Intelligent Human Systems Integration (IHSI 2024), Vol. 119, 2024, 548–559
https://doi.org/10.54941/ahfe1004520

Using Machine Learning for Anomaly Detection in German Public Budgeting Data

Anna-Katharina Dhungel[1], Lea Watermann[1], Christiane Wegner[2], and Moreen Heine[1]

[1]Institute for Multimedia and Interactive Systems, University of Lübeck, Germany
[2]MACH AG, Lübeck, Germany

ABSTRACT

Outlier detection can assist individuals involved in the public budgeting process by enabling them to focus on specific values within the budget plan, which is particularly valuable considering the extensive nature of these plans (the 2023 German federal budget comprised 3,214 pages). Through a human-centered development approach, this study evaluates the feasibility of implementing algorithms for outlier detection in the context of public budgeting in Germany. In addition, results of the three algorithms Isolation Forest, One-Class Support Vector Machine, and Local Outlier Factor are compared. Our results reveal two insights: 1. The quality and availability of data pose fundamental challenges for outlier detection using machine learning in Germany; 2. The tested algorithms are indeed proficient in detecting certain values within the budget plan as anomalous and they exhibit a certain level of consistency. Nevertheless, computing the measure of accuracy presents difficulties due to the complexity of discerning when a value is in accordance with political intent or constitutes an error. The study overall highlights the potential of outlier detection in public budgeting while emphasizing the requirement for appropriate datasets and ongoing evaluation by the target audience.

Keywords: Outlier detection, Public budgeting data, Feasibility analysis, Human-centered design, Digitization of the public sector

INTRODUCTION

One of the biggest challenges a state faces is deciding where to put how much money – for every political program that gets funded, there is an alternative that some people would prefer (Kirshner, 2003). The budget preparation for the upcoming year is a formidable task involving complex considerations, including the alignment of political interests and conflicting objectives. This process requires the management of a vast amount of data, with an emphasis on identifying relationships and dependencies. Additionally, external factors like EU-imposed debt limits (Maastricht Treaty) and other events such as natural disasters or pandemics must be considered. Ensuring early error detection by various stakeholders with differing levels of expertise is essential. Given these challenges, it is relevant to explore the feasibility of leveraging modern technologies to enhance public budgeting.

Machine learning techniques are expected to increase efficiency, accuracy, flexibility, and data utilization in the public sector (Alexopoulos et al., 2019). One of the most common tasks during public budgeting involves verifying the accuracy of numerical data, with the aim of promptly identifying any discrepancies. Regarding this challenge, algorithms can potentially provide assistance. Hence, the research questions of this study are:

- RQ1: Is it fundamentally feasible to employ machine learning for outlier detection in public budget planning in Germany?
- RQ2: What values are identified as outliers by Isolation Forest, Local Outlier Factor, and the One-Class Support Vector Machine?

We followed a human-centered development approach, including interviews, workshops, and a survey to assess the context of use. Our findings indicate that in Germany, machine learning for outlier detection in public budgeting faces feasibility challenges primarily related to data quality and availability. While testing algorithm on a small scale is feasible, limited data availability restricts widespread adoption. The analysis of Isolation Forest (iForest), Local Outlier Factor (LOF), and One-Class Support Vector Machine (OC-SVM) demonstrates their potential for detecting outliers in public budgeting data. However, defining outliers in this context is complex, as not every deviation constitutes an error. Continuous evaluation requires close collaboration with the target audience.

The paper is structured as follows: first, some information about public budgeting is given, followed by an overview of the current state of research. Then, the methodological approach is described. Based on this, the results of the feasibility analysis and the comparison of the algorithms are presented. The paper concludes by discussing the key findings and addressing next possible research endeavors.

PUBLIC BUDGETING

Public budgeting differs significantly from financial management in private organizations, making direct use of existing IT applications for corporate budgeting infeasible. In the private sector, the primary focus is profit maximization and cost reduction, predominantly driven by monetary incentives, competition, and market forces – the companies' shareholders act as the control organ. In contrast, the main objective in the public sector is not to maximize profit but to generate common good and public value. To meet this aim, costs have to be covered. Achieving a balanced budget is crucial in public budgeting, and decision incentives often involve non-monetary considerations tied to various political objectives. Unlike companies, governments do not face direct competition; instead, budget decisions are subject to oversight by the electorate or audit courts (Lorenz, 2012). Public budgets serve as a blueprint for government spending, funding activities, and managing borrowing. They are influenced by economic factors, such as inflation, and serve as historical records (Lewis and Hildreth, 2013). The public budgeting process involves diverse stakeholders with varying objectives, susceptibility to external factors, a separation between taxpayers and decision-makers, and other inherent constraints (Rubin, 1990).

