital library several recent papers on the power of AI and its implications for economies and societies.[2] Therefore, it is essential to understand how the solutions are structured.

Considering the above, this work presents a systematic literature mapping (SLM) that was carried out (i) to provide an overview of the aspects related to the data mining steps in the presented context and, (ii) to identify gaps that can be addressed and/or explored. Therefore, this study has the potential to significantly impact future research and practice in public planning and data mining. It is worth mentioning that many of the studies in the literature using data mining, machine learning, or deep neural networks, contextualize the solutions as AI solutions. However, AI encompasses several other areas such as computer vision, robotics, and fuzzy logic. Thus, in this study, AI refers to the data mining context.

The paper is structured as follows: Section II presents the concepts related to the data mining process; Section III presents the protocol used to perform the SLM; Section IV presents the results, as well as the analyses and discussions for each research question. Finally, Section V concludes the paper and discusses gaps that can be addressed and explored.

[1]The document is available at https://one.oecd.org/document/GOV/SBO(2024)14/en/pdf

[2]https://www.oecd-ilibrary.org/science-and-technology/oecd-artificial-intelligence-papers_dee339a8-en

II. THE DATA MINING PROCESS

Currently, a large amount of data is constantly generated and stored and is available in a variety of formats, such as text, video, and audio. To search for implicit patterns in data, a process consisting of steps must be executed. This process is known as data mining. Thus, data mining is characterized as an iterative and interactive process through which patterns (knowledge) are obtained [13], [14]. Several methodologies, including KDD, CRISP-DM and SEMMA, enable the execution of this process [15].

Fig. 1 presents all general steps of the data mining process. We believe that the steps described below, as in [13], are those that commonly appear in most methodologies. Therefore, the SLM was performed considering these steps, which are briefly described below based on [13]. At the end of the process, the knowledge and/or patterns extracted can be incorporated into an intelligent system.

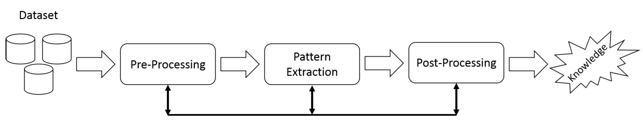


FIGURE 1. General flow of a data mining process.

In the **pre-processing** step the raw input data are transformed into a format suitable for subsequent analysis. According to [13], it is the most laborious and time-consuming step. As a number of different aspects can be considered here, they can be grouped in different ways. The categorization below is adopted in [13], namely:

- Sampling: the aim is to select a subset of instances[3] from the dataset to reduce, for example, the time required to process all data. To this end, there are many approaches, such as random sampling, stratified sampling, and progressive sampling.
- Dimensionality reduction: the aim is to reduce the number of features[4] by creating new attributes that are a combination of old attributes. To this end, many techniques have been developed, such as PCA and t-SNE.
- Feature selection: the aim is to reduce the number of features by selecting a subset of attributes that describe the instances. To this end, many techniques exist that can be grouped into three approaches: embedded, filter and wrapper.
- Discretization, binarization and encoding: the aim here may be to convert: (i) a continuous numerical feature into a categorical one (discretization); (ii) data into a binary format, where the values are transformed into "0" or "1"; and (iii) a categorical feature into a numerical feature (encoding). To this end, many techniques exist for each, such as equal width and equal frequency for discretization, threshold and binary

[3]The words instance and sample are used synonymously.

[4]The words feature, attribute and variable are used synonymously.

encoding for binarization, and one-hot and label encoding for encoding.

- Attribute (variable) transformation: the aim is to transform all the values of a variable by applying (i) a simple mathematical function (AT: Simple Functions), such as log or sqrt, or (ii) a normalization (such as z-score) or standardization (such as min-max) technique (AT: Normalization/Standardization).

Some other aspects are still found in [13], but were not listed above because they were not addressed in the papers selected in the SLM. Other important aspects such as removal of outliers, imputation of missing values, and treatment of duplicate data were considered by [13] as aspects of data quality, which must be evaluated before the pre-processing step. However, in this work, all these aspects are considered to belong to the pre-processing step. In these cases, more general data treatments are classified as data cleaning.

In the **pattern extraction** step, patterns are obtained using machine learning algorithms. To this end, algorithms are selected, set (with/without hyperparameter tuning), and executed. To perform this step, the main task must have already been defined before the start of the process, that is, classification, regression, association, clustering, among others. The aim of predictive tasks (classification, regression) is to induce a model capable of predicting the value of a target variable (dependent) from a set of explanatory variables (independent) [13]. On the other hand, descriptive tasks (association, clustering) aim to derive patterns that summarize the underlying relationships in the data [13]. For each type of task, a range of algorithm families is available, which must be selected depending, among other factors, on the problem in question.

Finally, in the **post-processing** step the obtained patterns are validated. This validation can be in terms of performance metrics (accuracy, mean squared error, etc.) or interpretability (through the model itself[5] or XAI techniques [16], [17]). To ensure a good estimate of performance metrics and, enable the configuration of algorithm hyperparameters, different validation strategies can be used, such as holdout and cross-validation.

As previously mentioned, several types of data are available today, including temporal data. A time series is a set of points observed over time. Thus, depending on the type of data involved in the pattern analysis, there are variations in the term data mining (DM), to explain the type of data involved in the process, such as temporal data mining (TDM), for time series data, text mining (TM), for textual data, and stream mining (SM), for stream data. Because economic time series constitute a valuable source of information in the context of public budgeting, several of the selected papers work with this type of data. Although, in general, the term DM is used when data are stored in tabular format, in this work the terms DM and TDM are used as synonyms because of the type of application addressed.

One of the tasks that stands out in relation to temporal data is forecasting, which aims to estimate future values based on the past values of the series. To this end, two approaches can be used: parametric approaches, based on statistical methods, in which knowledge about the nature of the data distribution is assumed, such as ARIMA, and non-parametric approaches, based on machine learning methods, in which the data distribution need not be known. In economics the use of econometric (statistical) methods is common, because it is a more consolidated area. [18] present an excellent survey of the traditional statistical methods. Thus, some of the papers selected in SLM use such methods as baselines to evaluate the results obtained when machine learning methods are used.

When a forecasting task is modeled via supervised machine learning (ML), it is necessary to include attributes that incorporate the temporal aspect, since ML algorithms assume that the pairs (X_i, y_i) are independent and identically distributed (i.i.d.). Thus, the series must be transformed into the format (X_i, y_i) considering, in some way, temporal aspects. This is generally performed through a sliding window procedure, as seen in the lines of Table 1 (in the example the window size was set to two). It is observed that the values obtained in t_1, t_2 and t_3, together, represent the first instance of the set (X_1, y_1) (line $t = 4$), t_2, t_3 and t_4 the second (X_2, y_2) (line $t = 5$), and so on. Note that to describe each point at instant t_i, two new attributes referring to its temporal aspect were included, i.e., referring to the last two events, in this case, t_{i-1} and t_{i-2}. Thus, once the series have been converted to tabular format (X_i, y_i), they can be modeled using ML algorithms, now described by a set of instances and a set of attributes.

The transformation presented above is categorized as a pre-processing step. In [19] four categories of temporal attributes are presented: date-time, lag (sliding window, Table 1, where window size defines the number of lags), rolling window statistics, and expanding window statistics. Other pre-processing aspects related to temporal data, which can also be applied to tabular data, are the imputation of missing values, as well as normalization or standardization of data [19].

Finally, it is worth mentioning another aspect of pre-processing that is not covered in [13] and [19]: data augmentation. When the dataset size is small, it is necessary, in some cases, to create new instances from the initially existing ones. Augmented datasets tend to improve model generalization and avoid problems such as overfitting. In this work, techniques for tabular data, such as SMOTE [20], and for temporal data, such as interpolation via upsampling [21] or scaling and jittering [22], are categorized as data augmentation techniques.

III. PROTOCOL

Systematic Literature Mapping (SLM) [23] is a process in which a set of studies, available in the literature, is analyzed based on a research question. The aim is to provide an overview

[5] White-box algorithms produce interpretable models, i.e., models that can be understood by humans, such as decision trees [16].

TABLE 1. Example of converting a 10-point time series $Z = (z_1, z_2, \ldots, z_{10})$ to tabular format via a sliding window of size 2.

Time Series					Tabular Format (X_i, y_i)
		X_1	X_2	X_3	y
t	Z	$z_{(t-2)}$	$z_{(t-1)}$	$z_{(t)}$	$z_{(t+1)}$
1	2.5	NA	NA	NA	2.5
2	2.1	NA	NA	2.5	2.1
3	2.3	NA	2.5	2.1	2.3
4	2.2	2.5	2.1	2.3	2.2
5	3.2	2.1	2.3	2.2	3.2
6	4.0	2.3	2.2	3.2	4.0
7	3.0	2.2	3.2	4.0	3.0
8	3.4	3.2	4.0	3.0	3.4
9	2.8	4.0	3.0	3.4	2.8
10	2.0	3.0	3.4	2.8	2.0

of the state of the art through the presentation and discussion of results considering the analyses carried out in relevant studies. To this end, a protocol is elaborated, which contains the following steps: (a) formulation of one or more research questions (Section III-A); (b) identification of the primary studies to be considered (for this purpose, the studies must be extracted and analyzed) (Section III-B); (c) extraction and synthesis of data (Section III-C); and (d) summary and discussion of the results (Section IV). Regarding this SLM it is worth mentioning that:

- this SLM focuses much more on computational aspects than on economic aspects, as the aim is to analyze the data mining process itself in this application domain;
- no other secondary study was found on this subject. Only a few studies that are partially related to the theme addressed here were identified [24], [25], [26];
- this SLM only considered studies that use tabular data or time series data, because the main difference between them, when using a data mining process, occurs in the pre-processing step regarding temporal features. Works that use other types of data (text [5], [27], [28], for example) were not considered because the pre-processing step becomes relatively different, making it difficult to summarize the results presented in Section IV (in fact, not only the pre-processing step, but the others as well, since other concepts can be addressed, such as language models).

A. RESEARCH QUESTIONS

The aim of this SLM was to retrieve and analyze primary studies that use data mining in the context of public planning and budgeting to understand how the process occurs, from data collection to the validation of the extracted patterns. To this end, the following questions were formulated.

RQ1. What are the main budgetary subjects explored? This question aims to identify the main subjects in which data mining techniques have been applied. To this end, the division proposed in [29] is used.

RQ2. What data mining methodologies have been used? Is it possible to access the implementations used to guarantee the reproducibility of experiments? This question aims to identify whether the works follow any data mining methodology and whether it is possible to guarantee the reproducibility of the results by making implementations available in repositories such as GitHub.[6]

RQ3. What are the characteristics of the datasets used? This question aims to identify the characteristics of the datasets used in the studies, namely: number of instances, number of features, extraction period, data extraction source, and number of countries considered in the analyses.

