Time Series Forecasting of Public Revenues and Expenditures: A Comparative Analysis of Individual, Local, and Global Approaches

Ruam Pastor[1], Hemir da C. Santiago[2], Paulo S. G. de Mattos Neto[1], Bernardo Silva[3], Newton Cerezini[3]

[1] Centro de Informática (CIn) – Universidade Federal de Pernambuco (UFPE)

[2]Escola Politécnica de Pernambuco (POLI) – Universidade de Pernambuco (UPE)

[3]Secretaria de Planejamento, Gestão e Desenvolvimento Regional de Pernambuco (SEPLAG-PE)

{rercp,psgmn}@cin.ufpe.br, hemir.santiago@upe.br, {bernardo.silva,newton.cerezini}@seplag.pe.gov.br

Abstract. *Public revenue and expenditure forecasting is essential for fiscal planning, yet remains relatively underexplored from a machine learning perspective. This study investigates the performance of three learning paradigms for multiple time series forecasting, individual, local, and global, applied to real public finance data from the State of Pernambuco, Brazil. The paper employs a range of traditional statistical models, machine learning methods, and foundation models, encompassing 13 different combinations. The results indicate that no single approach is universally superior, as performance depends on the specific dataset. While the evaluated approaches performed similarly for revenues, the global approach showed better results for expenditures. These findings highlight how foundation models can complement domain-specific training, providing practical evidence for choosing the most suitable forecasting strategy in public finance.*

1. Introdução

A previsão de receitas e despesas públicas desempenha papel fundamental no planejamento governamental e na sustentabilidade fiscal dos entes federativos [Congressional Budget Office 2024]. Estimativas confiáveis permitem definir limites orçamentários, apoiar a formulação de políticas públicas, orientar investimentos e reduzir riscos associados à execução financeira. A crescente disponibilidade de dados históricos provenientes dos sistemas corporativos do governo cria oportunidades para a adoção de técnicas avançadas de análise de dados e aprendizado de máquina capazes de complementar ou aprimorar métodos tradicionais de projeção orçamentária [Buettner and Kauder 2010]. A previsão de receitas e despesas pode ser formulada como um problema de séries temporais, no qual observações históricas são utilizadas para identificar tendências, sazonalidades e dependências estatísticas capazes de auxiliar a estimativa de valores futuros. Métodos clássicos, como os modelos da família de Box & Jenkins, permanecem amplamente utilizados devido à sua interpretabilidade e sólida fundamentação estatística [Box et al. 2015]. Entretanto, avanços recentes em aprendizado de máquina têm demonstrado que modelos como *Support Vector Regression* (SVR) [Barchi et al. 2026], redes neurais *Multi-Layer Perceptron* (MLP) [Masini et al. 2023], modelos de aprendizado profundo [Padilha et al. 2025]

métodos baseados em árvores de decisão [Petropoulos et al. 2022] e modelos de fundação [Ansari et al. 2024] podem alcançar desempenho superior em cenários caracterizados por não linearidade e heterogeneidade de padrões.

Nos últimos anos, um dos principais debates da literatura de previsão de múltiplas séries temporais tem se concentrado na escolha do paradigma de aprendizado mais adequado [Makridakis et al. 2022]. Nesse contexto, três abordagens têm se destacado na literatura [Makridakis et al. 2018]: individual, local e global. A abordagem individual ajusta um modelo específico para cada série temporal, explorando exclusivamente suas características particulares. A abordagem local treina um modelo a partir de um grupo de séries temporais com comportamentos semelhantes, permitindo capturar padrões temporais comuns dentro de subconjuntos de dados. Já a abordagem global treina um único modelo sobre todas as séries temporais disponíveis, explorando regularidades compartilhadas e favorecendo a transferência de conhecimento entre diferentes séries temporais [Montero-Manso and Hyndman 2021]. As tradicionais competições M4 [Makridakis et al. 2018] e M5 [Makridakis et al. 2022] evidenciaram o potencial dos modelos globais em bases extensas. Contudo, outros estudos [Montero-Manso and Hyndman 2021, Bandara et al. 2020, Ansari et al. 2024] indicam que abordagens locais podem apresentar vantagens em conjuntos heterogêneos ou compostos por séries curtas. Apesar dos avanços, não há consenso sobre qual paradigma favorece domínios fragmentados e heterogêneos. Essa lacuna é crítica nas finanças públicas, onde as séries apresentam escalas distintas, influências do calendário fiscal e baixa previsibilidade temporal.