In Germany, budgeting processes are influenced by federal structures, Maastricht Treaty criteria from the European Union, and the debt brake mandated by the German constitution. The federal government, states, municipalities, and other organizations in the public sector manage their own budgets, adhering to principles like economy and efficiency. They are required to provide complete and detailed documentation of all income, expenditures, and future obligations, specifying purpose and amount, which results in a high level of complexity (de Vries et al., 2019). For instance, the 2022 state budget of Schleswig-Holstein, with a volume of around 14 billion euros, comprises over 6,000 budget items at the lowest structural level.

LITERATURE REVIEW

While there is extensive research on artificial intelligence applications in public administrations (compare e.g. Gerrits, 2021; Sousa et al., 2019; Wirtz et al., 2021), the link between machine learning and public finance has been little explored (Fernandez-Cortez et al., 2020; Wirtz et al., 2021). Some researchers argue that the "study of public budgeting through AI techniques is practically non-existent" (Valle-Cruz et al., 2020). Nonetheless, a few research efforts are evident. For example, Fernandez-Cortez et al. adopt a macroeconomic approach, employing machine learning techniques to simulate optimal government spending allocation using data from various sources (Fernandez-Cortez et al., 2020). They also explore AI's potential in classifying budget allocations to achieve the highest GDP value with minimal inflation increase (Valle-Cruz et al., 2020). In a more recent study, they extend their investigation to reduce the Gini index while increasing GDP and reducing inflation, using a more extensive dataset (Valle-Cruz et al., 2022). Other studies touch on AI and public budgeting-related topics, such as Papagiannies et al., who analyze the relationship between regional classification and environment-driven development for optimal regional budget allocation (Papagiannis et al., 2020). Additionally, Martins et al. examine resource allocation in Brazilian universities aided by a decision support system (Martins et al., 2019). In contrast, the field of outlier detection in the private financial sector boasts a more substantial body of work, only a few examples are provided here. Ahmed et al. employ clustering-based anomaly detection techniques for fraud detection and highlight the ever-evolving nature of fraud detection (Ahmed et al., 2016). John and Naaz use machine learning algorithms to identify instances of fraudulent activity in credit card transactions, with the local outlier factor demonstrating higher accuracy (John and Naaz, 2019). Fan et al. describe the use of anomaly detection algorithms for bankruptcy prediction and find the isolation forest to be the most effective (Fan et al., 2017). However, none of these studies have yet taken the approach of using machine learning to identify outliers in public budgets. It therefore seems promising from a research perspective to address this gap.

METHODOLOGY

The subject was explored through the human-centered design process according to DIN EN ISO 9241-210, involving two workshops, eight interviews, and a nationwide survey (n = 69), all specifically targeted at individuals involved in diverse budget planning roles, such as departmental budget preparation, central budget planning with responsibility for verification, and state audit office personnel. Within the context analysis, the primary aim was to gain a comprehensive understanding of the process, roles, challenges, and the potential for machine learning support. Liberating Structures techniques were applied in workshops, interviews followed a guided format, and the survey included both open-ended qualitative responses and quantitatively assessable questions, reaffirming the relevance of outlier detection. To investigate the feasibility of machine learning for outlier detection, we leverage the results of this context analysis. To assess the outcomes achievable through machine learning, we compare the performance of iForest, LOF, and OC-SVM in identifying anomalies within both entire sections of a budget plan and individual items.

FEASIBILITY ASSESSMENT

The fundamental prerequisite for applying machine learning is the availability of data in a volume suitable to the context and of appropriate quality. However, public budgeting data availability and quality remains a significant challenge in Germany. A substantial portion of this data is exclusively accessible in pdf format, with limited availability in machine-readable formats like csv. Furthermore, instances of improperly saved csv files, as seen, for example, in Schleswig-Holstein in 2011, among other cases, can obscure the actual data structure. Historical data, often spanning a relatively short time frame of two to ten years, is common. While longer data periods were mentioned during the interviews, they remained undisclosed due to a lack of responsibility or technical barriers. Additionally, comprehensive explanations for budget items are frequently absent. Table 1 provides an illustrative summary of data availability for selected states and municipalities, highlighting the prevalence of short time periods, restricted to budget plans, with inconsistent provision of actual expenditure data as well as varying formats.