RQ4. What type of pre-processing has been applied to the samples? Pre-processing techniques are addressed in this question with the aim of verifying how data preparation has been carried out, as this affects the performance of the models.

RQ5. Which tasks and their respective algorithms are used? Several algorithms can be used to induce models; therefore, it is important to identify the most widely used ones in this domain, relating them to existing data mining tasks (classification, regression, etc.).

RQ6. How have the patterns obtained been evaluated? This question aims not only to identify the performance metrics used, but also the validation strategies, as well as the use of interpretability techniques.

B. IDENTIFICATION OF THE PRIMARY STUDIES

To identify the primary studies relevant to data extraction, it is necessary to define the search string, the databases for retrieving the papers, the inclusion and exclusion criteria to select or not a paper as relevant and the steps to make the selection.

1) SEARCH STRING

The search string was formulated with the aim of covering the issues of "public budget" (first part) and "data mining" (second part): *("public budget*" OR "public spend*" OR "public expen*" OR "public account*" OR "public financ*" OR "government budget*" OR "government spend*" OR "government expen*" OR "government account*" OR "government financ*" OR "national budget*" OR "national spend*" OR "national expen*" OR "national account*" OR "national financ*") AND ("data mining" OR "machine learning" OR "deep learning" OR "artificial intelligence").*

[6] https://github.com/

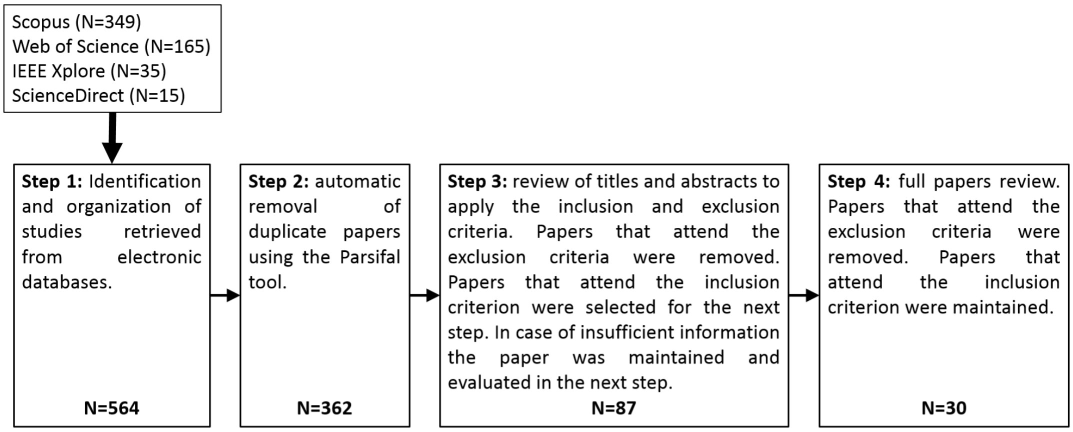


FIGURE 2. Selection steps.

To formulate this string we evaluated the words frequently used in different works, as well as their synonyms, and then we checked the papers retrieved with such string, in the selected databases, to calibrate it. The words "machine learning", "deep learning" and "artificial intelligence" were used as synonyms for "data mining" as the first two are general terms related to the pattern extraction step and the last one related to the large area where data mining is inserted (it is currently the term used to name the most diverse data science solutions).

2) SOURCE SELECTION

The search string was applied only to electronic databases, making the necessary adjustments to the syntax of each one. The following electronic databases were considered: Scopus,[7] ISI Web of Science,[8] IEEE Xplore[9] and ScienceDirect.[10] The string was applied to titles, abstracts and keywords. The period considered in the search was from 01/01/2014 to 31/12/2023 (10 years).[11]

3) INCLUSION AND EXCLUSION CRITERIA

The purpose of defining these criteria is to identify primary studies that provide direct evidence in relation to the research questions. Therefore, the studies to be selected for data extraction are those that do not meet any exclusion criteria. To this end, the following exclusion criteria were considered: (i) paper out of scope, as it does not address the topic of data mining in the context of the public budget[12]; (ii) the paper addresses the topic of data mining in the context of the public budget, but uses other types of data, such as textual data (as already mentioned, only papers related to tabular and temporal data were considered); (iii) the paper addresses the use of tools and/or software and not the data mining process itself; (iv) the paper is very general, covering topics such as system description, for example; (v) it is not possible to extract information from the paper in a clear manner; (vi) the paper is a copy or version of another paper already considered; (vii) the paper is not a primary study (grey literature, such as proceedings, technical report, etc.); (viii) the paper is not written in English; (ix) could not access the full paper.

4) SELECTION STEPS

To assist the process we used the Parsifal[13] tool and Microsoft Excel. Fig. 2 shows the steps used to select the papers, which are described in the figure itself. The values initially obtained in the searches, as well as the values obtained at each step, are also presented in the figure. A total of 30 papers were selected.

C. DATA EXTRACTION AND SYNTHESIS

Data extraction is concerned with collecting information from selected papers to answer research questions. Data extraction was performed by reading the selected papers. However, it is also important to extract and organize more general data from these papers. The extraction forms were built using Microsoft Excel.

IV. RESULTS AND ANALYSIS

Before presenting the analyses related to the research questions, more general data on the papers is presented. It is important to mention that all of the information below is available here.[14] Thus, it is possible to verify, for each of the following aspects, which papers (ID) make up a given item and/or category. For example, it is possible to determine which papers used a certain algorithm by consulting the respective link.

[7]www.scopus.com
[8]http://apps.webofknowledge.com/
[9]http://ieeexplore.ieee.org
[10]http://www.sciencedirect.com
[11]2024 was not considered as it has not yet been finalized (current year of extraction).
[12]Causal inference papers [30], [31], [32], [33] were considered out of scope because their primary aim was not to obtain and evaluate a model per se. Furthermore, purely econometric papers [34], [35] were also considered out of scope because they did not use ML methods to induce the models.
[13]https://parsif.al
[14]https://bit.ly/3Csxy5K

TABLE 2. Selected papers.

Paper ID	Reference/Title
p1	[36] A machine learning model of national competitiveness with regional statistics of public expenditure
p2	[37] Application of data mining technology in financial intervention based on data fusion information entropy
p3	[38] Application of deep learning to forecast the South African unemployment rate: A multivariate approach
p4	[39] Application of Support Vector Machines and Holt-winters model in local finance forecast
p5	[7] Assessing the performance of open contracting in Colombia through data mining
p6	[40] Boosting tax revenues with mixed-frequency data in the aftermath of COVID-19: The case of New York
p7	[41] Comparison and application of Logistic Regression and Support Vector Machine in tax forecasting
p8	[42] Data-driven approach for predicting and explaining the risk of long-term unemployment
p9	[43] Enhancing cash management using machine learning
p10	[44] Evaluation of public procurement efficiency of the EU countries using preference learning TOPSIS method
p11	[45] Forecasting unemployment in the euro area with machine learning
p12	[46] Forecasting with many predictors using bayesian additive regression trees
p13	[47] From e-budgeting to smart budgeting: Exploring the potential of artificial intelligence in government decision-making for resource allocation
p14	[48] Impact of COVID-19 on G20 countries: Analysis of economic recession using data mining approaches
p15	[49] Point break: Using machine learning to uncover a critical mass in women's representation
p16	[50] Prediction of out-of-pocket health expenditures in Rwanda using machine learning techniques
p17	[9] Prioritization of public expenditure for a better return on social development: A data mining approach
p18	[4] Public budget simulations with machine learning and synthetic data: Some challenges and lessons from the Mexican case
p19	[51] Tax default prediction using feature transformation-based machine learning
p20	[52] The nexus between financial development and economic growth: An application of the VECM model and machine learning algorithms
p21	[53] The optimized k-means clustering algorithms to analyzed the budget revenue expenditure in Padang
p22	[11] Use of machine learning and multilevel analysis in hierarchical approaches of public expenditure forecasting
p23	[54] Clustering and classification analysis in financial reporting of Philippine government business enterprises
p24	[55] Budgeting for SDGs: Quantitative methods to assess the potential impacts of public expenditure
p25	[56] Analysis factors affecting Egyptian inflation based on machine learning algorithms
p26	[57] Analysis of judiciary expenditure and productivity using machine learning techniques
p27	[58] The forecasting of a leading country's government expenditure using a recurrent neural network with a gated recurrent unit
p28	[59] Modelling the non-linear dependencies between government expenditures and shadow economy using data-driven approaches
p29	[60] Customs revenues prediction using ensemble methods (statistical modelling vs machine learning)
p30	[61] Machine learning for economic modeling: An application to South Africa's public expenditures

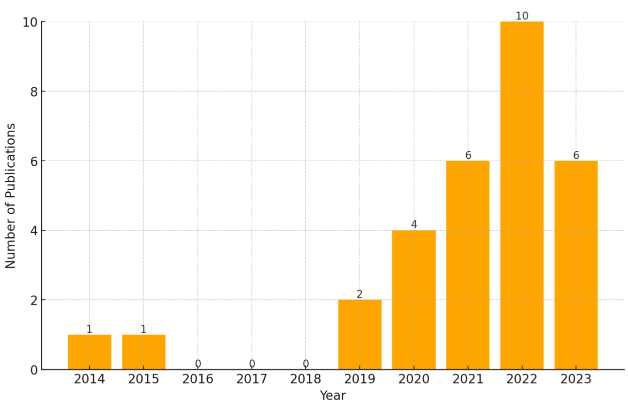


FIGURE 3. Number of publications over the years.

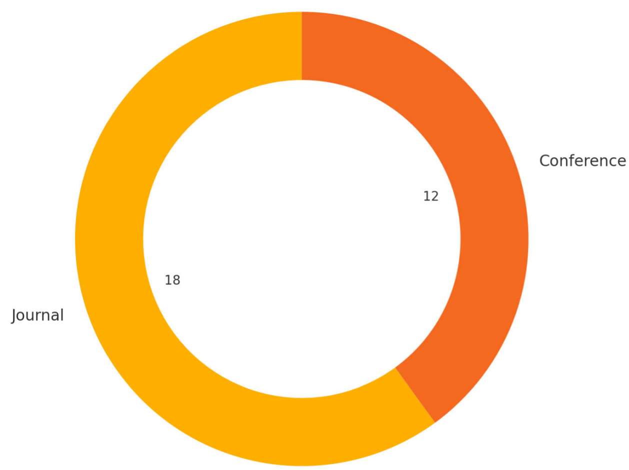


FIGURE 4. Distribution of publication types.