Este trabalho investiga o desempenho relativo das abordagens individual, local e global na previsão de séries temporais de receitas e despesas públicas do estado de Pernambuco, avaliando modelos estatísticos clássicos (ARIMA), métodos de aprendizado de máquina (Regressão Ridge, SVR, MLP, *Random Forest* e *Light Gradient Boosting Machine* (LightGBM)) e modelos de fundação (Chronos). O objetivo central é identificar como a heterogeneidade e a granularidade dos dados fiscais influenciam a eficácia de cada paradigma de aprendizado. As principais contribuições deste estudo são: **(1) Comparação sistemática de paradigmas:** realiza uma avaliação entre modelos individuais, locais e globais em um domínio de alta complexidade e heterogeneidade, como as finanças públicas governamentais; **(2) Avaliação de modelos de fundação:** investiga o potencial de generalização do modelo *Chronos* em séries fiscais, analisando sua capacidade complementar aos modelos treinados nos dados do domínio; **(3) Análise de desempenho por categoria fiscal:** identifica que, enquanto abordagens individuais e locais mantêm competitividade em séries de receita, a abordagem global demonstra maior robustez e consistência na previsão de despesas.

O restante deste artigo está organizado da seguinte forma: a Seção 2 revisa a literatura sobre previsão fiscal e aprendizado de máquina, enquanto a Seção 3 detalha os paradigmas de previsão para múltiplas séries temporais. A Seção 4 descreve a metodologia proposta e a Seção 5 apresenta o protocolo experimental. Os resultados são discutidos na Seção 6 e, por fim, a Seção 7 encerra o trabalho com as considerações finais e sugestões para pesquisas futuras.

Os dados e códigos utilizados ao longo do trabalho estão disponíveis em repo-

sitório público[1]. Ferramentas de inteligência artificial foram utilizadas como apoio na implementação e organização do código deste trabalho[2].

2. Previsão de Finanças Públicas

A previsão de receitas e despesas governamentais constitui um problema clássico da economia fiscal, desempenhando papel central na formulação de políticas públicas, no planejamento orçamentário e na sustentabilidade das contas públicas. Nesse contexto, métodos sistemáticos e formalizados de previsão tendem a reduzir erros e aumentar a transparência dos processos orçamentários, contribuindo para uma gestão fiscal mais eficiente e previsível [Buettner and Kauder 2010]. De forma complementar, Frankel (2011) identifica evidências recorrentes de viés otimista em projeções fiscais oficiais, enquanto Jonung e Larch (2006) argumentam que a utilização de previsões independentes pode atuar como importante mecanismo de fortalecimento institucional e mitigação de vieses políticos [Jonung and Larch 2006].

O avanço das técnicas de aprendizado de máquina tem ampliado as possibilidades de modelagem de fenômenos econômicos complexos. Em contraste com abordagens econométricas tradicionais, esses métodos são capazes de capturar relações não lineares, interações de alta ordem e padrões temporais complexos presentes em grandes volumes de dados [Duarte et al. 2025]. No campo das finanças públicas, modelos baseados em aprendizado de máquina podem produzir ganhos substanciais de precisão em previsões fiscais regionais [Batóg and Batóg 2021]. De maneira semelhante, revisões apontam que abordagens híbridas, combinando modelos estatísticos e algoritmos de aprendizado de máquina, apresentam elevado potencial para lidar com a volatilidade, a sazonalidade e a heterogeneidade frequentemente observadas em séries fiscais e orçamentárias [Masini et al. 2023].

Paralelamente, o desenvolvimento de arquiteturas de aprendizado profundo tem transformado a área de previsão de séries temporais ao possibilitar o aprendizado automático de representações complexas em conjuntos de dados massivos [Ansari et al. 2024]. Esses avanços estimularam o surgimento de modelos globais capazes de explorar padrões compartilhados entre múltiplas séries temporais, frequentemente superando abordagens locais em cenários com grande disponibilidade de dados [Masini et al. 2023, Ansari et al. 2024]. Apesar desse progresso, a aplicação dessas técnicas ao contexto da previsão orçamentária governamental ainda permanece relativamente limitada. Em particular, estudos que exploram séries fiscais altamente desagregadas, abrangendo múltiplas categorias de receitas e despesas, continuam escassos, especialmente em países em desenvolvimento e no contexto brasileiro.