Furthermore, datasets frequently contain missing values, and budget structures undergo regular changes over time, complicating the tracking of expenditures for specific outputs. While there are guidelines for budget plan structure and numbering, not all entities adhere to them. During the interviews, it became evident that, to a certain extent, the shifting of certain expenditures between budget items and the resulting reduction in transparency is sometimes a deliberate political strategy. Moreover, behind the values in the budget plans lies a political decision-making process that is not discernible from the values themselves. Hence, a collaboration with stakeholders is imperative for long-term machine learning deployment in this context. This collaboration is equally vital for establishing criteria for identifying outliers in the context of public budgeting, as this is a challenging task due to the fact that not every statistical deviation necessarily signifies an error.

Political decisions and global developments can significantly influence public finances without necessarily indicating an error. As a result, it is often challenging to establish a clear definition of an outlier that captures all relevant aspects of the process. Additionally, employees engaged in public budgeting exhibit significant diversity in their technological preferences and capabilities. For example, during the survey, one respondent noted that their municipality still employs traditional pen-and-paper methods for budget preparation. This diversity should be taken into account when selecting individuals for evaluations. In summary, widespread implementation of machine-driven outlier analysis in public budgeting is presently unfeasible. Nonetheless, test runs can be conducted, particularly in municipalities and regions with adequate data availability, serving as reference points for potential future adoption of machine learning in budget planning.

Table 1. Public budgeting data in German federal states (S) and municipalities (M).

Entity	Period	Planned/actual data	Format*
Schleswig-Holstein (S)	2011–2021	planned/actual	csv (since 2011) xls (2014-2017) xlsx (since 2021)
Lower Saxony (S)	1999–2023	planned	pdf
North Rhine-Westphalia (S)	2012–2021	planned/actual	csv (2012-2017) xlsx (since 2018)
Berlin (S)	2010–2023	planned	xls (2010-2011) xlsx (since 2012) csv (2014-2015)
Aachen (M)	2022–2023	planned	csv
Duisburg (M)	2015–2022	planned/actual	pdf (since 2015) csv (since 2019)
Glückstadt (M)	2019–2022	planned/actual	csv
Kiel (M)	2015–2022	incomprehensible	csv
Moers (M)	2013–2022	planned/actual	xml (2013-2017) csv (since 2018)
Wuppertal (M)	2012–2021	planned/actual	csv (2012-2017) xlsx (since 2018)

* If no period is specified, the total period is available in the respective format.

OUTLIER DETECTION WITH MACHINE LEARNING

Dataset

Schleswig-Holstein has provided its budget data in csv format since 2012, as the 2011 data is unusable. This dataset includes earmarked funds, the actual and planned values for the previous year, and the preliminary planning for the current budget. As an e-government research group with an extensive network in digitization, we've chosen to analyze section 14 of Schleswig-Holstein's budget data, focusing on Information and Communication Technologies (ICT). This strategic selection facilitates access to the appropriate individuals for future evaluations. During the observed period, section 14 contains a total of 148 budget items. However, only 26 of these items are consistently filled with values each year, while the rest frequently contain missing values. Budget items with values in just one or two years are unsuitable for analysis. Nevertheless, using only these 26 items, which represent just 17.5% of all items, appears inadequate. To strike a balance, we consider all items with continuous values for at least seven years, tolerating some missing data. Consequently, we end up with a dataset that includes planned data from 2012 to 2021 and actual data from 2012 to 2020 for 48 budget items. Figure 1 depicts the distribution of values for these items.

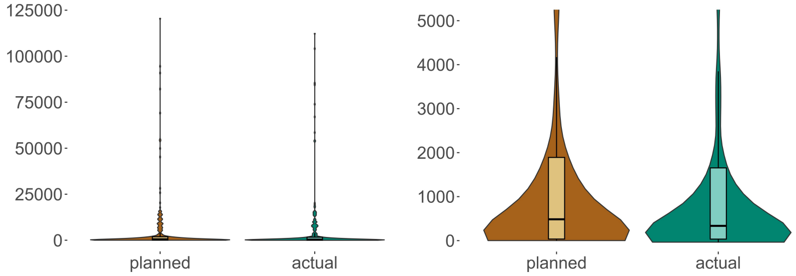

Figure 1: Planned and actual values in our subset of Schleswig-Holstein's Section 14 budget from 2012 to 2021 in k euros (left: overall distribution, right: up to five million euros).