A. OVERVIEW OF THE STUDIES

This section presents some general aspects of the selected papers, which are listed in Table 2.

1) PUBLICATION YEAR

As already mentioned, the period considered was from 01/01/2014 to 31/12/2023 (10 years). We can observe through Fig. 3 an increase in the number of publications since 2019. 93.33% (28 studies)[15] of the papers were published from this year onwards, with 2022 being the year with the most

[15]We will always try to present relative and absolute values using "()" or ",".

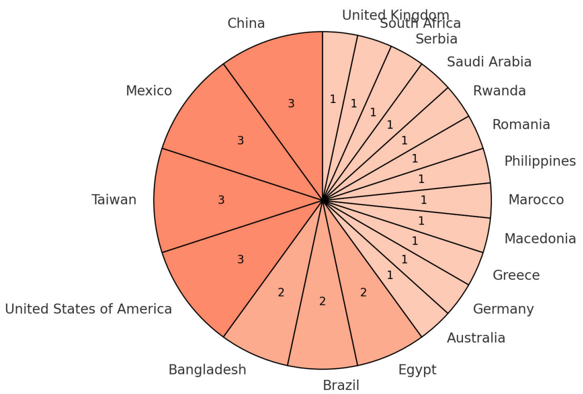


FIGURE 5. Number of publications by country.

publications (10, 33.33%). It can be noticed that interest in the presented context has been growing, even though the number of papers is not high.

2) PAPER TYPE

In Fig. 4 it is possible to notice that most papers were found in journals (18, 60%), while the rest (12, 40%) appeared in conferences.

3) GEOGRAPHICAL DISPERSION

Fig. 5 shows the dispersion of papers by country, considering the country of the first author. It can be noted that the public budget is an important government planning tool throughout the world.

B. RQ1. WHAT ARE THE MAIN BUDGETARY SUBJECTS EXPLORED?

The aim of this question was to identify, in a generic way, considering the logical division proposed in [29], in which budgetary subjects, themes and subthemes, data mining techniques have been applied.

The logical division comprises three elements: (i) the logical whole, constituted, in this case, by budgetary matters; (ii) the basis or fundamental principle of the division, expressed by its metaphysical aspect, that is, the point of view from which the division is made, presented, in this case, as an essential division of the artificial genus (budgetary themes) into artificial species (budgetary subthemes); and (iii) the dividing members or species resulting from the logical division, in this case, the budget subthemes presented. Although the rules of applied logical division are observed, namely: (a) using only one principle, (b) the constituent species being mutually exclusive, and (c) the division being collectively exhaustive, it is understood that this division, like all positive divisions, is based on empirical knowledge, demanding frequent revisions given that subsequent investigations may prove that previous conclusions were incomplete, inadequate or misleading [62]. It is also important to emphasize that the studies were classified subjectively, considering the best understanding of the main subject addressed in the works, thus not constituting a definitive classification, given the possibility of categorizing the same study into different categories.

The public budget comprises several interrelated subjects that are essential for managing public policies. Fig. 6 presents the identified subjects: "Budget process" and "The State in the economy". The figure presents the themes addressed in each subject (x-axis) and their respective subthemes (y-axis). The subjects are represented (differentiated) by the colors of the graph.

The subject "Budget process" was divided into the following themes: (i) "Budget and financial execution", (ii) "Control and evaluation of budget execution", (iii) "Discussion, voting, and approval of the budget law" and (iv) "Preparation of the budget proposal". "The State in the economy" was divided into: (i) "Economic responsibilities of the State" and (ii) "The growth of public expenditure".

Belonging to the subject "Budget process", the subtheme "Preparation of fiscal studies and projections" (13, 43.33%) consists of one of the stages of the process of "Preparation of the budget proposal", in which fiscal studies and projections are carried out to support the definition and validation of budgetary limits. The subtheme "Execution of expenditure" (3, 10%), related to the theme "Budget and financial execution", within the scope of public administration, particularly with the acquisition of goods and the contracting of services and works, depends on compliance with bidding rules, a formal administrative procedure that ensures equal competition among competitors and seeks the most advantageous proposal for public administration [12], [29]. The theme "Control and evaluation of budget execution" aims to verify the administration's integrity, the safekeeping and legal use of public funds and compliance with the Budget Law, with the Legislative Branch being the holder of "External control" (subtheme) (1, 3.33%). Finally, there is the subtheme "Amendments" (1, 3.33%), referring to parliamentary amendments related to the theme "Discussion, Voting and Approval of the Budget Law".

Connected with the subject "The State in the Economy", the subthemes related to fiscal functions (budgetary functions), referring to the theme "Economic responsibilities of the State", are divided into [63]: (i) "Allocative Function" (4, 13.33%), to promote adjustments in the allocation of resources; (ii) "Distributive Function", to promote adjustments in the distribution of income; (iii) "Stabilizing Function" (5, 16.67%), to maintain economic stability, having among its main macroeconomic objectives the maintenance of a high level of employment and stability in price levels [29]. Several doctrinal currents seek to explain the "Reasons for the growth of public expenditure" (3, 10%), a subtheme related to the theme "The growth of public expenditure". According to [64], neoclassical and Keynesian interpretations are based on the law of continuous growth of state activities, attributed to the German economist Adolf Wagner, in which the relative participation of the government in the economy grows with the country's own pace of economic growth, constituting the

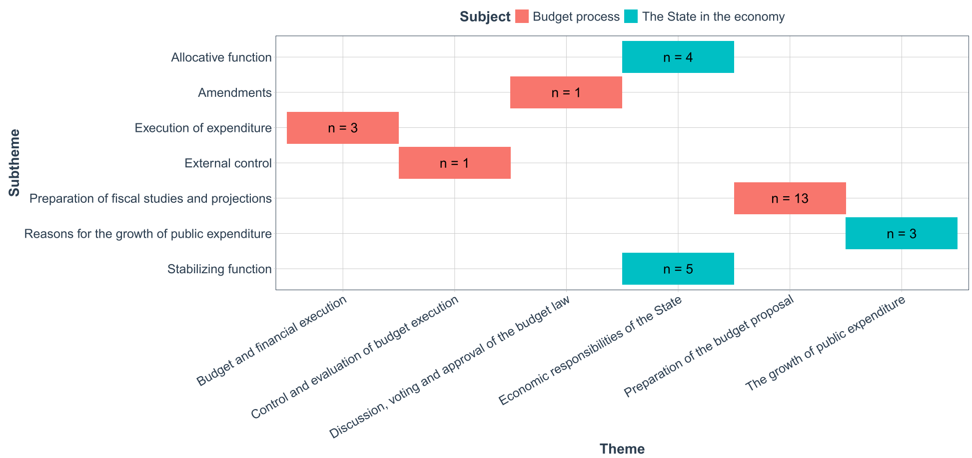


FIGURE 6. Budgeting issues explored divided into themes and subthemes.

determining causes pointed out by Richard Bird: (i) the growth of administrative and security functions; (ii) the growing demands for greater social welfare, especially education and health; and (iii) the more significant direct and indirect intervention of the government in the production process.

Given the above, there is a diversity of subjects in which the most varied data mining processes, techniques, and algorithms are applied, seeking to improve the efficiency of government planning in its different budgetary matters. Although other budgetary issues are of great importance, such as "Budget classifications", it is clear that only some studies are using computational techniques to effectively improve the quality of public spending based on evidence-based analysis. In this sense, academia needs to conduct studies to improve public expenditure efficiency based on budgetary, economic, and social parameters to commit to the quality of public policies.

C. RQ2. WHAT DATA MINING METHODOLOGIES HAVE BEEN USED? IS IT POSSIBLE TO ACCESS THE IMPLEMENTATIONS USED TO GUARANTEE THE REPRODUCIBILITY OF EXPERIMENTS?

The aim of this question was to identify whether the works follow any data mining methodology and whether the papers make their implementations available in a repository such as GitHub to reproduce the results.

Regarding the use of methodologies, only 2 (6.67%) works specify the methodology adopted, namely: CRISP-DM (p5) and SEMMA (p17). The absence of a formal methodology implies an ad-hoc strategy for the data mining process. Although this is not an issue per se, the formal and explicit use of existing methodologies can help in the analysis, reproducibility and understanding of the limitations of an investigation, as it has already been tested in different applications or even domains.

Regarding reproducibility, only 1 (3.33%) (p29) work provides the code used to perform the experiments described in the paper. However, even in this case, the Excel spreadsheet used to load the data is unavailable in the repository. This fact, along with other issues raised in the next sections, makes reproducibility difficult (more on reproducibility in machine learning see [65]). Access to implementations would contribute to, for example, (i) analyze the feasibility of using the proposals in other countries, which could support administrative decision-making; (ii) compare results already obtained with results generated from other models proposed to solve the same type of problem.

D. RQ3. WHAT ARE THE CHARACTERISTICS OF THE DATASETS USED?

The aim of this question was to identify the characteristics of the datasets used in the studies, namely: number of instances, number of features, extraction period, data extraction source and number of countries involved in the analyses.

Fig. 7 shows the characteristics of the datasets in relation to the number of instances considered. For a better understanding, the number of instances was grouped into ranges. Initially, it is worth mentioning the difficulty in obtaining this information: in 8 (26.67%) papers it was not possible to identify this data (label "ND" (Not Defined)).[16] Furthermore, even in the papers in which extraction was possible, in some of them the information was not accurate and, therefore, in each interval, it is shown how many works of the total fell into this situation (label "Doubts"). It is noted that among the works in which extraction was possible (22, 73.33%), 8 (36.36%) used a relatively small sample size ([1;100[). Furthermore, most of these works (13, 59.09%) reached, at most, sets with up to 300 instances (ranges [1;100[+ [100;200[+ [200;300[).

Fig. 8 shows the characteristics of the datasets in relation to the number of features considered.[17] For better understanding, the number of features was also grouped into ranges. Obtaining this information was also difficult, albeit to a lesser extent. It is

[16]When it was not possible to extract a certain piece of information, or when the information was unavailable, the label "ND" (Not Defined) was used.

[17]When possible, the number of features reported is the one obtained after the pre-processing step.