3. Paradigmas de Aprendizado para Múltiplas Séries Temporais

A previsão de múltiplas séries temporais [Makridakis et al. 2018, Makridakis et al. 2022, Aksu et al. 2024] tornou-se uma das áreas mais ativas da pesquisa em aprendizado de máquina nos últimos anos. Diferentemente da previsão univariada tradicional, em que cada série é modelada de forma independente, os cenários modernos frequentemente envolvem centenas ou milhares de séries correlacionadas, criando oportunidades para compartilhamento

[1]Código, dados e resultados deste trabalho estão disponíveis em: `https://github.com/Ruamemanuel/Time-Series-Forecasting-of-Public-Revenues-and-Expenditures`.

[2]Os autores assumem integral responsabilidade pelos resultados, análises e conclusões apresentados neste trabalho.

de informação entre diferentes processos temporais [Montero-Manso and Hyndman 2021]. Nesse contexto, a literatura organiza os métodos de previsão em três abordagens principais: individual, local e global [Montero-Manso and Hyndman 2021, Bandara et al. 2020]. Na abordagem individual, cada série recebe um modelo próprio treinado exclusivamente com seus dados históricos, preservando características específicas, mas podendo sofrer com escassez de dados quando as séries são curtas ou altamente ruidosas.

Em contraste, a abordagem global utiliza um único modelo treinado simultaneamente sobre todas as séries disponíveis, explorando padrões compartilhados para aumentar a capacidade de generalização e reduzir problemas de sobreajuste. Esse paradigma ganhou destaque após as competições M4 e M5, nas quais modelos globais baseados em *gradient boosting* e redes neurais profundas alcançaram desempenhos competitivos em diversos cenários de previsão [Makridakis et al. 2018, Makridakis et al. 2022]. Entre esses extremos encontra-se a abordagem local, que busca equilibrar especialização e compartilhamento de conhecimento por meio do agrupamento de séries semelhantes e do treinamento de modelos específicos por *cluster* [Bandara et al. 2020]. O sucesso dessa estratégia depende da capacidade de identificar padrões efetivamente compartilhados entre as séries.

Um estudo recente [Aksu et al. 2024] mostrou que não existe uma abordagem universalmente superior para previsão de múltiplas séries temporais. O desempenho de modelos individuais, locais e globais depende de fatores como o número de séries disponíveis, seu comprimento, o grau de heterogeneidade dos dados e a existência de padrões compartilhados [Montero-Manso and Hyndman 2021]. Evidências da competição M5 indicam que abordagens globais tendem a ser competitivas em bases extensas, enquanto modelos individuais permanecem robustos em cenários mais heterogêneos [Makridakis et al. 2018, Makridakis et al. 2022]. Além disso, o equilíbrio entre capacidade de generalização e especialização continua sendo um dos principais desafios da área [Petropoulos et al. 2022, Masini et al. 2023].

4. Metodologia Proposta

A Figura 1 ilustra a metodologia proposta, que se estrutura em duas fases: Treinamento e Previsão. Na etapa de treinamento, o conjunto de treinamento dos dados brutos D, que compreende todas as séries temporais, é submetido a procedimentos de pré-processamento para gerar o conjunto D_{pr}. Este conjunto refinado é estruturado de forma a viabilizar a modelagem computacional e a extração de padrões temporais pelas abordagens individual, local e global. Nesse sentido, o teste de Ljung-Box [Box et al. 2015] é aplicado ao conjunto D_{pr}. Séries temporais que não apresentam correlação linear estatisticamente significativa são excluídas da modelagem. Após a geração de D_{pr}, as funções de auto-correlação (FAC) e auto-correlação parcial (FACP) são utilizadas para definir os valores candidatos de $lags$ (k).

Na fase de previsão, uma janela temporal de entrada $D_q = \{d_q, d_{q-1}, \ldots, d_{q-k}\}$, previamente submetida às etapas de pré-processamento, é apresentada aos modelos individual ($\mathrm{M_I}$), local ($\mathrm{M_L}$) e global ($\mathrm{M_G}$). Com base nessa janela, cada modelo gera uma previsão para o horizonte h, denotada por $\hat{D}_{q+h}$.

