

Detecção e Análise de Anomalias em Séries Temporais de Despesas Municipais

Marco Túlio Dutra^{1,2}, Lucas G. L. Costa¹,
Gabriel P. Oliveira¹, Mariana O. Silva¹, Gisele L. Pappa¹

¹Universidade Federal de Minas Gerais (UFMG) – Belo Horizonte, MG – Brasil

²Universidade Federal de Ouro Preto (UFOP) – Ouro Preto, MG – Brasil

marco.dutra@aluno.ufop.edu.br, lucas-lage@ufmg.br
{gabrielpoliveira,mariana.santos,glpappa}@dcc.ufmg.br

Abstract. *This paper presents an approach for detecting and analyzing anomalies in time series of public expenditures. The approach employs a combination of advanced techniques for anomaly detection in time series, including statistical analysis and machine learning. In addition to anomaly detection, the approach allows the ranking of expenditures based on the number of detected anomalies and the monetary values involved, facilitating the prioritization of audits. Our approach is evaluated on a real dataset containing over one million municipal expenditure records from the state of Minas Gerais. Overall, our findings indicate that the approach effectively identifies and prioritizes cases with a high potential for irregularities.*

Resumo. *Este artigo apresenta uma abordagem para a detecção e análise de anomalias em séries temporais de despesas públicas. A abordagem emprega uma combinação de técnicas avançadas para detecção de anomalias em séries temporais, incluindo análise estatística e aprendizado de máquina. Além da detecção de anomalias, a abordagem permite o ranqueamento das despesas conforme a quantidade de anomalias identificadas e nos valores envolvidos, facilitando a priorização das auditorias. A abordagem é avaliada em um conjunto de dados reais contendo mais de um milhão de registros de despesas municipais do estado de Minas Gerais. Os experimentos indicam que a abordagem é capaz de identificar e priorizar casos com alto potencial de irregularidade.*

1. Introdução

A gestão transparente dos recursos públicos é fundamental para garantir a eficiência e a integridade da administração governamental. No Brasil, a Lei Federal nº 12.527/11,¹ conhecida como “Lei de Acesso à Informação”, determina que dados sobre receitas e despesas públicas devem ser acessíveis aos cidadãos, permitindo o acompanhamento e a fiscalização da alocação de recursos. No entanto, a grande quantidade de informações contidas nesses registros representa um desafio para a identificação de irregularidades, como despesas injustificadas e contratos superfaturados [Silva et al. 2023].

A análise da evolução temporal dos gastos públicos é fundamental para identificar desvios que podem indicar irregularidades, sendo as séries temporais uma ferramenta estatística central nesse processo. Nesse sentido, a detecção de anomalias

¹Lei nº 12.527/11: https://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/112527.htm

em séries temporais é uma abordagem promissora para identificar padrões incomuns [Zamanzadeh Darban et al. 2024], no contexto de despesas públicas as anomalias podem indicar fraudes ou erros administrativos. No entanto, a análise desses dados é complexa devido ao seu alto volume e à variabilidade nos padrões de gasto entre diferentes municípios e funções orçamentárias. Métodos tradicionais de auditoria, baseados exclusivamente em inspeção manual, são insuficientes para processar eficientemente essa quantidade de informações, tornando necessária a aplicação de técnicas automatizadas para identificação de potenciais anomalias [Gomide et al. 2023].

Nesse contexto, este trabalho propõe uma abordagem para detecção e análise de anomalias em séries temporais de despesas públicas municipais. A abordagem combina diferentes técnicas, incluindo análise estatística, aprendizado de máquina e métodos probabilísticos, permitindo não apenas a detecção de anomalias, mas também o ranqueamento das despesas com base na quantidade de ocorrências e nos valores envolvidos. Esse ranqueamento possibilita a priorização dos casos mais críticos, facilitando o direcionamento de auditorias e investigações. trabalho da auditoria.

Este artigo está estruturado da seguinte forma. A Seção 2 apresenta trabalhos relacionados na área de detecção de anomalias e auditoria de despesas públicas, enquanto a Seção 3 descreve a abordagem utilizada na construção do processo de detecção de anomalias e no ranqueamento. A Seção 4 apresenta os resultados dos experimentos realizados com dados de despesas de municípios do Estado de Minas Gerais. Por fim, a Seção 5 discute as conclusões, possíveis limitações e sugestões para trabalhos futuros.

2. Trabalhos Relacionados

O combate a fraudes no setor público é uma prioridade global, sendo essencial para garantir a integridade das instituições governamentais [Silva et al. 2020]. Um aspecto em comum dos trabalhos que abordam essa tarefa é a necessidade de lidar com grandes volumes de dados, geralmente utilizando técnicas de inteligência artificial e aprendizado de máquina para processá-los [Mongwe et al. 2020, Handoko et al. 2022]. No contexto brasileiro, estudos recentes investigam os procedimentos de detecção de fraudes e o impacto das tecnologias na consciência anti-fraude no setor público [Silva et al. 2020].