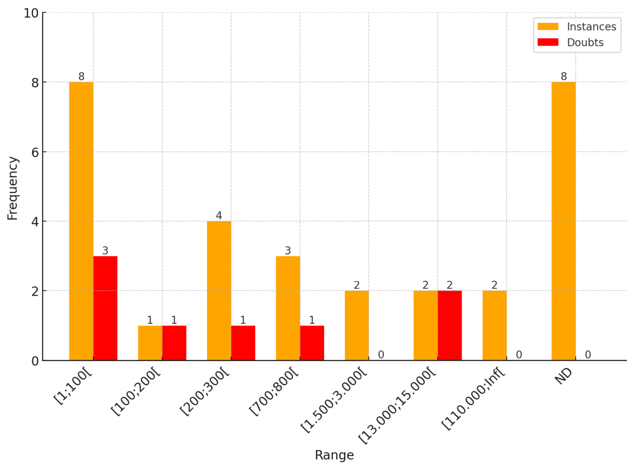


FIGURE 7. Number of instances grouped by ranges.

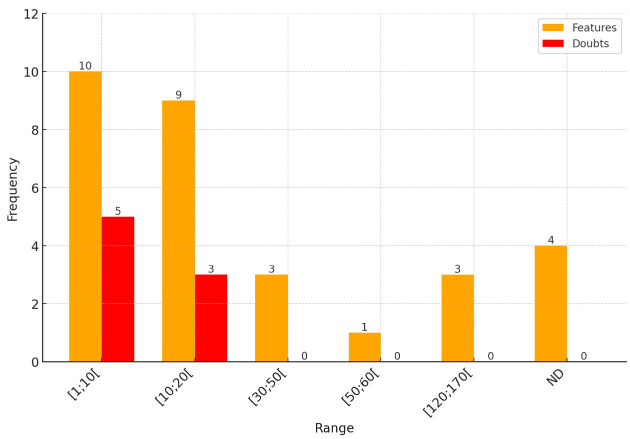


FIGURE 8. Number of features grouped by ranges.

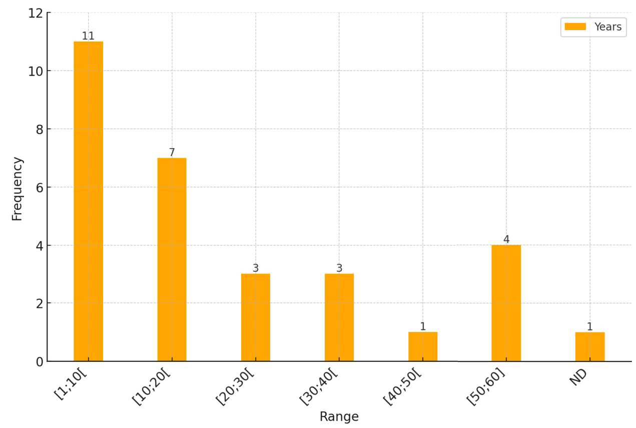


FIGURE 9. Number of years grouped by ranges.

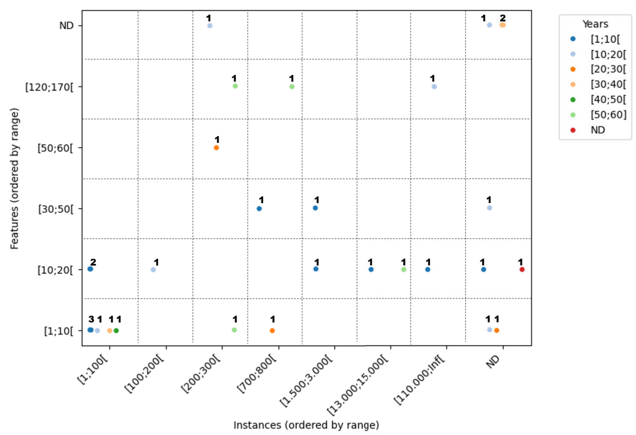


FIGURE 10. Distribution of papers in relation to the last three characteristics: number of instances, number of attributes and number of years.

noted that among the works in which extraction was possible (26, 86.67%), 10 (38.46%) used up to 10 features ([1;10[). Furthermore, most of these works (19, 73.08%) reached, at most, sets with up to 20 features (ranges [1;10[+ [10;20[).

Fig. 9 shows the characteristics of the datasets in relation to the extraction period. For better understanding, the number of years was also grouped into ranges. It is noted that among the works in which extraction was possible (29, 96.67%), 11 (37.93%) considered a period of up to 10 years ([1;10[). Furthermore, most of these works (18, 62.07%) considered a period of up to 20 years (intervals [1;10[+ [10;20[). In this case, the period is not expected to be long, since many data related to the public budget are annual (e.g. GDP). This explains, at least in part, why datasets contain a relatively small number of instances.

Fig. 10 relates the three previous graphs into a single graph. It is possible to notice a cluster of 6 works in the grid [1;100[x [1:10[, corroborating the results presented above. In this cluster the works vary over time, with most referring to a period of up to 10 years ([1:10[).

Finally, it is worth mentioning two more characteristics of the datasets: in 28 (93.33%) papers (exceptions p2, p7) the data explored were obtained from official internet websites (e.g. World Bank). Furthermore, in 22 (73.33%) papers (exceptions p10, p11, p13, p14, p15, p17, p27, p28) the data explored involved data from only 1 country.

Given the results presented above, it follows that:

- the difficulty in identifying the characteristics of the datasets prevents an understanding of how the data mining process was applied;
- different from econometric methods, the sample size, when applying data mining processes, impacts generalization [13], [66]. Although it is understandable, to some extent, why the samples are small, since much of the data related to public budgeting are annual, strategies could be used to solve this problem, such as data augmentation (see Section IV-E). Only 2 (6.67%) papers used this strategy (p8, p18);
- although the data explored by the papers are, for the most part, open access (e.g. World Bank), the authors do not make available the data they downloaded, which makes it difficult to know exactly which data were downloaded and how they were manipulated (see Sections IV-C and IV-E).

E. RQ4. WHAT TYPE OF PRE-PROCESSING HAS BEEN APPLIED TO THE SAMPLES?

The aim of this question was to identify the pre-processing techniques used in the selected papers. Data often needs to be pre-processed not only to improve the quality of the data itself, because this step impacts the performance of the models, but also to adapt to a standard demanded by a specific data mining task [13].

Fig. 11 presents the identified techniques. Although this step is one of the most important in the data mining process [13], 46.67% (14) of the papers do not specify the type of pre-processing performed (label "ND"). The identified techniques are shown in the figure, namely: attribute transformation (AT) through normalization/standarlization (9, 30%), missing values (7, 23.33%), feature selection (5, 16.67%), encoding (3, 10%), attribute transformation (AT) through simple functions (2, 6.66%), data augmentation (2, 6.66%), data cleaning (1, 3.33%), dimensionality reduction (1, 3.33%) and sampling (1, 3.33%). In relation to some works related to time series, there is a conversion of the series to a tabular format via a sliding window, in which the lags are defined manually, via hyperparameter tuning, among others. As seen in the figure, 10 (33.33%) papers performed such conversion. However, since conversion is tied to certain tasks, more details are presented in the next section (Section IV-F).

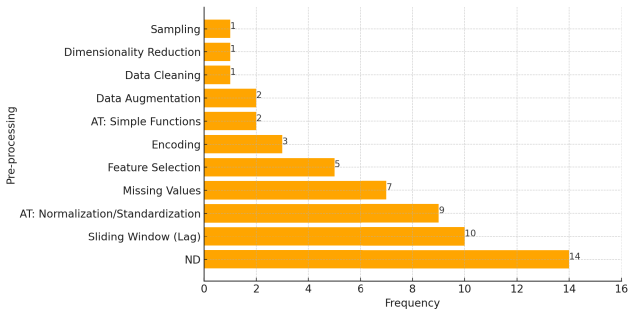


FIGURE 11. Pre-processing techniques.

Given the results presented above, it follows that:

- this step was considered so trivial that its details were not presented (46.67%). The reason for the lack of information may have occurred, in some cases, because of the characteristics of the algorithms used, since according to [67] some algorithms already have implicit pre-processing steps. However, it is necessary to describe/discuss in detail each step of the process in order to ensure not only understanding of the work, but also its reproducibility;
- the pre-processing techniques identified are, in general, those commonly used, with emphasis on normalization/ standardization (9, 56.25%), missing values (7, 43.75%) and feature selection (5, 31.25%) (in this case, the absolute values were divided by 16, the number of papers that mention the techniques used). Data augmentation is a technique that deviates from the general standard;
- for data augmentation, two different techniques were used: oversampling (p8, SMOTE) and interpolation via upsampling (p18). Since samples related to public budgeting are generally small, it was expected that more studies would use techniques like these. As observed in the previous section, the sample size that stands out in the works is between [1;100[(see Fig. 7 and 10). This point could be better explored, as in [4], [68]. According to [69] there are three ways to deal with the problem of short time series (generally common in the economic area): (i) choosing simple models; (ii) combining several models (ensemble); and (iii) extending the dataset via synthetic samples. On the other hand, according to [70], the three ways are: (i) increasing (expanding) the available data; (ii) using a selection approach to choose the best forecasting model among several models; and (iii) adjusting the parameters of a given forecasting model to obtain the highest possible accuracy. In the selected papers there are no explanations regarding the choice of algorithms to try to solve the problem presented here.

F. RQ5. WHICH TASKS AND THEIR RESPECTIVE ALGORITHMS ARE USED?

The aim of this question was to identify the algorithms used, relating them to existing data mining tasks (classification, regression, etc.).

Fig. 12 initially presents the proportion of works in each of the identified tasks, namely: classification with 20% (6), regression with 73.33% (22) and others with 6.67% (2). Each task is represented by 3 bars: the first indicates the total number of papers that used it, the second the papers that used only that task and the last, a complement to the previous one, the papers that used it in addition to other tasks and/or techniques. It is noted that the task that stands out the most is regression.

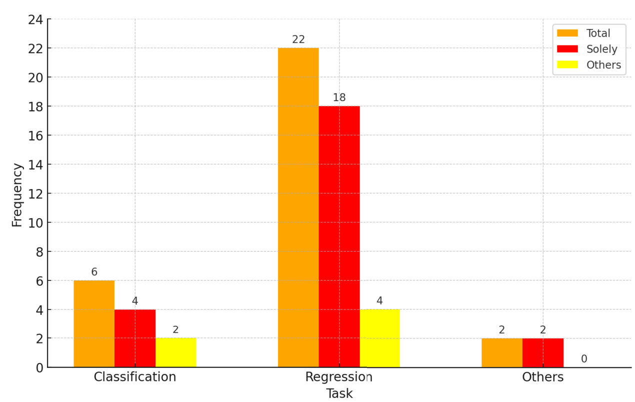


FIGURE 12. Data mining tasks.