4.1. Abordagem individual

Na abordagem individual, cada série temporal é tratada como uma tarefa de previsão independente. Os modelos são treinados exclusivamente com os dados históricos da

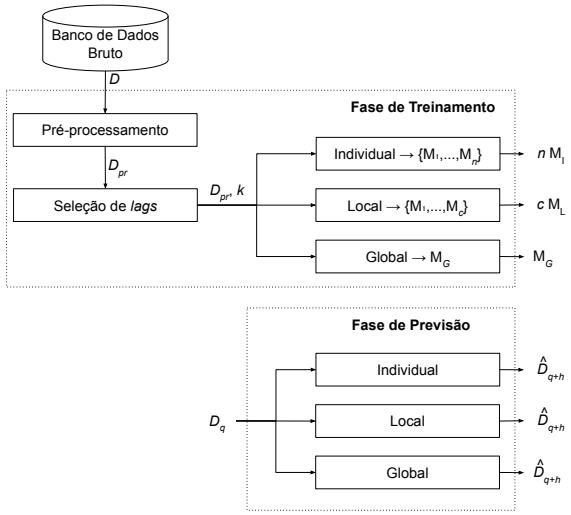


Figura 1. Metodologia proposta para avaliação das abordagens individual, local e global. Fases de treinamento e teste.

própria série, sem incorporar informações provenientes das demais séries disponíveis, ou dados exógenos. Para realizar previsões de múltiplos passos à frente, foi adotada a estratégia recursiva [Duarte et al. 2025], em que cada valor previsto em t é utilizado como entrada para a previsão do próximo instante $t + 1$, até que o horizonte h seja alcançado.

4.2. Abordagem Local

A abordagem local recebe como entrada o conjunto de treinamento (D_{pr}). A partir desse conjunto, a proposta gera um vetor compacto de atributos estatísticos com o objetivo de cobrir diferentes dimensões que determinam a similaridade estrutural e a previsibilidade das séries temporais [Aghabozorgi et al. 2015]. As características são: a magnitude da série temporal é capturada pelo logaritmo da média absoluta, que separa séries de escalas muito diferentes. A dispersão é medida pelo coeficiente de variação, que quantifica a volatilidade. A tendência aparece na inclinação linear normalizada pela escala, indicando crescimento ou queda. A dependência serial é descrita pelas autocorrelações nos *lags* 1, 2 e 3, que quantificam a memória temporal. Adicionalmente, a sazonalidade é capturada pela força sazonal, uma medida que reflete a importância relativa dos padrões cíclicos recorrentes presentes na série temporal. Por fim, a forma da distribuição é representada pelo coeficiente de assimetria, que permite identificar comportamentos marcados por concentrações assimétricas, picos ocasionais ou eventos extremos, frequentemente observados em fluxos fiscais e orçamentários.

A escolha desse conjunto de atributos busca equilibrar capacidade descritiva e parcimônia. Embora compacto, o vetor resultante incorpora informações complementares sobre nível, variabilidade, tendência, dependência temporal, sazonalidade e distribuição dos dados. Após isso, um modelo é treinado para cada grupo gerado a partir das séries que compõem o *cluster*.

4.3. Abordagem Global

A abordagem global treina um único modelo sobre todas as séries simultaneamente [Montero-Manso and Hyndman 2021]. A ideia central é que um modelo exposto

a diferentes comportamentos aprende padrões comuns a todas elas e, ao mesmo tempo, distingue cada série quando recebe informação sobre sua identidade. O modelo recebe como entrada: a série temporal (D_{pr}); atributos de calendário (mês e trimestre); e a identidade da série, dada por um identificador numérico. O identificador é o elemento que permite ao modelo aprender ajustes específicos de cada série sem abrir mão dos padrões compartilhados, o que torna a abordagem global competitiva mesmo entre séries heterogêneas [Montero-Manso and Hyndman 2021, Makridakis et al. 2022].

5. Protocolo Experimental

Os conjuntos de dados provêm do Sistema de Gerenciamento Orçamentário (SGO) de Pernambuco e reúnem séries temporais mensais de receitas realizadas e despesas liquidadas, de janeiro de 2019 a outubro de 2025. As Receitas Realizadas (RR) somam 367 séries, cada uma referente a uma fonte de receita orçamentária. As Despesas Liquidadas (DL) somam 97 séries, que englobam diferentes tipos de despesa. Após o teste de ruído branco, restaram 133 séries em RR e 19 em DL. Entre as séries temporais de despesas restantes, três eram compostas por conjunto de treinamento com menos de 15 observações, o que inviabiliza a validação cruzada dos modelos individuais e, por isso, foram excluídas, restando 16 séries para a experimentação. Cada série temporal foi dividida em 80% das observações para o conjunto de treinamento e 20% das observações para o conjunto de teste.