No contexto da fiscalização de despesas municipais, a literatura apresenta diversos trabalhos que abordam a identificação de irregularidades em licitações e gastos governamentais. Por exemplo, Braz et al. (2023) propuseram metodologias para analisar irregularidades em licitações públicas, com foco em empresas de pequeno porte. Um dos métodos verifica se o faturamento anual dos licitantes excede o limite permitido, enquanto outro investiga possíveis vínculos com outras entidades jurídicas. Já Oliveira et al. (2023b) introduziram um conjunto de 19 trilhas de auditoria para identificar diferentes tipos de fraudes e irregularidades em processos licitatórios. Além disso, Silva et al. (2023) propõem um arcabouço para desambiguação de itens licitados, e análises estatísticas para detectar sobrepreço nesses itens, ainda nessa linha, Costa et al. (2024) desenvolveram um sistema de consulta de itens de compras públicas e detecção de sobrepreço.

Especificamente na detecção de anomalias em despesas municipais, Gomide et al. (2023) propuseram uma heurística baseada em mineração de dados para identificar municípios cujas despesas estão significativamente acima da média de municípios com características semelhantes. Essa abordagem agrupa municípios com base nas faixas po-

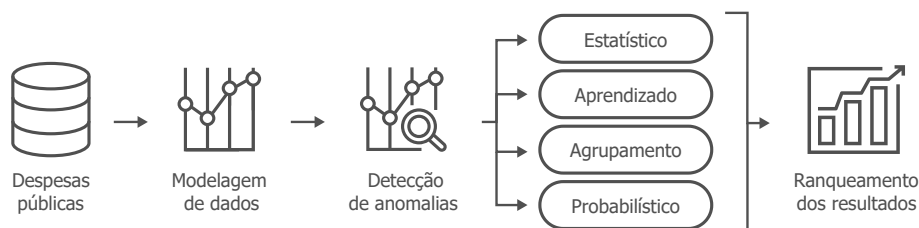


Figura 1. Fluxo da detecção de anomalias em séries temporais de despesas.

pulacionais do Censo 2010² e em sua localização geográfica, permitindo a detecção de padrões suspeitos de gasto público. Além disso, Oliveira et al. (2023a) propuseram métodos para avaliar a qualidade dos dados de receitas e compras públicas municipais.

No campo da análise de séries temporais, diferentes abordagens têm sido exploradas para detecção de anomalias. Mendes et al. (2023) aplicaram o algoritmo Soft-DTW k-means para agrupar séries temporais de faturamento de empresas cujos sócios realizaram doações para campanhas políticas, com o objetivo de identificar possíveis favorecimentos em processos licitatórios. Já Blázquez-García et al. (2021) realizaram uma revisão abrangente sobre técnicas de detecção de anomalias em séries temporais, destacando métodos não supervisionados. No contexto de previsão de séries temporais, Oliveira et al. (2024) investigaram o uso de aprendizado de máquina para prever preços de quartos de hotéis, analisando a eficiência dos diferentes métodos.

Mais recentemente, Zamanzadeh Darban et al. (2024) apresentaram uma revisão sistemática sobre técnicas avançadas de aprendizado profundo aplicadas à detecção de anomalias em séries temporais. O estudo fornece uma taxonomia baseada nas estratégias de detecção utilizadas, bem como nos modelos de aprendizado profundo empregados. Essa revisão destaca a importância da combinação de diferentes técnicas para melhorar a identificação de padrões atípicos em séries temporais complexas.

Com base nesses estudos, este trabalho propõe uma abordagem híbrida para a detecção de anomalias em séries temporais de despesas municipais, combinando métodos estatísticos, aprendizado de máquina e técnicas probabilísticas. Diferentemente das abordagens existentes, a metodologia proposta não apenas identifica anomalias, mas também as prioriza por meio de um ranqueamento, permitindo uma auditoria mais direcionada.

3. Metodologia

Nesta seção, apresentamos a metodologia adotada para avaliar o desempenho de diferentes métodos de detecção de anomalias em séries temporais. O objetivo é identificar os casos mais anômalos nos dados temporais, aplicando diferentes técnicas de detecção de anomalias. Por fim, os resultados de cada método são agregados e ranqueados para identificar os casos mais críticos. A abordagem segue três etapas principais ilustradas na Figura 1: modelagem dos dados (Seção 3.1), detecção de anomalias em séries temporais (Seção 3.2) e ranqueamento dos resultados (Seção 3.3).

²Censo 2010 - IBGE: <https://censo2010.ibge.gov.br/>

3.1. Modelagem de Dados

No modelo orçamentário brasileiro, a dotação orçamentária é uma verba prevista para despesas públicas específicas, cuja existência é obrigatória para qualquer pagamento.³ Ela é composta por quatro classificações: institucional, funcional, de natureza da despesa e por fonte de recursos. A abordagem proposta considera as classificações funcionais, de forma que as despesas são agrupadas a nível de função. A função se relaciona com a missão institucional do órgão como, por exemplo, educação, saúde, cultura, transporte, direitos da cidadania, etc.

Dessa forma, a modelagem proposta é fundamentada na teoria dos conjuntos e utiliza cinco conjuntos base que representam as entidades de interesse: Despesa Pública (D), Ano (A), Mês (M), Cidade (C), Função (F). A seguir, são detalhadas tais entidades e suas definições formais.