Since a forecasting task can be modeled as a regression task, in Fig. 12 the count refers to both actual regression tasks and forecasting tasks. However, as they have different aims, Fig. 13 presents the proportion of each: of the 22 works, 14 (63.63%) refer to forecasting and 8 (36.36%) refer to regression. While forecasting aims to predict future values based on temporal patterns, regression does not. In other words, forecasting uses time series data, where each observation depends on the previous ones (the data are ordered and time dependent). On the other hand, regression does not consider time dependence and assumes that observations are independent (see Section II). Therefore, the analyses below are divided into the following tasks: classification, regression, forecasting and others.

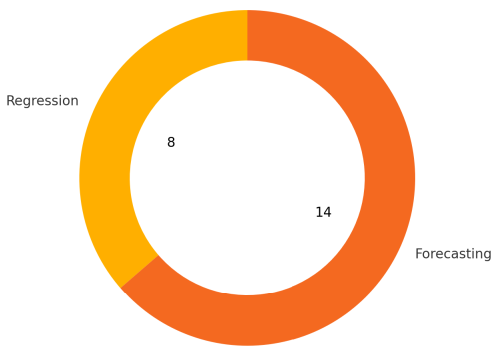


FIGURE 13. **Regression task breakdown (distinction between regression and forecasting).**

Since the forecasting task was the one that stood out the most, the results related to it are initially presented. A first point analyzed in relation to this task is whether, in fact, the papers related to this task were treating the data as a time series. Fig. 14 presents the results: of the 14 works, 8 (57.14%) treat the data as a time series (label "TS") (p3, p6, p12, p18, p22, p27, p29, p30), 3 (21.43%) treat them as non-temporal data (label "Not TS") (p7, p9, p21) and in 3 (21.43%) it is not possible to extract such information (label "ND") (p4, p20, p25). Thus, a certain inconsistency can be noted in works p7, p9 and p21. In these works the authors perform a regression task, although they mention that it is a forecasting task; however, they do not work with temporal variables (lags)[18] (see Section II). Therefore, there is some confusion regarding the concepts used. This was an important point, as all 14 papers should work with temporal variables. Regarding works p4, p20 and p25, they were classified in the "ND" category because they used econometric methods as baselines, or some other representation that made reference to time series analysis, even without it being possible to identify the use of lags (a deduction was made from certain elements of the text). Finally, of the 8 works that treat the data as time series, in 6 (75%) (p3, p6, p12, p18, p29, p30) the authors inform, even if briefly, how the lags were defined via a sliding window (manually, via hyperparameter tuning, etc.) and in 2 (25%) (p22, p27) it is not possible to identify how they were defined.

[18]It was decided to keep them in the forecasting category since this is the way in which the authors describe the problem and to highlight some confusion between the concepts used.

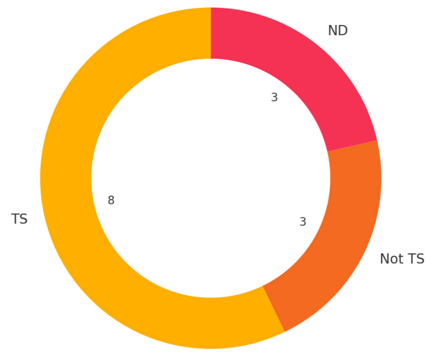


FIGURE 14. **Forecasting papers regarding the use of data as time series (TS).**

Regarding the algorithms used in the works that perform forecasting (14), it can be seen in Fig. 15[19] that the algorithms that stand out the most are Support Vector Regression (SVR) (7, 50%) and Random Forest (RF) (4, 28.57%). However, a diversity of algorithms is observed, which could be grouped into families such as those based on ensemble (RF, GB, Ens, XGBoost, AdaBoost, BART), those based on (deep) neural networks (LSTM, ANN, GRU, MLP, CNN, DAN2, ELM) and those based on linear regression (LASSO, LR, BR, MLR, RR, SG-LASSO) (acronyms in the figure). It was decided to keep the analysis at the algorithm level, and not at the family level, since a given group may stand out in relation to individual algorithms, distorting the frequency of use of certain algorithms. Note that SVR and RF together cover 11 (78.57%) works (actually 9 (64.29%), since in 2 of them both algorithms are used).

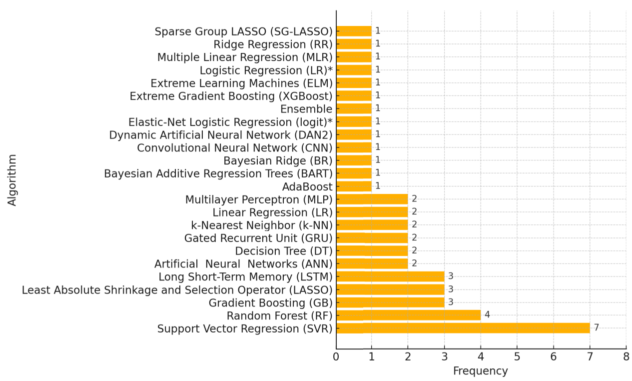


FIGURE 15. **Forecasting algorithms.**

Still in relation to the forecasting task, it is worth mentioning that of the 14 works, 9 (64.29%), 7 from the "TS" category (p3, p6, p12, p22, p27, p29, p30) and 2 from the "ND" category (p4, p20), use some econometric method to perform a comparison with the results obtained via machine learning,

[19]The "logit" and "LR" algorithms were marked with "*" because they are traditionally used in classification tasks; however, they were used in an unusual way in the forecasting task.

whose methods are presented in Fig. 16. It is noted that ARIMA was the one that stood out the most. This is an interesting aspect, since the authors were concerned with using some state-of-the-art method to validate the results, which in this domain is carried out by econometric methods. However, only 2 of them (p4, p20) present stationary tests (and/or others), such as Augmented Dickey-Fuller (ADF), which are necessary for the application of such methods, since some assumptions must be met.

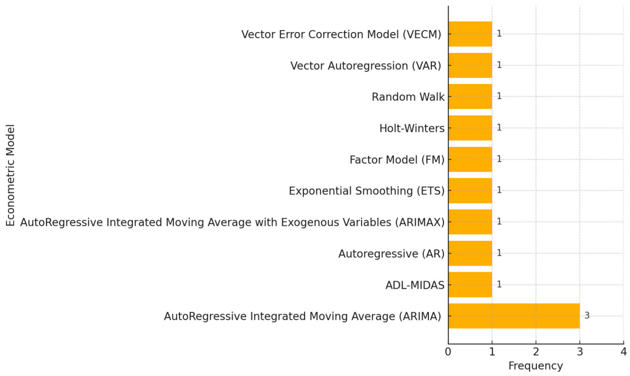


FIGURE 16. Econometric methods.

Regarding the algorithms used in the works that perform regression (8), it can be seen in Fig. 17[20] that the algorithms that stand out the most are Decision Tree (DT), Linear Regression (LR) and Random Forest (RF), all with 3 (37.5%) occurrences. The Multilayer Perceptron (MLP) also stood out with 2 (25%) occurrences. As in the forecasting task, a diversity of algorithms is observed, which could be grouped into families such as those based on ensemble (RF, XGBoost, GB, Treenet) and those based on (deep) neural networks (MLP, ANN, AN*, NN*, Perceptron) (acronyms in the figure). However, as previously, it was decided to maintain the analysis at the algorithm level for the reasons already presented.

Regarding the algorithms used in the works that perform classification (6), it can be seen in Fig. 18 that the algorithms that stand out the most are Decision Tree (DT) and Random Forest (RF), both with 4 (66.67%) occurrences, as well as Logistic Regression (LR) and Support Vector Machine (SVM), both with 3 (50%) occurrences. Extreme Gradient Boosting (XGBoost) also stood out with 2 (33.33%) occurrences. As in the above tasks, a diversity of algorithms is observed, which could also be grouped into families. However, as previously, it was decided to maintain the analysis at the algorithm level for the reasons already presented. This task was the one that presented the greatest variation in the algorithms that stood out. It is also worth mentioning that in 2 (33.33%) papers (p11, p24) the authors treated the data as time series, making it possible to identify how the lags were defined (manually,

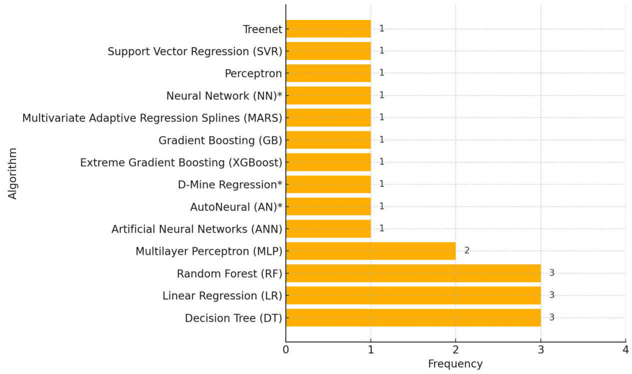


FIGURE 17. Regression algorithms.

via hyperparameter tuning, etc.). In these cases, the output is a binary variable that assumes either "0" or "1" (binary classification).

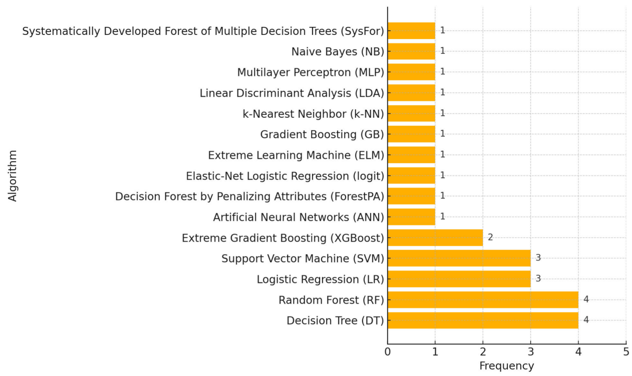


FIGURE 18. Classification algorithms.

Regarding the other tasks and/or techniques, Fig. 19 shows those used by the selected works. It is noted that the most used task was clustering (4, 50%), having been used in works that combine it with other tasks such as regression (p14, p26), forecasting (p21) and classification (p24). The clustering algorithms used were Agglomertive Hierarchical Clustering (Ward) (p14), k-means (p21, p23, p26), Self-Organizing Map (SOM) (p23) and DBSCAN (p23). The other tasks and/or techniques occurred only once (12.50%). The association task was not used in conjunction with any other task, and the algorithm used was Apriori (p5). The MCDA technique [71] was also not used in conjunction with any other task, and the method used was TOPSIS [72] (p10). Regarding the use of genetic algorithm (p13) and agents (p24), they were used in works that combine them, respectively, with regression and classification tasks.