The planned values exhibit a median of 485 million euros (1st quartile: 36 million euros, 3rd quartile: 1891 million euros), while the actual values have a median of 335 million euros (1st quartile: 35 million euros, 3rd quartile: 1653 million euros). This indicates that 75% of the data is below 2 million euros, but outliers with values as high as 1.2 billion euros are present. To identify outliers statistically, the threefold interquartile range was applied, resulting in the removal of two budget items, which consistently appeared as outliers for each year. Other items were identified as outliers but only for specific years, so they were retained. As a result, the dataset was reduced from 48 to a total of 46 budget items. The dataset comprises budget item ID, year, actual value, and planned value. Through our interviews with stakeholders, we have learned that both absolute and relative changes are considered concerning the previous year's data. Therefore, we have incorporated the following features into our dataset:

- absolute difference between planned and actual values for the same year;
- percentage change between the current actual value and the previous year, as well as between the current actual value and the value two years prior;
- percentage change between the current planned value and the previous year, as well as between the current planned value and the value two years prior.

LOF and OC-SVM cannot handle missing values, so we set them to zero. This adjustment was unnecessary for iForest, as the algorithm interprets missing values as information. We use the entire dataset for both training and evaluation, as we lack ground truth information for anomaly assignment.

Consequently, we cannot partition the data into training and test sets to assess algorithm accuracy in this analysis.

Algorithms

A method that does not require distance or density measures for anomaly detection is Isolation Forest (iForest). It follows the approach to separate data points from each other in an ensemble of binary tree structures, called isolation tree. As anomalies are few and different in their attribute values compared to normal data points, they are prone to isolation. Therefore, outlier data points are more likely to be located near the root of the tree, while inliers are more likely to be isolated deeper in the tree (Liu et al., 2008).

LOF is a density-based algorithm that assesses the density of a data point in relation to its nearest neighbors. It operates on the principle that outliers are atypical in proximity to other data points, whereas inliers tend to be in relatively uniform environments. A value indicating how much that point deviates from its surroundings is calculated for each data point and is called Local Outlier Factor (LOF). An LOF exceeding 1 indicates that the data point is unusual because its local density is lower than that of its neighbors. A higher LOF suggests a greater likelihood of the data point being an outlier (Breunig et al., 2000).

OC-SVM operates by learning a decision boundary, typically a hyperplane, that encloses the majority of data points, categorizing them as the 'normal' class, and flagging data lying beyond this boundary as potential outliers. The goal of OC-SVM is to optimize the margin around the normal data points while minimizing the likelihood of false positives among the outliers (Shin et al., 2005).

Outlier Analysis for the Entire Dataset

In Table 2, the R packages and parameter settings used for outlier identification in the complete dataset by each algorithm are outlined. Unused parameters were set to their default values.

Table 2. Packages and parameters used for outlier analysis of the entire dataset.

	iForest	**LOF**	**SVM**
R package:	h2o	dbscan	e1071
Parameter settings:	*ntrees = 100*	*k = 18*	*type = "one-class"*
	sample_size = 256	*outlier: LOF ≥ 3.2*	*kernel = "linear"*
	s >= 0.6		*nu = 0.9*

Initially, we assessed the budget items that the three algorithms flagged as outliers in various years. iForest identified 13 values from 6 distinct budget items, while LOF marked 20 values from 11 items, and OC-SVM labeled 25 values from 12 items as outliers. Next, we analyzed the level of concurrence among the algorithms based on title and year, as well as solely by title (see Table 3).

Table 3. Overlap of the algorithm results by title and year as well as by individual title.

	Matching title and year	Matching title
iForest & LOF & OC-SVM	1	3
iForest & LOF	2	3
iForest & OC-SVM	8	6
LOF & OC-SVM	4	5

The labeling of outliers displays variability, with only one budget item in a given year identified by all three algorithms. However, when considering only the item title, we identify three such instances. Additionally, focusing solely on title-based matches reveals that all the outliers identified by iForest are also detected by OC-SVM, with the latter identifying additional outliers. Moreover, when examining alignment by both title and year, iForest and OC-SVM exhibit a higher level of similarity compared to other combinations, with 8 identical results. The lowest intersection is observed between iForest and LOF. Out of the 11 titles labeled as outliers by LOF, at least 5 of them are also included by OC-SVM.