- $D = \{d \mid d \in \text{Despesa Pública}\}$: representa todas as despesas públicas;
- $M = \{m \mid m \in \text{Meses}\}$: conjunto de todos os meses;
- $C = \{c \mid c \in \text{Cidade}\}$: conjunto de todas cidades analisadas;
- $F = \{f \mid f \in \text{Função}\}$: conjunto de todas funções da dotação orçamentaria.

Além disso, cada despesa pública d possui os seguintes atributos:

- $d[m] \in M$: mês da despesa pública;
- $d[c] \in C$: cidade da despesa pública;
- $d[f] \in F$: função da despesa pública;
- $d[v] \in \mathbb{R}^+$: valor da despesa pública.

No contexto do presente trabalho, uma série temporal de despesas públicas corresponde a uma sequência, ordenada ao longo do tempo, dos valores dessas despesas. Como cada despesa está associada a uma cidade e uma função específica, tem-se uma série temporal para cada par cidade-função, definida como $S_{c,f} = \text{ord}_m(\{d \in D \mid d[c] = c \wedge d[f] = f\})$, onde a função ord_m retorna a sequência das despesas ordenadas crescentemente pelo atributo mês.

3.2. Detecção de Anomalias

Os métodos de detecção de anomalias em séries temporais buscam identificar padrões ou valores atípicos que desviam do comportamento esperado [Zamanzadeh Darban et al. 2024]. Cada método de detecção de anomalia (md_x) está associado a uma função valor ($fv_x(d)$) e uma função limiar ($fl_x(D, \dots)$). A função valor calcula uma métrica específica para cada item da série temporal, e a função limiar define um limite máximo ou mínimo para determinar se um item é anômalo. A função valor é sempre aplicada individualmente a cada item d , enquanto a função limiar pode levar em consideração um contexto mais amplo, como toda a série ou múltiplas séries temporais.

Uma despesa d é considerada uma anomalia em um método md_x se o resultado de sua função valor exceder o resultado da função limiar, i.e., $d \in A_x \iff fv_x(d) > fl_x(D, \dots)$, onde A_x é o conjunto de anomalias identificadas pelo método md_x .

³<https://www12.senado.leg.br/manualdecomunicacao/guia-de-economia/dotacao-orcamentaria>

Para a abordagem proposta, foram selecionados nove métodos de detecção de anomalia que se dividem em quatro categorias principais: *estatístico*, *aprendizado*, *agrupamento* e *probabilístico*. Na Tabela 1, são apresentados os nove métodos de detecção de anomalia considerados neste trabalho, junto às definições formais da função valor e da função limiar de cada um. Em seguida, cada método é descrito em detalhes, destacando suas principais características e vantagens, bem como suas funções valor e limiar.

Tabela 1. Métodos de detecção de anomalia, junto às definições formais de suas funções valor e limiar.

x	Método md_x	Função valor $fv_x(d)$	Função limiar $fl_x(D, \dots)$
<i>Estatístico</i>			
1	Z-Score (ZSC)	$fv_1(d) = d[v]$	$fl_1(D, f, m) = \mu(D^{f,m}[v]) + 3\sigma(D^{f,m}[v])$
2	Interquartile Range (IQR)	$fv_2(d) = d[v]$	$fl_2(D, f, m) = q3(D^{f,m}[v]) + 1.5\,iqr(D^{f,m}[v])$
3	Seasonal Decomposition Using Moving Averages (SDMA)	$fv_3(d) = rd(d[v])$	$fl_3(D, c, f) = \mu(rd(D^{c,f}[v])) + 3\sigma(rd(D^{c,f}[v]))$
4	Level Shift Anomaly Detection (LSAD)	$fv_4(d) = d[v]$	$fl_4(D, w, c, f) = q3(D^{w,c,f}[v]) + 1.5\,iqr(D^{w,c,f}[v])$
<i>Aprendizado</i>			
5	Isolation Forests (ISF)	$fv_5(d) = cce(d[v])$	$fl_5(D, c, f) = \mu(cce(D^{c,f}[v]))$
6	Long Short-Term Memory (LSTM)	$fv_6(d) = ep(d[v])$	$fl_6(D, c, f) = \mu(ep(D^{c,f}[v])) + 3\sigma(ep(D^{c,f}[v]))$
7	Local Outlier Factor (LOF)	$fv_7(d) = -lrd(d[v])$	$fl_7(D, d, c, f) = -\mu(lrd(D^{d,c,f}[v]))$
<i>Agrupamento</i>			
8	Soft-DTW k-Means (SDKM)	$fv_8(d) = d[v]$	$fl_8(D, k, f, m) = \mu(D^{k,f,m}[v]) + 3\sigma(D^{k,f,m}[v])$
<i>Probabilístico</i>			
9	Autoregressive Integrated Moving Average (ARIMA)	$fv_9(d) = ep(d[v])$	$fl_9(D, c, f) = \mu(ep(D^{c,f}[v])) + 3\sigma(ep(D^{c,f}[v]))$

Z-Score (ZSC). O Z-Score é um método estatístico que calcula a distância entre um valor e a média em termos de desvios padrão. Esse método é muito simples, sensível a pontos fora da curva e não é eficaz para dados não normais. Nessa metodologia, valores com um Z-Score superior a 3 são considerados anomalias. Portanto a função valor ($fv_1(d)$) é o próprio valor da despesa ($d[v]$). E há um limiar para cada par função - mês (f, m), dessa forma, a função limiar ($fl_1(D, f, m)$) é calculada considerando um sub conjunto de todas despesas de uma mesma função e mês ($D^{f,m} \subset D$). Portanto a função limiar é a média dos valores dos elementos do sub conjunto ($\mu(D^{f,m}[v])$), mais 3 vezes o desvio padrão dos valores dos elementos do sub conjunto ($\sigma(D^{f,m}[v])$). Ou seja, os dados são agrupados por função e mês e são considerados anomalia os registros com Z-Score maior que 3 dentro de cada grupo.