Finally, it is worth mentioning one last aspect of the works: problem modeling. Most of the works (24, 80%) aim to predict the value of a single output variable (#O:1), be it continuous (regression, forecasting) or discrete (classification, be it binary or multiclass), as seen in Fig. 20. Only 4 works aim to minimize the error of more than one variable. However,

[20]The "AN", "D-Mine" and "NN" algorithms have been marked with "*" since they refer to specific implementations of "SAS Enterprise Miner" (https://www.sas.com/en_th/software/enterprise-miner.html), which is based on the SEMMA methodology.

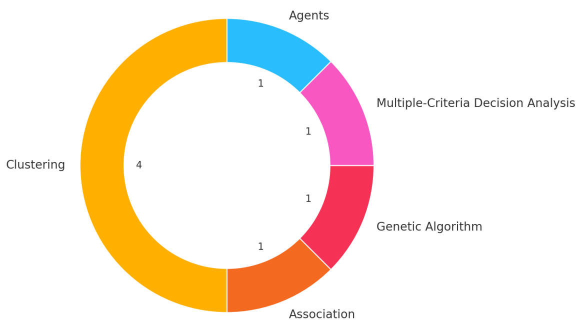


FIGURE 19. Other tasks used in the papers.

in 3 (10%) of them (p12 (#O:9), p15 (#O:3), p18 (#O:49)) a model is induced for each output variable, that is, they do not work with multi-output models, where the aim is to predict multiple dependent variables simultaneously. Only 1 (3.33%) paper (p13) focuses on minimizing the error of 3 variables simultaneously (#O:3). It was decided to keep this information in this section since the modeling of multi-output problems is dependent on the algorithm used. Some algorithms are inherently designed to work with only one output variable, such as SVM and SVR, while others are not, such as DT, RF, and MLP. The 2 works labeled with "ND" are the two that used other tasks and/or techniques in isolation (p5, p10) (see above).

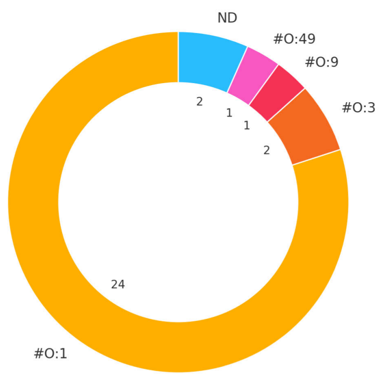


FIGURE 20. Number of output (#O) variables.

Given the results presented above, it follows that:

- the task that stands out the most is forecasting. When it is used, the authors are concerned with using some econometric method as a baseline;
- there were few works that did not use classification, regression and/or forecasting tasks (2, 6.67%). Thus, depending on how the problem is modeled, other possibilities are still plausible to be explored;
- there is a certain confusion in some works between regression and forecasting tasks, in which the authors perform regression but refer to it as forecasting. However, to work with forecasting it is necessary to treat the data as time series and, therefore, define how the lags will be incorporated into the process, i.e., how the lags will be defined. Therefore, it is necessary for the works to describe in more detail how the modeling was carried out so that the process is truly understood and the results validated;
- although there is no standard regarding the algorithms used, some stood out in the tasks presented, namely: (i) forecasting, SVR and RF; (ii) regression, DT, LR and RF; and (iii) classification, DT, RF, LR and SVM. It is noted that RF appears in all tasks, DT in 2, LR in 2 and support vector in 2 (SVR+SVM). They are all traditional and consolidated algorithms in the area of machine learning;
- although there is currently a trend towards using deep neural network models, including for temporal data, they have been little used. This is probably because of the sample size, which is generally small for these models. Thus, as seen above, the works opted to use more traditional algorithms;
- There are few works that perform multi-output modeling. However, several budgetary problems can be designed in a multi-objective manner, so that different modeling possibilities can be explored.

G. RQ6. HOW HAVE THE PATTERNS OBTAINED BEEN EVALUATED?

The aim of this question was to identify not only the performance metrics used, but also the validation strategies, as well as the use of interpretability techniques.

As in the previous question, the results are presented by task. Regarding the metrics, Fig. 21, 22 and 23 show, respectively, the metrics used in the forecasting (14), regression (8) and classification (6) tasks. The most used metrics in forecasting were MAE (7, 50%), MAPE (6, 42.86%), RMSE (6, 42.86%) and MSE (5, 35.72%). In regression, MSE, R^2 (R Squared), RMSE and SSE appear with the same frequency (2, 25%). Finally, in classification, Accuracy (6, 100%), F1-Score (4, 66.67%) and AUC (3, 50%). There is an intersection between forecasting and regression metrics, since, as already mentioned, the forecasting task can be modeled as a regression task. Both the MSE and RMSE stand out in these tasks. Only 2 works do not specify the metric used: 1 for forecasting (p21) and 1 for regression (p14), since they also use clustering, giving a different focus to the analyses. In this case, it follows that: (i) in relation to forecasting tasks, more specific metrics could also be used, such as MASE [73]; (ii) in relation to classification tasks, it is necessary for the works to inform under which curve the AUC metric is computed (ROC or other), so that the results are interpreted correctly. It is also worth mentioning in relation to the forecasting task that 7 (50%) works present, in addition to performance measures, forecasting plots (p4, p6, p7, p22, p25, p27, p29). This graphical representation is very interesting to support the understanding of the behavior of predictions, complementing the analysis of results.

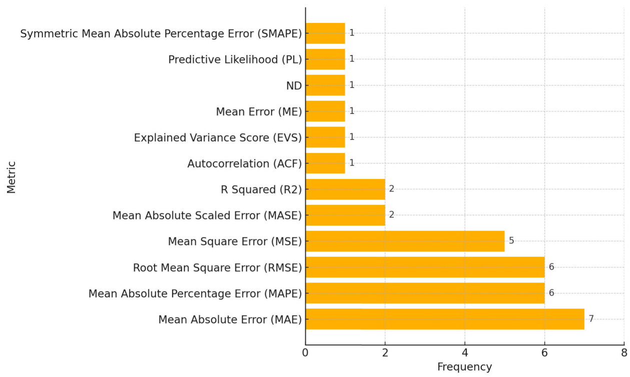


FIGURE 21. Forecasting metrics.

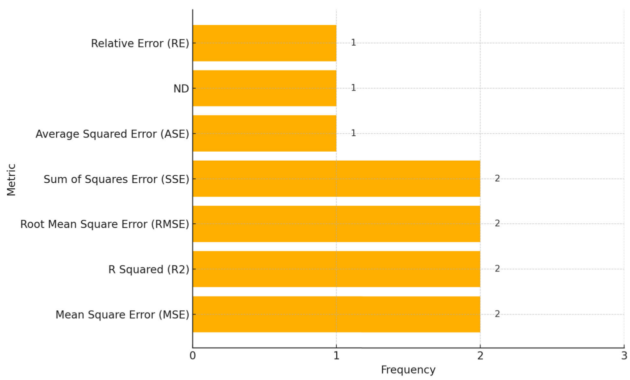


FIGURE 22. Regression metrics.

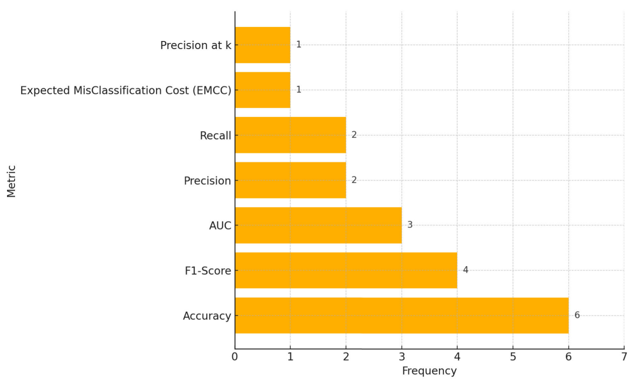


FIGURE 23. Classification metrics.

Regarding validation strategies, the works mostly use holdout, namely:

- of the 14 forecasting works, 10 (71.43%) use holdout and in 4 (28.57%) the strategy is not specified;
- of the 8 regression works, 5 (62.50%) use holdout and in 3 (37.50%) the strategy is not specified;
- of the 6 classification works, all (100%) use holdout.

The data above is noteworthy, as other means of validation exist. In fact, some works used cross-validation, not to estimate metrics, but rather to perform hyperparameter tuning (only the one used to estimate the metrics was considered here). However, when working with time series, a wide range of methods exist, as seen in [74]. Some possibilities for the works having chosen the holdout: (i) sample size, which are not so large; (ii) generally used to estimate metrics after performing a cross-validation for hyperparameter setting; otherwise, a double cross-validation would have to be performed (see [13]).

Finally, regarding the evaluation of the knowledge obtained, it is worth mentioning that 14 (46.67%) works (p6, p8, p9, p11, p13, p15, p16, p17, p19, p23, p24, p25, p26, p28), almost half of all those selected, were concerned with applying some interpretability technique (SHAP, feature importance, etc.). Since the use of data mining processes is not yet widely used in the area, as most studies are recent, the processes are designed to support decision making. However, for decision makers to use the results generated by systems that incorporate machine learning models, tools are needed to support them in understanding the decision. This was an important point identified in the works. The aspect of interpretability should, in fact, be incorporated, in some way, into all works.

V. CONCLUDING REMARKS

This work presented a SLM to identify and analyze the primary studies available in the literature to address some research questions on the use of data mining for public planning and budgeting. The period considered was from 01/01/2014 to 31/12/2023 (10 years). In general, it was observed that:

- interest in the topic has been growing since 2019 in a wide range of countries;
- among the few studies surveyed, the "Preparation of fiscal studies and projections" was the budgetary process that most commonly employed data mining techniques (see Fig. 6). Conversely, only one study specifically addressed "External control" and parliamentary "Amendments";
- there are few works that actually use a DM methodology;
- there are few works that enable the reproducibility of their processes/experiments through repositories such as GitHub;
- the datasets used generally contain few instances owing to the problem addressed here, i.e., public budgeting;
- many of the works do not describe how the pre-processing step was carried out; however, when the information is available, it is possible to note the use of standard techniques (normalization/standardization, etc.);
- the most used task was forecasting, owing to the application domain itself; however, regression and classification tasks were also used;
- works related to forecasting tend to use econometric methods as baselines;
- the RF algorithm excelled in all tasks;
- there are few works that address budgetary issues via multi-outputs, i.e., in which several output variables are modeled simultaneously;

- the works seek to evaluate the patterns not only in terms of performance metrics, estimated via holdout, but also in terms of interpretability.