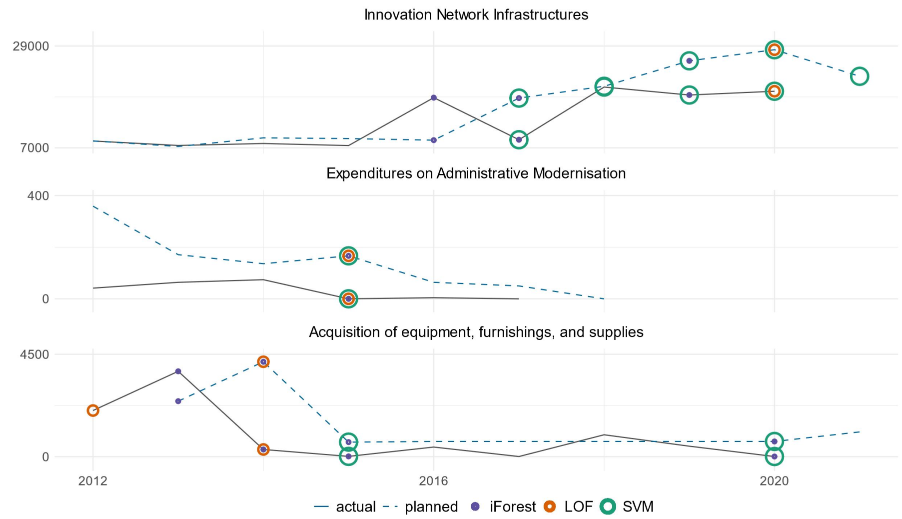

Figure 2: Similarities of outliers detected in three budget items within section 14 of Schleswig-Holstein budget data (in k euros).

A closer analysis of the three items in which all algorithms have identified outliers reveals a noteworthy degree of convergence in the outlier labeling (see Figure 2). Once again, it becomes evident that at this juncture, the necessary subsequent step must involve an evaluation of the target group in order to determine which of the years among these titles are actual outliers.

Outlier Analysis for Individual Budget Items

We also aimed to compare the values identified as outliers for individual budget items. Due to the limited timeframe, the number of observations is reduced to n = 10 for each budget item, necessitating parameter adjustments (see Table 4).

Table 4. Packages and parameters used for outlier analysis of single budget item.

	iForest	LOF	OCSVM
R package:	h2o	dbscan	e1071
Parameter settings:	*ntrees = 100*	$k = 6$	*type = "one-class"*
	sample_size = 7	*outlier:* $LOF \geq 1.2$	*kernel = "linear"*
	$s >= 0.6$		$nu = 0.0004$
	max_depth = 4		

We selected budget items with clear expense definitions and intentionally included those with non-linear trends to increase the likelihood of encountering outliers. The following three budget items were compared (item ID, title):

- 140251802: Business Trips
- 140251144: Expenses for the Use of Smartphones (Mobile Communication)
- 140251802: Rent and Leases for Machinery, Equipment, and Vehicles

In the case of Business Trips, all methods reach a consensus, identifying outliers in both 2021 and 2020. LOF additionally detects anomalies in 2019, while OC-SVM identifies an outlier in 2014. For the Rents/Leases budget item, 2021 stands out as a significant year, as all three methods flag it as an outlier. iForest and LOF also identify outliers in 2012, while LOF and OC-SVM detect anomalies in 2020. OC-SVM further marks 2013 as an outlier. Regarding the Smartphones budget item, iForest and LOF consistently label 2019 and 2020 as outliers. iForest additionally identifies 2018 as an anomaly, while LOF and OC-SVM jointly detect outliers in 2021. These findings are summarized in Figure 2.