A comparação contempla sete modelos que cobrem o espectro consolidado da área de previsão de séries temporais: do estatístico clássico aos modelos de fundação [Masini et al. 2023]. O modelo ARIMA é o *benchmark* estatístico clássico [Box et al. 2015]. Como os *lags* de uma série temporal podem ser fortemente correlacionados entre si, a Regressão Ridge foi escolhida porque acrescenta uma penalização ℓ_2 que regulariza esse efeito e estabiliza as estimativas sob multicolinearidade [Safari et al. 2025]. O SVR usa uma margem ε-insensível que filtra ruído e, com diferentes *kernels*, pode modelar relações lineares e não-lineares [Duarte et al. 2025]. O MLP é um aproximador universal de funções, incluído para capturar não-linearidades [Duarte et al. 2025]. O *Random Forest* (RF) é um *ensemble* por *bagging*, robusto a *outliers* [Breiman 2001]. O LightGBM é um *ensemble* por *boosting*, modelo global de referência na competição M5 [Makridakis et al. 2022]. O Chronos representa os modelos de fundação, que transferem conhecimento via pré-treinamento em larga escala [Ansari et al. 2024]. A Tabela 1 indica quais modelos foram avaliados em cada abordagem. A individual usa ARIMA, MLP, Ridge e SVR; a local usa LightGBM, MLP, Ridge e SVR; e a global usa LightGBM, MLP, Ridge, RF e o Chronos. O ARIMA e o SVR ficam fora da global porque não acomodam naturalmente o identificador de série, mecanismo central do treinamento global [Montero-Manso and Hyndman 2021], e o RF foi reservado à abordagem global, onde o maior volume de dados justifica seu custo. O Chronos é pré-treinado sobre um amplo corpus de séries e aplicado em modo *zero-shot*, sem nenhum ajuste aos dados deste estudo. Sua previsão segue o mesmo protocolo recursivo dos demais modelos. Os *lags* para cada série temporal foram selecionados a partir da FACP [Box et al. 2015]. O maior *lag* significativo define a janela de entrada, com o teto de 12 meses. Esse teto tem dupla justificativa: 12 *lags* cobrem um ciclo sazonal anual completo, a periodicidade dominante em séries fiscais regidas pelo calendário orçamentário, e limitar a janela mantém o número de atributos baixo frente ao treino curto destas séries, reduzindo o sobreajuste. Todos os modelos foram treinados para prever de 3 a 12 passos à frente, dependendo do tamanho da

série.

Tabela 1. Espaço de busca de hiperparâmetros por modelo. Os mesmos valores são usados em todas as abordagens em que o modelo é utilizado.

Modelo	Hiperparâmetro	Valores
ARIMA	p, q, d	$\{0, 1, 2, 3\}$
Ridge	α	{0,01; 0,1; 1; 10; 100}
SVR	C	{0,1; 1; 10; 100}
	ε	{0,01; 0,1; 1}
	kernel	{RBF, linear}
MLP	camadas	$\{(50), (100), (50{,}50), (100{,}50)\}$
	α	$\{10^{-4}, 10^{-3}, 10^{-2}\}$
	η	{0,001; 0,01}
RF	Estimadores	$\{100, 300, 500\}$
	profundidade	$\{\infty, 5, 10\}$
	min_samples_leaf	$\{1, 5, 10\}$
LightGBM	*num_leaves*	$\{15, 31, 63\}$
	min_child_samples	$\{5, 20, 50\}$
Chronos	modo	*zero-shot* (sem ajuste)

As métricas de avaliação são calculadas no conjunto de teste. A métrica primária é a mediana da Raiz do Erro Quadrático Médio (RMSE) normalizado pela magnitude média do período de treino de cada série, o que torna os erros comparáveis entre séries de escalas diferentes [Hewamalage et al. 2023]. Utiliza-se a mediana, e não a média, por ser mais robusta à assimetria das distribuições, complementada pela contagem de vitórias por série. Adicionalmente, o *Mean Absolute Percentage Error* (MAPE) foi calculado por ser interpretável e independente de escala.

6. Resultados

Os resultados da previsão das séries temporais de receitas e despesas são apresentados nas Seções 6.1 e 6.2, respectivamente.