Interquartile Range (IQR). O IQR é definido como a diferença entre o terceiro quartil (Q3) e o primeiro quartil (Q1). Esse método utiliza essa diferença para detectar anomalias. Ele é mais robusto que o Z-Score, pois não é afetado por pontos fora da curva. Assim

como no método do Z-Score, a função valor ($fv_2(d)$) é o próprio valor da despesa, ou seja, o elemento da série temporal ($d[v]$). E há um limiar para cada par função - mês (f, m). Dessa forma, a função limiar ($fl_2(D, f, m)$) é calculada considerando um subconjunto de todas as despesas de uma mesma função e mês ($D^{f,m} \subset D$). Portanto, a função limiar é o terceiro quartil dos valores dos elementos do subconjunto ($q3(D^{f,m}[v])$), mais 1,5 vezes o IQR dos valores dos elementos do subconjunto ($iqr(D^{f,m}[v])$).

Seasonal Decomposition Using Moving Averages (SDMA) [Hyndman et al. 2018, Chapter 6]. Este método utiliza a decomposição sazonal na série de dados para decompor a série em tendência, sazonalidade e resíduos, e, em seguida, aplica médias móveis aos resíduos para detectar anomalias. Ele é eficiente para dados com tendências e sazonalidades fortes. Nessa metodologia, apenas os registros classificados como resíduos são considerados na detecção de anomalias. Portanto, a função valor ($fv_3(d)$) é a função $rd(d[v])$ que retorna o próprio valor do $d[v]$ se ele for considerado um resíduo, ou retorna zero caso contrário. E há um limiar por série temporal, ou seja, para cada par cidade - função (c, f). Dessa forma, a função limiar ($fl_3(D, c, f)$) é calculada considerando um subconjunto de todas as despesas de uma mesma cidade e função ($D^{c,f} \subset D$). Portanto, a função limiar é a média dos valores dos elementos resíduos do subconjunto ($\mu(rd(D^{c,f}[v]))$) mais 3 vezes o desvio padrão dos valores dos elementos resíduos do subconjunto ($\sigma(rd(D^{c,f}[v]))$).

Level Shift Anomaly Detection (LSAD) [Hethu Avinash et al. 2024]. Este método detecta anomalias utilizando uma técnica para detectar mudanças bruscas em uma série temporal. Essas mudanças ocorrem quando há uma alteração repentina e significativa no valor médio da série temporal. A série temporal é dividida em janelas deslizantes, que são intervalos de tempo consecutivos e sobrepostos. Cada janela (w) tem um tamanho fixo, que é definido pelo usuário. O cálculo da anomalia é feito agrupando os valores dentro de cada janela. Dessa forma, a função valor ($fv_4(d)$) é o próprio valor da despesa ($d[v]$), e há um limiar para cada janela de cada série temporal, ou seja, um limiar para cada trio janela - cidade - função (w, c, f). Dessa forma, a função limiar ($fl_4(D, w, c, f)$) é calculada considerando um subconjunto de todas as despesas de uma mesma janela, cidade e função ($D^{w,c,f} \subset D$). Portanto, a função limiar é o terceiro quartil dos valores dos elementos do subconjunto ($q3(D^{w,c,f}[v])$) mais 1,5 vezes o IQR dos valores dos elementos desses subconjunto ($iqr(D^{w,c,f}[v])$).

Isolation Forest (ISF) [Liu et al. 2008]. Este método utiliza árvores de isolamento para detectar anomalias. Cada árvore é treinada para isolar uma instância de dados, e as instâncias que são isoladas com mais frequência são consideradas anomalias. O número de divisões necessárias para isolar uma anomalia determina o comprimento do caminho (cc) até o nó folha da instância. O cálculo da anomalia é feito comparando o comprimento do caminho esperado para uma instância (cce), com a média do comprimento do caminho entre todas as árvores de isolamento ($\mu(cc)$). Dessa forma, a função valor ($fv_5(d)$) é o comprimento do caminho esperado para a despesa ($cce(d[v])$). E há um limiar para cada série temporal, ou seja, um limiar para cada par cidade - função (c, f). Portanto, a função limiar ($fl_5(D, c, f)$) é calculada considerando um subconjunto de todas as despesas de uma mesma cidade e função ($D^{c,f} \subset D$). Sendo assim, a função limiar é a média do comprimento do caminho entre todas as árvores de isolamento ($\mu(cc(D^{c,f}[v]))$).