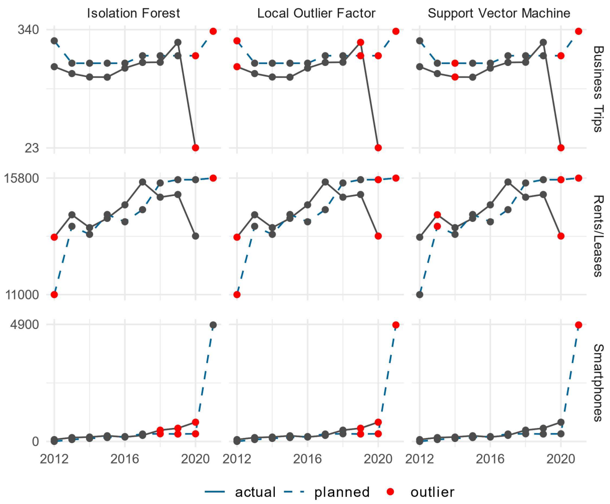

Figure 3: Comparative outlier detection results for budget items business trips, rents/leases, and smartphones (in k euros).

The results underscore the complexity of evaluating these outliers within the context of public budgeting. Regarding the Smartphones category, values for the years 2020 and 2021 were predominantly flagged as outliers. It is worth noting that these data points coincide with the period of the COVID-19 pandemic, during which there was a deliberate political effort to accelerate digitalization to facilitate measures such as remote working. Consequently, an increase in expenditures for mobile communication during this period appears to be justifiable and not necessarily indicative of errors in the budget. Similarly, the sharp decrease in expenses for business trips, also attributed to the pandemic, provides a reasonable explanation for the 2020 actual spending. Notably, despite the ongoing pandemic, there were plans for increased business trip expenditures in 2021, as identified as outliers by all three algorithms. Given the additional contextual knowledge of the societal environment at that time, these planned expenditures indeed appear dubious. In the case of rents and leases, it is surprising that none of the algorithms flagged the values for the year 2017 as outliers, despite a noticeable surge in actual expenditures that significantly exceeded the corresponding budgeted amount.

DISCUSSION AND OUTLOOK

The results should be discussed from both a technical and a domain-specific perspective. The technical perspective includes considerations such as whether replacing missing values with zeros is the correct approach, as this may lead the algorithms to flag endpoints as outliers. Furthermore, questions arise about whether the parameter settings yield optimal results and why there is such variability in the agreement among the algorithms. From a domain-specific viewpoint, the discussion should focus on which values can genuinely be defined as outliers in the budget data, as only then can the accuracy of the individual algorithms be determined. Hence, a corresponding evaluation involving stakeholders from the target audience is scheduled in the near future. In further tests and prototypes, the issue of explainable artificial intelligence (XAI) should also be taken into account, as it is of paramount importance in the public sector that algorithmic decision support systems are comprehensible and transparent (Barbado et al., 2022).

CONCLUSION

It has become evident that budgetary planning within public administration represents a domain where the application of machine learning could be fundamentally advantageous and beneficial – given that the data is not only well-structured but also of high quality, and a development process is conducted via target group evaluation. Actors involved in the budgetary planning process are required to oversee a substantial amount of data while typically operating under significant time constraints. Consequently, a technological assistant capable of flagging anomalies within the planned data would serve as a considerable support. Moreover, from a scientific perspective, exploring this distinctive form of outliers and determining the most effective algorithm in this context is of noteworthy interest. Hence, delving deeper into the application of machine learning for detecting anomalies within budgetary planning data appears both fitting and justified.

REFERENCES

Ahmed, M., Mahmood, A. N., Islam, Md. R. (2016) A survey of anomaly detection techniques in financial domain. Future Generation Computer Systems 55, 278–288. https://doi.org/10.1016/j.future.2015.01.001

Alexopoulos, C., Lachana, Z., Androutsopoulou, A., Diamantopoulou, V., Charalabidis, Y., Loutsaris, M. A. (2019) How Machine Learning is Changing e-Government, in: Proceedings of the 12th International Conference on Theory and Practice of Electronic Governance. ACM, Melbourne VIC Australia, pp. 354–363. https://doi.org/10.1145/3326365.3326412

Barbado, A., Corcho, Ó., Benjamins, R. (2022) Rule extraction in unsupervised anomaly detection for model explainability: Application to OneClass SVM. Expert Systems with Applications 189, 116100. https://doi.org/10.1016/j.eswa.2021.116100

Breunig, M. M., Kriegel, H.-P., Ng, R. T., Sander, J. (2000) LOF: identifying density-based local outliers. SIGMOD Rec. 29, 93–104. https://doi.org/10.1145/335191.335388

de Vries, M. S., Nemec, J., Špaček, D. (2019) Performance-based budgeting in the public sector. Springer Berlin Heidelberg, New York, NY.