Other issues observed through SLM refer to problems that can be solved and/or explored to use data mining in the presented context, namely:

- it is important that the works provide a clearer description of each of the steps of the data mining process, to enable the understanding of the process; in fact, the use of a methodology, such as CRISP-DM, facilitates such details;
- along with the previous item, it is necessary for authors to make implementations available in repositories such as GitHub, as well as the data used, to guarantee not only the reproducibility of the experiments, but also their use and/or adaptation in other countries that wish to evaluate the proposed modeling. Even though the data are available on official websites, such as the World Bank, it is still not possible to guarantee that the data were downloaded and processed correctly just by describing the process;
- regarding the use of ML algorithms in small samples, the use of data augmentation methods is interesting, since the sample size affects the generalization of the models. This is an important point addressed in only two papers (p8, p18) as a pre-processing step. This treatment could not only improve the performance of the models, but also enable the use of other algorithms, such as (deep) neural networks. However, as mentioned earlier, other solutions are also possible (see Section IV-E);
- in relation to the pre-processing step, regarding the extraction of temporal features, other possibilities besides the use of lags are possible, such as rolling window statistics and expanding window statistics [19], not explored in the works. Furthermore, other temporal features can be extracted using various packages (such as tsfresh),[21] enabling the capture of other potentially relevant information that can contribute to the induction of models with better performance. However, in some cases, interpretability may be compromised, since attributes do not necessarily have inherently interpretable semantics;
- due to the application domain, modeling via multi-outputs could be explored;
- in relation to the forecasting task, there are several ways to estimate performance metrics, which, depending on the modeling, become more appropriate, such as those described in [74]. The works focused only on the holdout;
- the aspect of interpretability should be addressed in all works, since the solutions presented here are to be used as a means of support for decision makers and, therefore, become essential in terms of understanding the process.

Considering the above, we believe that this work has the potential to significantly impact future research and practice in public planning and data mining. Finally, it is important to mention that all the information described here is available here.[22] Thus, it is possible to verify, for each of the aspects, which papers (ID) make up a given item and/or category. For example, it is possible to determine which papers used a certain algorithm by consulting the respective link.

[21] https://tsfresh.readthedocs.io/en/latest/

[22] https://bit.ly/3Csxy5K

REFERENCES

[1] N. Bennett and G. J. Lemoine, "What a difference a word makes: Understanding threats to performance in a VUCA world," *Bus. Horizons*, vol. 57, no. 3, pp. 311–317, May 2014.

[2] N. Tasche, A.-K. Dhungel, and M. Heine, "How to design user-centered decision support systems in public budgeting? Guidelines and a web-based prototype with first insights from a mixed-methods study," in *Proc. Mensch und Comput.*, M. Mühlhäuser, C. Reuter, B. Pfleging, T. Kosch, A. Matviienko, K. Gerling, S. Mayer, W. Heuten, T. Döring, F. Müller, and M. Schmitz, Eds., Sep. 2022, pp. 425–430.

[3] V. Fernandez-Cortez, D. Valle-Cruz, and J. R. Gil-Garcia, "Can artificial intelligence help optimize the public budgeting process? Lessons about smartness and public value from the Mexican federal government," in *Proc. 7th Int. Conf. EDemocracy EGovernment (ICEDEG)*, Apr. 2020, pp. 312–315.

[4] D. Valle-Cruz, V. Fernandez-Cortez, A. López-Chau, and R. Rojas-Hernández, "Public budget simulations with machine learning and synthetic data: Some challenges and lessons from the Mexican case," in *Electronic Governance With Emerging Technologies*, F. Ortiz-Rodríguez, S. Tiwari, M.-A. Sicilia, and A. Nikiforova, Eds., Cham, Switzerland: Springer, 2022, pp. 141–160.

[5] C. Vuppalapati, A. Ilapakurti, S. Vissapragada, V. Mamaidi, S. Kedari, R. Vuppalapati, S. Kedari, and J. Vuppalapati, "Application of machine learning and government finance statistics for macroeconomic signal mining to analyze recessionary trends and score policy effectiveness," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2021, pp. 327–3283.

[6] R. B. Velasco, I. Carpanese, R. Interian, O. C. G. Paulo Neto, and C. C. Ribeiro, "A decision support system for fraud detection in public procurement," *Int. Trans. Oper. Res.*, vol. 28, no. 1, pp. 27–47, Jan. 2021.

[7] K. A. C. Castro, I. Ting, and G. M. T. Bermúdez, "Assessing the performance of open contracting in Colombia through data mining," in *Proc. Int. Workshop Learn. Technol. Educ. Challenges*, in Communications in computer and information science, vol. 1595, Cham, Switzerland, L. Uden and D. Liberona, Eds., Springer, 2022, pp. 68–80.

[8] A. Soylu, Ó. Corcho, B. Elvesæter, C. Badenes-Olmedo, F. Yedro-Martínez, M. Kovacic, M. Posinkovic, M. Medveššek, I. Makgill, C. Taggart, E. Simperl, T. C. Lech, and D. Roman, "Data quality barriers for transparency in public procurement," *Information*, vol. 13, no. 2, p. 99, Feb. 2022.

[9] H. M. Abdelsalam, A. Al-shaar, A. M. Zaki, N. El-Sebai, M. Saleh, and M. H. Khodeir, "Prioritization of public expenditure for a better return on social development: A data mining approach," in *Advanced Machine Learning Technologies and Applications*, A. E. Hassanien, M. F. Tolba, and A. Taher Azar., Cham, Switzerland: Springer, 2014, pp. 523–530.

[10] M. Kasy, "Optimal taxation and insurance using machine learning–sufficient statistics and beyond," *J. Public Econ.*, vol. 167, pp. 205–219, Nov. 2018.

[11] J. I. C. Neto, F. B. D. L. Neto, and F. L. João de Oliveira, "Use of machine learning and multilevel analysis in hierarchical approaches of public expenditure forecasting," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, Oct. 2022, pp. 1155–1160.

[12] D. Carvalho, *Orçamento E Contabilidade Pública*. Amsterdam, The Netherlands: Elsevier, 2010.

[13] P.-N. Tan, M. Steinbach, A. Karpatne, and V. Kumar, *Introduction to Data Mining*, 2nd ed., London, U.K.: Pearson Education Limited, 2020.

[14] S. O. Rezende, *Sistemas Inteligentes: Fundamentos E Aplicaçẽs*. Manila, Philippines: Editora Manole Ltda, 2003.

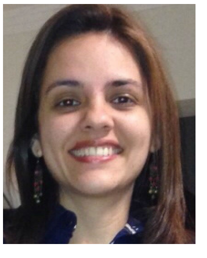
[15] V. Plotnikova, M. Dumas, and F. Milani, "Adaptations of data mining methodologies: A systematic literature review," *PeerJ Comput. Sci.*, vol. 6, p. e267, May 2020.

[16] C. Molnar, *Interpretable Machine Learning*, 2nd ed., 2022. [Online]. Available: https://christophm.github.io/interpretable-ml-book/cite.html