Long Short-Term Memory (LSTM) [Malhotra et al. 2015]. O LSTM é um tipo de rede neural recorrente que é capaz de aprender padrões complexos em séries temporais. O LSTM é treinado para prever o próximo valor em uma série temporal, e o erro de previsão (ep) é utilizado no cálculo de detecção de anomalias. A ideia é que o modelo aprende o padrão e tenta prever o valor seguindo esse padrão. Se o valor real estiver muito distante do valor previsto, pode indicar uma anomalia. Portanto, a função valor ($fv_6(d)$) é o erro de previsão do valor da despesa ($ep(d[v])$). E há um limiar para cada série temporal, ou seja, um limiar para cada par cidade - função (c, f). Dessa forma, a função limiar ($fl_6(D, c, f)$) é calculada considerando um subconjunto de todas as despesas de uma mesma cidade e função ($D^{c,f} \subset D$). Portanto, a função limiar é a média dos erros de previsão dos valores dos elementos do subconjunto ($\mu(ep(D^{c,f}[v]))$) mais 3 vezes o desvio padrão dos erros de previsão dos valores dos elementos do subconjunto ($\sigma(ep(D^{c,f}[v]))$).

Local Outlier Factor (LOF) [Breunig et al. 2000]. O LOF é um método que calcula a densidade local de cada ponto de dados e compara com a densidade local de seus vizinhos. O cálculo da anomalia é feito calculando a densidade local (lrd) de cada ponto de dados, e os que possuem um lrd menor que os seus vizinhos são considerados anomalias. Portanto, a função valor ($fv_7(d)$) é o oposto da densidade local do valor da despesa ($-lrd(d[v])$). E há um limiar para cada despesa de cada série temporal, ou seja, um limiar para cada trio despesa - cidade - função (d, c, f). Portanto, a função limiar ($fl_7(D, d, c, f)$) é calculada considerando um subconjunto de todas as despesas que são vizinhas de outra despesa e são da mesma cidade e função ($D^{d,c,f} \subset D$). Sendo assim, a função limiar é o oposto da média da densidade local de todas as despesas vizinhas de d na mesma série temporal ($-\mu(lrd(D^{d,c,f}[v]))$). Tanto para a função valor quanto para a função limiar, são usados o oposto dos valores, pois, na nossa abordagem, para uma despesa ser considerada anomalia, sua função valor deve ser maior que a função limiar. No LOF, para a despesa ser considerada anomalia, sua densidade deve ser menor que a dos vizinhos. Assim, para inverter o sinal da desigualdade, é necessário usar o oposto dos valores.

Soft-DTW k-Means (SDKM) [Cuturi et al. 2017]. O Soft-DTW k-Means é um algoritmo de clusterização que combina a técnica de Dynamic Time Warping (DTW) com o algoritmo k-Means. O DTW é uma técnica utilizada para comparar sequências de séries temporais. O k-Means é um algoritmo de clusterização que divide um conjunto de dados em k grupos baseados na similaridade entre os dados. No contexto dessa metodologia, é feita uma rodada de agrupamento para cada função. Dessa forma, cada série estará agrupada em um grupo (k). O cálculo da anomalia é feito agregando os valores por grupo (k). Então, a função valor ($fv_8(d)$) é o próprio valor da despesa ($d[v]$), e há um limiar para cada grupo de cada função e de cada mês, ou seja, um limiar para cada trio grupo - função - mês (k, f, m). Dessa forma, a função limiar ($fl_8(D, k, f, m)$) é calculada considerando um subconjunto de todas as despesas de um mesmo grupo, função e mês ($D^{k,f,m} \subset D$). Portanto, a função limiar é a média dos valores dos elementos do subconjunto ($\mu(D^{k,f,m}[v])$) mais 3 vezes o desvio padrão dos valores dos elementos do subconjunto ($\sigma(D^{k,f,m}[v])$).

Autoregressive Integrated Moving Average (ARIMA) [Hyndman et al. 2018, Chapter 8]. O ARIMA é um modelo estatístico que combina três componentes: autoregressivo, integrado e média móvel. O ARIMA é utilizado para modelar a série temporal e prever valores futuros, e a partir daí, detectar anomalias. Para detectar anomalias, precisamos

comparar o valor real da série temporal com o valor previsto pelo ARIMA. Se o erro de previsão (ep) entre os dois valores for maior que um certo limite, podemos considerar o valor observado como uma anomalia. Portanto, a função valor ($fv_9(d)$) é o erro de previsão do valor da despesa ($ep(d[v])$). E há um limiar para cada série temporal, ou seja, um limiar para cada par cidade - função (c, f). Dessa forma, a função limiar ($fl_9(D, c, f)$) é calculada considerando um subconjunto de todas as despesas de uma mesma cidade e função ($D^{c,f} \subset D$). Portanto, a função limiar é a média dos erros de previsão dos valores dos elementos do subconjunto ($\mu(ep(D^{c,f}[v]))$) mais 3 vezes o desvio padrão dos erros de previsão dos valores dos elementos do subconjunto ($\sigma(ep(D^{c,f}[v]))$).