Fan, S., Liu, G., Chen, Z. (2017) Anomaly detection methods for bankruptcy prediction, in: 2017 4th International Conference on Systems and Informatics (ICSAI). IEEE, Hangzhou, pp. 1456–1460. https://doi.org/10.1109/ICSAI.2017.8248515

Fernandez-Cortez, V., Valle-Cruz, D., Gil-Garcia, J. R. (2020) Can Artificial Intelligence Help Optimize the Public Budgeting Process? Lessons about Smartness and Public Value from the Mexican Federal Government, in: 2020 Seventh International Conference on EDemocracy & EGovernment (ICEDEG). IEEE, Buenos Aires, Argentina, pp. 312–315. https://doi.org/10.1109/ICEDEG48599.2020.9096745

Gerrits, L. (2021) Soul of a new machine: Self-learning algorithms in public administration. IP 26, 237–250. https://doi.org/10.3233/IP-200224

John, H., Naaz, S. (2019) Credit Card Fraud Detection using Local Outlier Factor and Isolation Forest. International Journal of Computer Sciences and Engineering 7, 1060–1064. https://doi.org/10.26438/ijcse/v7i4.10601064

Kirshner, J. (2003) Money is politics. Review of International Political Economy 10, 645–660. https://doi.org/10.1080/09692290310001601911

Lewis, C. W., Hildreth, W. B. (2013) Budgeting: politics and power, Second edition. ed. Oxford University Press, New York.

Liu, F. T., Ting, K. M., Zhou, Z.-H. (2008) Isolation Forest, in: 2008 Eighth IEEE International Conference on Data Mining. IEEE, Pisa, Italy, pp. 413–422. https://doi.org/10.1109/ICDM.2008.17

Lorenz, C. (2012) The genesis of performance budgeting, in: The Impact of Performance Budgeting on Public Spending in Germany's Laender. Gabler Verlag, Wiesbaden, pp. 27–53. https://doi.org/10.1007/978-3-8349-3483-3_3

Martins, C. L., Teixeira de Almeida, A., Morais, D. C. (2019) Design of a Decision Support System for Resource Allocation in Brazil Public Universities: International Journal of Decision Support System Technology 11, 20–34. https://doi.org/10.4018/IJDSST.2019010102

Papagiannis, F., Gazzola, P., Burak, O., Pokutsa, I. (2020) An intelligent environmental plan for sustainable regionalisation policies: The case of Ukraine. Environmental Science & Policy 108, 77–84. https://doi.org/10.1016/j.envsci.2020.03.010

Rubin, I. S. (1990) The Politics of Public Budgeting: Getting and Spending, Borrowing and Balancing, 1. [publ.]. ed, Chatham House series on change in American politics. Chatham House Publ, Chatham, NJ.

Shin, H. J., Eom, D.-H., Kim, S.-S. (2005) One-class support vector machines—an application in machine fault detection and classification. Computers & Industrial Engineering 48, 395–408. https://doi.org/10.1016/j.cie.2005.01.009

Sousa, W. G. de, Melo, E. R. P. de, Bermejo, P. H. D. S., Farias, R. A. S., Gomes, A. O. (2019) How and where is artificial intelligence in the public sector going? A literature review and research agenda. Government Information Quarterly 36, 101392. https://doi.org/10.1016/j.giq.2019.07.004

Valle-Cruz, D., Fernandez-Cortez, V., Gil-Garcia, J. R. (2022) From E-budgeting to smart budgeting: Exploring the potential of artificial intelligence in government decision-making for resource allocation. Government Information Quarterly 39, 101644. https://doi.org/10.1016/j.giq.2021.101644

Valle-Cruz, D., Gil-Garcia, J. R., Fernandez-Cortez, V. (2020) Towards Smarter Public Budgeting? Understanding the Potential of Artificial Intelligence Techniques to Support Decision Making in Government, in: The 21st Annual International Conference on Digital Government Research. ACM, Seoul Republic of Korea, pp. 232–242. https://doi.org/10.1145/3396956.3396995

Wirtz, B. W., Langer, P. F., Fenner, C. (2021) Artificial Intelligence in the Public Sector - A Research Agenda. International Journal of Public Administration 44, 1103–1128. https://doi.org/10.1080/01900692.2021.1947319