[17] N. Burkart and M. F. Huber, "A survey on the explainability of supervised machine learning," 2020, *arXiv:2011.07876*.
[18] F. Dama and C. Sinoquet, "Time series analysis and modeling to forecast: A survey," 2021, *arXiv:2104.00164*.
[19] F. Lazzeri, *Machine Learning for Time Series Forecasting With Python*. Hoboken, NJ, USA: Wiley, 2020.
[20] A. Fernández, S. García, M. Galar, R. C. Prati, B. Krawczyk, and F. Herrera, *Learning From Imbalanced Data Sets*, vol. 10. Cham, Switzerland: Springer, 2018.
[21] J. Brownlee, *Introduction to Time Series Forecasting with Python: How to Prepare Data and Develop Models to Predict the Future*. Montpelier, VT, USA: Machine Learning Mastery, 2017.
[22] G. Iglesias, E. Talavera, Á. González-Prieto, A. Mozo, and S. Gómez-Canaval, "Data augmentation techniques in time series domain: A survey and taxonomy," *Neural Comput. Appl.*, vol. 35, no. 14, pp. 10123–10145, May 2023.
[23] B. Kitchenham and S. Charters, "Guidelines for performing systematic literature reviews in software engineering," Keele Univ. Durham Univ., Keele, U.K., Joint Rep. EBSE 2007-001, 2007. [Online]. Available: https://www.bibsonomy.org/bibtex/aed0229656ada843d3e3f24e5e5c9eb9 and https://www.elsevier.com/__data/promis_misc/525444systematic reviewsguide.pdf
[24] D. Valle-Cruz, J. R. Gil-Garcia, and R. Sandoval-Almazan, "Artificial intelligence algorithms and applications in the public sector: A systematic literature review based on the PRISMA approach," in *Research Handbook on Public Management and Artificial Intelligence*. Camberley, U.K.: Elgaronline, 2024, pp. 8–26.
[25] R. C. Borges, R. S. E. Silva, H. V. D. Lima, K. D. A. Batista, and D. S. D. Souza, "Optimizing data mining processes in government databases: A comprehensive approach to strategic information analysis," *Caderno Pedagógico*, vol. 21, no. 7, p. e5639, Jul. 2024.
[26] F. A. Amani and A. M. Fadlalla, "Data mining applications in accounting: A review of the literature and organizing framework," *Int. J. Accounting Inf. Syst.*, vol. 24, pp. 32–58, Feb. 2017.
[27] L. Dun, S. Soroka, and C. Wlezien, "Dictionaries, supervised learning, and media coverage of public policy," *Political Commun.*, vol. 38, nos. 1–2, pp. 140–158, Mar. 2021.
[28] D. Guariso, O. A. Guerrero, and G. Castañeda, "Automatic SDG budget tagging: Building public financial management capacity through natural language processing," *Data Policy*, vol. 5, p. e31, Sep. 2023.
[29] J. Giacomoni, *Orçamento Público*. White Plains, NY, USA: Atlas Ltda, 2021.
[30] L. P. Garcia, I. J. C. Schneider, C. de Oliveira, E. Traebert, and J. Traebert, "What is the impact of national public expenditure and its allocation on neonatal and child mortality? A machine learning analysis," *BMC Public Health*, vol. 23, no. 1, p. 11, Apr. 2023.
[31] S. Berset, M. Huber, and M. Schelker, "The fiscal response to revenue shocks," *Int. Tax Public Finance*, vol. 30, no. 3, pp. 814–848, Jun. 2023.
[32] K. Baret, "Fiscal rules' compliance and social welfare," *Ann. Econ. Statist.*, vol. 2023, no. 150, pp. 25–66, Jun. 2023.
[33] F. F. Adedoyin, N. Mavengere, and A. Mutanga, "A simulation experiment on ICT and patent intensity in South Africa: An application of the novel dynamic ARDL machine learning model," *Technological Forecasting Social Change*, vol. 185, Dec. 2022, Art. no. 122044.
[34] Y. Wang and Y. Li, "Chinese economic growth and sustainable development: Role of artificial intelligence and natural resource management," *Resour. Policy*, vol. 85, Aug. 2023, Art. no. 103996.
[35] A. T. Budiman, Y. Ruldeviyani, and A. N. Hidayanto, "Predictive analytics comparison of Indonesia government budget with or without COVID-19 pandemic," in *Proc. Int. Conf. Smart Inf. Syst. Technol. (SIST)*, Apr. 2022, pp. 1–5.
[36] A. Zaragoza-Ibarra, G. G. Alfaro-Calderón, V. G. Alfaro-García, F. Ornelas-Tellez, and R. Gómez-Monge, "A machine learning model of national competitiveness with regional statistics of public expenditure," *Comput. Math. Org. Theory*, vol. 27, no. 4, pp. 451–468, Dec. 2021.
[37] C. Gu, "Application of data mining technology in financial intervention based on data fusion information entropy," *J. Sensors*, vol. 2022, pp. 1–10, Jan. 2022.
[38] R. Mulaudzi and R. Ajoodha, "Application of deep learning to forecast the South African unemployment rate: A multivariate approach," in *Proc. IEEE Asia–Pacific Conf. Comput. Sci. Data Eng. (CSDE)*, Dec. 2020, pp. 1–6.
[39] J. Chen, "Application of support vector machines and holt-winters model in local finance forecast," in *Proc. Asia–Pacific Conf. Commun. Technol. Comput. Sci. (ACCTCS)*, Jan. 2021, pp. 349–353.
[40] K. Lahiri and C. Yang, "Boosting tax revenues with mixed-frequency data in the aftermath of COVID-19: The case of new York," *Int. J. Forecasting*, vol. 38, no. 2, pp. 545–566, Apr. 2022.
[41] X. Hui, "Comparison and application of logistic regression and support vector machine in tax forecasting," in *Proc. Int. Signal Process., Commun. Eng. Manage. Conf. (ISPCEM)*, Nov. 2020, pp. 48–52.
[42] L. F. Zhao, "Data-driven approach for predicting and explaining the risk of long-term unemployment," in *Proc. Int. Conf. Energy Big Data Low-carbon Develop. Manage. (EBLDM)*, vol. 214, 2020, pp. 1–16.
[43] Z. Moubariki, L. Beljadid, M. El Haj Tirari, M. Kaicer, and R. O. H. Thami, "Enhancing cash management using machine learning," in *Proc. 1st Int. Conf. Smart Syst. Data Sci. (ICSSD)*, Oct. 2019, pp. 1–6.
[44] M. Milos, R. Sandro, and D. Boris, "Evaluation of public procurement efficiency of the EU countries using preference learning TOPSIS method," *Econ. Comput. Econ. Cybern. Stud. Res.*, vol. 55, no. 3, pp. 187–202, Sep. 2021.
[45] P. Gogas, T. Papadimitriou, and E. Sofianos, "Forecasting unemployment in the euro area with machine learning," *J. Forecasting*, vol. 41, no. 3, pp. 551–566, Apr. 2022.
[46] J. Prüser, "Forecasting with many predictors using Bayesian additive regression trees," *J. Forecasting*, vol. 38, no. 7, pp. 621–631, Nov. 2019.
[47] D. Valle-Cruz, V. Fernandez-Cortez, and J. R. Gil-Garcia, "From e-budgeting to smart budgeting: Exploring the potential of artificial intelligence in government decision-making for resource allocation," *Government Inf. Quart.*, vol. 39, no. 2, Apr. 2022, Art. no. 101644.
[48] O. Taylan, A. S. Alkabaa, and M. T. Yılmaz, "Impact of COVID-19 on G20 countries: Analysis of economic recession using data mining approaches," *Financial Innov.*, vol. 8, no. 1, p. 81, Sep. 2022.
[49] K. D. Funk, H. L. Paul, and A. Q. Philips, "Point break: Using machine learning to uncover a critical mass in women's representation," *Political Sci. Res. Methods*, vol. 10, no. 2, pp. 372–390, Apr. 2022.
[50] R. Muremyi, H. Dominique, K. Ignace, and F. Niragire, "Prediction of out-of-pocket health expenditures in Rwanda using machine learning techniques," *Pan Afr. Med. J.*, vol. 37, no. 1, p. 357, 2020.
[51] M. Z. Abedin, G. Chi, M. M. Uddin, Md. S. Satu, Md. I. Khan, and P. Hajek, "Tax default prediction using feature transformation-based machine learning," *IEEE Access*, vol. 9, pp. 19864–19881, 2021.
[52] R. Parvin and R. Ferdows, "The Nexus between financial development and economic growth: An application of the VECM model and machine learning algorithms," in *Proc. 25th Int. Conf. Comput. Inf. Technol. (ICCIT)*, Dec. 2022, pp. 360–365.
[53] D. Novaliendy, Y. Hendriyani, C. Yang, and H. Hamimi, "The optimized K-means clustering algorithms to analyzed the budget revenue expenditure in Padang," in *Proc. Int. Conf. Electr. Eng., Comput. Sci. Informat. (EECSI)*, vol. 2, Jun. 2015, pp. 61–66.
[54] K. J. A. Bergantiños and M. J. C. Samonte, "Clustering and classification analysis in financial reporting of Philippine government business enterprises," in *Proc. IEEE 15th Int. Conf. Humanoid, Nanotechnol., Inf. Technol., Commun. Control, Environ., Manage. (HNICEM)*, Nov. 2023, pp. 1–6.
[55] D. Guariso, G. Castañeda, and O. A. Guerrero, "Budgeting for SDGs: Quantitative methods to assess the potential impacts of public expenditure," *Develop. Eng.*, vol. 8, Nov. 2023, Art. no. 100113.
[56] M. F. Abd El-Aal, "Analysis factors affecting Egyptian inflation based on machine learning algorithms," *Data Sci. Finance Econ.*, vol. 3, no. 3, pp. 285–304, 2023.
[57] F. F. Vasconcelos, R. M. Sátiro, L. P. L. Fávero, G. T. Bortoloto, and H. L. Corrêa, "Analysis of judiciary expenditure and productivity using machine learning techniques," *Mathematics*, vol. 11, no. 14, p. 3195, Jul. 2023.
[58] C.-H. Yang, T. Molefyane, and Y.-D. Lin, "The forecasting of a leading country's government expenditure using a recurrent neural network with a gated recurrent unit," *Mathematics*, vol. 11, no. 14, p. 3085, Jul. 2023.
[59] C.-F. Ivaşcu and S. E. Ştefoni, "Modelling the non-linear dependencies between government expenditures and shadow economy using data-driven approaches," *Sci. Ann. Econ. Bus.*, vol. 70, no. 1, pp. 97–114, Mar. 2023.
[60] J. Simonov and Z. Gligorov, "Customs revenues prediction using ensemble methods (Statistical modelling vs machine learning)," *World Customs J.*, vol. 15, no. 2, pp. 111–128, Sep. 2021.

[61] M. Ghiassi and B. D. Simo-Kengne, "Machine learning for economic modeling: An application to South Africa's public expenditures," *Int. J. Public Admin. Digit. Age*, vol. 8, no. 1, pp. 1–17, Mar. 2022.
[62] I. M. Joseph, *O Trivium: As Artes Liberais Da Lógica, Da Gramática E Da Retórica* (Educação Clássica). É Realizações Editora Livraria e Distribuidora LTDA, 2018. [Online]. Available: https://books.google.com.br/books/about/O_Trivium.html?id =8NNfDwAAQBAJ&redir_esc=y
[63] R. A. Musgrave and A. B. Simões, *Teoria Das Finanças Públicas: Um Estudo De Economia Governamental*. São Paulo, Brazil: Atlas, 1976.
[64] F. A. R. da Silva, *Finanças Públicas*. White Plains, NY, USA: Atlas, 1978.
[65] R. Tatman, J. Vanderplas, and S. Dane, "A practical taxonomy of reproducibility for machine learning research," in *Proc. Reproducibility Mach. Learn. Workshop ICML*, Jun. 2018, p. 5.
[66] V. Cerqueira, L. Torgo, and C. Soares, "A case study comparing machine learning with statistical methods for time series forecasting: Size matters," *J. Intell. Inf. Syst.*, vol. 59, no. 2, pp. 415–433, Oct. 2022.
[67] S.-A.-N. Alexandropoulos, S. B. Kotsiantis, and M. N. Vrahatis, "Data preprocessing in predictive data mining," *Knowl. Eng. Rev.*, vol. 34, pp. 1–33, Jan. 2019.
[68] A. L. de Rojas, "Data augmentation in economic time series: Behavior and improvements in predictions," *AIMS Math.*, vol. 8, no. 10, pp. 24528–24544, 2023.
[69] L. A. Kone, A. O. Leonteva, M. T. Diallo, A. Haouba, and P. Collet, "Short time series forecasting method based on genetic programming and Kalman filter," in *Complex Computational Ecosystems*, P. Collet, L. Gardashova, S. El Zant, and U. Abdulkarimova, Eds., Cham, Switzerland: Springer, 2023, pp. 77–89.
[70] S. J. Fong, G. Li, N. Dey, R. Gonzalez-Crespo, and E. Herrera-Viedma, "Finding an accurate early forecasting model from small dataset: A case of 2019-nCoV novel coronavirus outbreak," *Int. J. Interact. Multimedia Artif. Intell.*, vol. 6, no. 1, p. 132, 2020.
[71] J. Wątróbski, J. Jankowski, P. Ziemba, A. Karczmarczyk, and M. Zioło, "Generalised framework for multi-criteria method selection," *Omega*, vol. 86, pp. 107–124, Jul. 2019.
[72] C.-L. Hwang and K. Yoon, *Methods for Multiple Attribute Decision Making*. Berlin, Germany: Springer, 1981.
[73] P. H. Franses, "A note on the mean absolute scaled error," *Int. J. Forecasting*, vol. 32, no. 1, pp. 20–22, Jan. 2016.
[74] V. Cerqueira, L. Torgo, and I. Mozetic, "Evaluating time series forecasting models: An empirical study on performance estimation methods," *Mach. Learn.*, vol. 109, no. 11, pp. 1997–2028, Nov. 2020.

JOSÉ CLÁUDIO GUEDES DAS NEVES received the master's degree in communications management from Coventry University, U.K. He was a Planning and Budget Analyst with Brazilian Federal Budget Secretariat. His research interests include data mining, machine learning, and public finance.



VERONICA OLIVEIRA DE CARVALHO received the Ph.D. degree in computer science from USP, Brazil. She is currently a Professor in computer science with UNESP. She has published papers at several international conferences. Her research interests include data mining and machine learning.