3.3. Ranqueamento dos Resultados

Esta seção descreve a estratégia de ranqueamento adotada para agregar os resultados de cada método, proposta originalmente por Oliveira et al. (2023) e adaptada ao contexto de despesas municipais. Para cada despesa, é gerado um indicador de alerta a partir da combinação dos resultados individuais de cada método. Esse indicador, denominado I_d , é calculado para a despesa d como a soma das anomalias detectadas por cada método x , multiplicada pelo logaritmo, na base 10, do valor da despesa. O valor da despesa é incluído no cálculo, pois a auditoria de despesas de maior valor são consideradas mais prioritárias. Formalmente, tal indicador é calculado pela Equação 1.

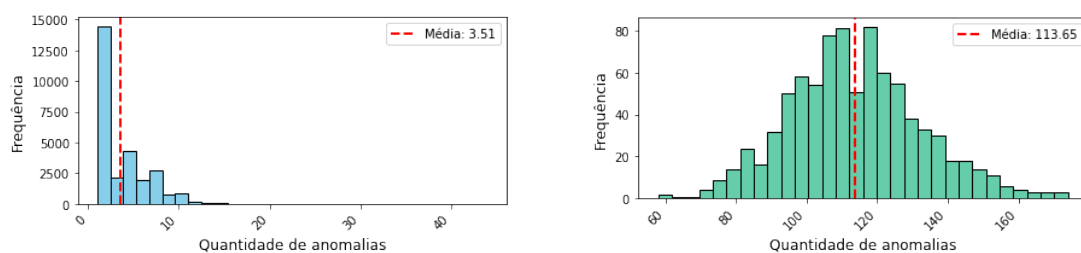
$$I_d = \log_{10}(d[v]) \cdot \sum_{x=1}^9 \begin{cases} 1, & \text{se } d \in A_x; \\ 0, & \text{caso contrario} \end{cases} \quad (1)$$

Onde $d[v]$ representa o valor da despesa e A_x é o conjunto de despesas identificadas como anômala pelo método x . Após o cálculo do indicador de alerta I_d , para cada despesa d , as despesas são ranqueadas de forma decrescente de acordo com o valor de I_d . O conjunto resultante, \hat{D} , contém as despesas ranqueadas com base no indicador de alerta. Dessa forma, temos que $\hat{D} \subseteq D$ e $\hat{D} = \{d \mid \forall d \implies I_d > I_{(d+1)}\}$.

4. Análise Experimental

Para avaliar a abordagem proposta, foi realizado um experimento considerando um conjunto de dados de despesas municipais do Estado de Minas Gerais, conforme detalhado na Seção 4.1. Em seguida, na Seção 4.2, os resultados dos métodos propostos são caracterizados e analisados, a fim de aprofundar a compreensão sobre elas. Por fim, na Seção 4.3, a análise experimental é complementada ao examinar qualitativamente as características do processo de ranqueamento das despesas com base nos indicadores de alerta.

O estudo [Gomide et al. 2023] conduziu um experimento de detecção de anomalia em despesas municipais na mesma base de dados, portanto foi utilizado como *baseline* para avaliar a abordagem proposta. Nesse estudo, são propostos dois métodos de detecção, que funcionam agrupando os municípios através de heurísticas. Uma vez agrupados, é realizado um cálculo estatístico para identificar os municípios com despesas consideravelmente superiores à média para seu respectivo grupo populacional. O primeiro método (BL_{fx}) utiliza uma heurística que agrupa os municípios com base na faixa populacional do Censo 2010. No entanto, para esse experimento as faixas foram atualizadas conforme o Censo 2022. Já o segundo método (BL_{geo}) utiliza uma heurística que faz o agrupamento com base na localização geográfica do município.



(a) Quantidade de anomalias por série temporal

(b) Quantidade de anomalias por município

Figura 2. Frequência da quantidade de anomalias.

4.1. Conjunto de Dados

Os dados utilizados neste artigo foram extraídos do Sistema Informatizado de Contas dos Municípios (SICOM),⁴ onde há uma base de dados abertos mantida pelo Tribunal de Contas do Estado de Minas Gerais (TCE-MG)⁵. Essa base contém registros detalhados de transparência fiscal de 853 municípios mineiros, incluindo informações sobre licitações, contratos, notas fiscais, despesas e receitas. Para este experimento, foram consideradas 1.045.189 registros de despesas municipais no período de janeiro de 2018 a julho de 2024. As despesas foram agrupadas por função orçamentária, conforme definido pelo modelo orçamentário brasileiro, totalizando 29 funções distintas. Cada município, portanto, possui 29 séries temporais, uma para cada função, permitindo uma análise temporal dos padrões de gastos públicos.

4.2. Análise Quantitativa

Nesta seção, apresentamos os resultados quantitativos do experimento, destacando a frequência e distribuição das anomalias identificadas nas séries temporais de despesas. A Figura 2 ilustra dois histogramas que representam a quantidade de anomalias (a) por série temporal e (b) por município.

Na análise por série temporal (Figura 2(a)), a média de anomalias é 3,51. Visto que cada série contém 79 meses de dados, uma média 3,51 meses com anomalias corresponde a 4,44% da serie. Algumas séries apresentaram índices significativamente mais altos, com três casos em que mais de 40% dos valores foram identificados como anômalos. Esses casos podem indicar problemas de qualidade na base de dados, corroborando resultados de estudos anteriores que analisaram o mesmo conjunto de dados [Oliveira et al. 2023a].

Na análise por município (Figura 2(b)), cada cidade possui 29 séries temporais de 79 registros cada, totalizando 2.291 despesas analisadas por município. A média de anomalias detectadas por cidade foi de 113,65, representando 4,96% das despesas analisadas. O município com a maior quantidade de anomalias apresentou 174 registros classificados como anômalos, o que equivale a 7,60% de suas despesas.

Na Tabela 2, as células nas laterais representam a quantidade total de anomalias detectadas por cada método, tanto os da abordagem proposta (md_x) quanto os do *baseline* (BL_{fx} , BL_{geo}), os valores entre parênteses indicam o percentual de anomalias em

⁴SICOM: <https://portalsicom1.tce.mg.gov.br/>

⁵Os dados brutos do SICOM estão disponíveis publicamente em <https://dadosabertos.tce.mg.gov.br/>

Tabela 2. Quantidade de anomalias detectadas por método, incluindo *baselines*.

Método	ZSC	IQR	SDMA	LSDA	ISF	LSTM	LOF	SDKM	ARIMA	Total
BL_{fx}	60	232	3.250	1.726	3.872	3.819	2.895	4.675	3.633	19.258 (1,842%)
BL_{geo}	74	298	1.246	1.315	1.764	1.690	1.412	2.012	1.572	28.143 (2,692%)
Total	174 (0,017%)	1.858 (0,178%)	16.113 (1,542%)	35.064 (3,355%)	26.374 (2,523%)	24.583 (2,352%)	24.839 (2,377%)	20.595 (1,970%)	23.353 (2,234%)	

Tabela 3. Top 10 despesas anômalas de acordo com o indicador de alerta (I_d).

#	Cidade	Função	Data	Valor Despesa	I_d
1	Belo Horizonte	Transporte	07/2023	248.194.900	58,76
2	Montes Claros	Administração	08/2018	20.938.310	51,24
3	Extrema	Administração	03/2022	16.196.690	50,46
4	Itabirito	Desporto e Lazer	08/2023	15.266.440	50,28
5	Congonhas	Habitação	03/2023	11.324.470	49,38
6	Poços de Caldas	Encargos Especiais	12/2018	8.273.214	48,42
7	Itabirito	Comércio e Serviços	02/2022	7.127.600	47,97
8	Cambuí	Administração	03/2023	5.880.813	47,38
9	Extrema	Judiciária	06/2019	5.665.788	47,27
10	Araporã	Administração	09/2020	5.392.563	47,12

relação ao total de 1.045.189 despesas analisadas. As demais representam a interseção da quantidade de anomalias detectadas entre os métodos da abordagem com os do *baseline*.

Os métodos Z-score e IQR identificaram um número muito inferior de anomalias em comparação com os demais algoritmos, o que pode ser atribuído à sua simplicidade. Em contrapartida, os outros métodos apresentaram quantidades de anomalias na mesma ordem de grandeza, sugerindo maior robustez na detecção de padrões atípicos.

No total, foram classificadas como anômalas 87.779 despesas, i.e., 8,40% do conjunto de dados. Desse total, 44.844 despesas (4,29%) foram consideradas anômalas por pelo menos um dos métodos do *baseline*. Além disso, 9.425 despesas foram detectadas simultaneamente pela abordagem proposta e pelo *baseline*, o que indica que 21,01% das anomalias identificadas pelo *baseline* também foram capturadas pela abordagem proposta.

4.3. Análise Qualitativa

Os resultados qualitativos deste experimento focam na priorização das despesas anômalas mais críticas, utilizando o indicador de alerta (I_d) definido na Equação 1. A Tabela 3 apresenta, a título de exemplificação, as despesas anômalas mais críticas em um ranqueamento de tamanho 10, de acordo com o indicador I_d .

Conforme mencionado na Seção 4.2, a interseção entre as anomalias identificadas pela abordagem proposta e aquelas detectadas pelos métodos do *baseline* é de apenas 21,01%. No entanto, essa sobreposição aumenta significativamente quando analisamos as anomalias mais críticas no topo do ranqueamento.

A Figura 3 ilustra a taxa de interseção entre os métodos à medida que o tamanho do ranqueamento aumenta. Observa-se que, para um ranqueamento de tamanho até 5, a interseção atinge 100%, ou seja, todas as despesas ranqueadas também foram detecta-

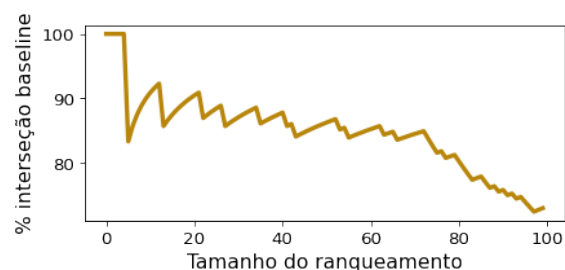


Figura 3. Percentual de interseção entre o *baseline* e a abordagem proposta para diferentes tamanhos de ranqueamento.

das pelo *baseline*. À medida que o tamanho do ranqueamento aumenta, essa interseção diminui gradativamente, chegando a 73% em um ranqueamento de tamanho 100.

No geral, esses resultados indicam que a abordagem proposta, aliada ao mecanismo de ranqueamento, é eficaz na priorização dos casos mais relevantes. Isso possibilita um direcionamento mais eficiente dos esforços de auditoria, permitindo que os recursos sejam alocados para as despesas com maior potencial de irregularidade, em consonância com estudos prévios sobre detecção de anomalias em gastos públicos.

5. Conclusão

Este trabalho propôs uma abordagem de detecção de anomalias em séries temporais de despesas municipais e uma forma de ranqueamento dos casos mais críticos, dessa forma nossa contribuição serve para identificar e priorizar casos para uma investigação e uma auditoria mais eficientes. A abordagem emprega uma combinação de técnicas avançadas para detecção de anomalia em séries temporais, incluindo análise estatística, aprendizado de máquina e técnicas probabilísticas.

Para validar a eficácia da abordagem, realizamos um experimento com um conjunto de dados extenso de despesas públicas municipais de Minas Gerais. Os resultados demonstram que a abordagem é capaz de identificar casos com alto potencial de irregularidade, especialmente quando combinados com o método de ranqueamento. Esses resultados sugerem que a abordagem pode ser uma ferramenta valiosa para auditores, oferecendo uma análise de triagem inicial diante do grande volume de dados disponíveis. Além disso, a abordagem proposta ressalta a importância de uma abordagem integrada e multifacetada para a análise de dados na auditoria, permitindo um levantamento automatizado de possíveis irregularidades em despesas municipais.

Limitações e Trabalhos Futuros. A principal limitação deste trabalho é a falta de dados rotulados para uma análise mais precisa dos resultados. Portanto, como trabalhos futuros, planeja-se fazer avaliações com auditores especialistas para analisar as anomalias ranqueadas. Além disso, pretende-se explorar técnicas de detecção de anomalia em séries temporais multidimensionais e executar a abordagem em um nível mais detalhado, analisando cada órgão público dentro de um município.

Agradecimentos. Ao Ministério Público do Estado de Minas Gerais (MPMG) pelo apoio através do Programa Capacidades Analíticas. Ao CNPq e Amazon Web Services (AWS) pelo financiamento recebido através do projeto da parceria entre ambos. Por fim, à CAPES e FAPEMIG pelo apoio aos pesquisadores envolvidos.

Referências

- Blázquez-García, A. et al. (2021). A review on outlier/anomaly detection in time series data. *ACM Computing Surveys*, 54(3):1–33.
- Braz, C. S. et al. (2023). Análise de irregularidades em licitações públicas com foco em empresas de pequeno porte. In *WCGE*, pages 94–105. SBC.
- Breunig, M. M. et al. (2000). Lof: identifying density-based local outliers. *SIGMOD Rec.*, 29(2):93–104.
- Costa, L. G. L. et al. (2024). Quanto Custa: Banco de Preços de Compras Públicas do Estado de Minas Gerais. In *SBBD DS-CoPS*. SBC.
- Cuturi, M. et al. (2017). Soft-dtw: a differentiable loss function for time-series. In *ICML*, pages 894–903. PMLR.
- Gomide, L. D. et al. (2023). Mineração de dados sobre despesas públicas de municípios mineiros para gerar alertas de fraudes. In *SBBD*, pages 378–383. SBC.
- Handoko, B. L. et al. (2022). The effect of skepticism, big data analytics to financial fraud detection moderated by forensic accounting. In *ICEEG*, pages 59–66.
- Hethu Avinash, D. et al. (2024). Integrating level shift anomaly detection for fault diagnosis of battery management system for lithium-ion batteries. *IEEE Access*, 12:116071–116084.
- Hyndman, R. et al. (2018). *Forecasting: Principles and Practice*. OTexts, 2nd edition.
- Liu, F. T. et al. (2008). Isolation forest. In *ICDM*, pages 413–422. IEEE Comp. Society.
- Malhotra, P. et al. (2015). Long short term memory networks for anomaly detection in time series. In *ESANN*.
- Mendes, B. M. A. et al. (2023). Impacto de Doações Eleitorais no Faturamento de Empresas: Um Estudo nas Eleições Municipais em Minas Gerais. In *SBBD*, pages 420–425. SBC.
- Mongwe, W. T. et al. (2020). The efficacy of financial ratios for fraud detection using self organising maps. In *SSCI*, pages 1100–1106. IEEE.
- Oliveira, F. B. et al. (2024). Machine learning and time series analysis to forecast hotel room prices. In *BRACIS*, pages 358–371. Springer.
- Oliveira, G. P. et al. (2023a). Assessing data quality inconsistencies in brazilian governmental data. *Journal of Information and Data Management*, 14(1).
- Oliveira, G. P. et al. (2023b). Ranqueamento de licitações públicas a partir de alertas de fraude. In *BraSNAM*, pages 1–12. SBC.
- Silva, C. A. et al. (2020). Consciência, Prevenção e Detecção contra Fraude: Estudo com Auditores Internos e Contadores do Setor Público Brasileiro. In *USP International Conference on Accounting*.
- Silva, M. O. et al. (2023). Análise de sobrepreço em itens de licitações públicas. In *WCGE*, pages 118–129. SBC.
- Zamanzadeh Darban, Z. et al. (2024). Deep learning for time series anomaly detection: A survey. *ACM Comput. Surv.*, 57(1).