6.1. Avaliação de desempenho para as séries temporais de receita

A Tabela 2 apresenta os resultados obtidos pelas abordagens individual, local e global na previsão de séries temporais de receita, considerando as métricas RMSE, MAPE, desvio padrão (DP) e a frequência de vitórias de cada modelo. Na abordagem individual, o modelo ARIMA apresentou o melhor desempenho mediano, alcançando os menores valores de RMSE e MAPE, enquanto a MLP registrou a maior frequência de vitórias, sendo o melhor modelo em 48 séries (36,1%). Esse resultado mostra que o ARIMA foi mais estável, enquanto a MLP obteve o melhor resultado em mais conjuntos de dados. Na abordagem local, o LightGBM obteve o menor RMSE e a maior frequência de vitórias (39,1%), evidenciando maior capacidade preditiva em comparação aos demais modelos dessa categoria. O Ridge destacou-se por apresentar o menor desvio padrão (1,61) e o menor MAPE (41,2%), sugerindo maior estabilidade dos seus resultados. Para a abordagem global, o modelo Chronos alcançou os melhores resultados e a maior frequência de vitórias entre os modelos globais, sendo o melhor em 48 séries (36,1%). Os modelos LightGBM, MLP, Random Forest e Ridge apresentaram desempenhos relativamente próximos entre si. O Ridge manteve o menor desvio padrão (1,61), indicando maior consistência dos resultados.

As linhas denominadas “Melhor” representam o desempenho obtido ao selecionar, para cada série, o modelo com melhor resultado dentro de cada abordagem. É importante salientar que essa análise é feita a partir de um resultado teórico, uma vez que a avaliação é feita no conjunto de teste. Nessa análise, a abordagem global apresentou o melhor desempenho geral, com RMSE de 0,459, MAPE de 22,4% e 69 vitórias (51,9%), superando as abordagens individual (RMSE = 0,559; MAPE = 24,5%) e local (RMSE = 0,563; MAPE = 33,7%). Esses resultados evidenciam que os modelos globais (em especial o Chronos) foram capazes de capturar de forma mais eficiente os padrões compartilhados entre as séries temporais analisadas, proporcionando maior acurácia preditiva e predominância em mais da metade dos casos avaliados.

Tabela 2. Resultados em termos de RMSE, MAPE e Desvio Padrão (DP) das abordagens individual, local e global para a previsão de séries temporais de receita. Os melhores desempenhos por abordagem estão em negrito. A coluna “Vitória” indica a contagem e o percentual de séries em que cada modelo obteve o melhor desempenho dentro de sua respectiva abordagem.

Abordagem	Modelo	RMSE	DP	MAPE	Vitória
Individual	ARIMA	**0,649**	**1,57**	**27,4%**	43 (32,3%)
	MLP	0,696	9,68	30,9%	**48 (36,1%)**
	Ridge	0,727	18,10	33,1%	13 (9,8%)
	SVR	0,759	1,61	35,3%	29 (21,8%)
	Melhor	0,559	1,57	24,5%	40 (30,1%)
Local	LightGBM	**0,687**	1,65	42,8%	**52 (39,1%)**
	MLP	1,164	1,74	57,3%	14 (10,5%)
	Ridge	0,854	**1,61**	**41,2%**	45 (33,8%)
	SVR	0,958	1,78	45,9%	22 (16,5%)
	Melhor	0,563	1,63	33,7%	24 (18,0%)
Global	Chronos	**0,669**	1,63	**27,8%**	**48 (36,1%)**
	LightGBM	0,792	1,72	40,6%	18 (13,5%)
	MLP	0,794	1,69	42,5%	25 (18,8%)
	RF	0,798	1,70	48,8%	26 (19,5%)
	Ridge	0,840	**1,61**	42,5%	16 (12,0%)
	Melhor	0,459	1,60	22,4%	69 (51,9%)

Na Tabela 3 e Figura 2 observa-se que não há uma abordagem claramente dominante entre os modelos avaliados. Os três melhores desempenhos em termos de rank médio foram obtidos por modelos pertencentes a paradigmas distintos: o ARIMA, representando a abordagem individual $(5,74)$, o LightGBM local $(5,81)$ e o Chronos global $(6,08)$. Esse resultado sugere que tanto estratégias tradicionais quanto abordagens baseadas em aprendizado de máquina e modelos fundacionais podem alcançar níveis competitivos de desempenho na previsão de receitas públicas. Outro aspecto relevante é a diferença entre consistência e capacidade de obtenção do melhor resultado pontual. Embora os modelos globais RF e MLP apresentem ranks médios relativamente baixos $(6,57$ e $6,91$, respectivamente), eles registram as maiores taxas de vitória da amostra, alcançando o melhor desempenho em 16,5% e 14,3% das séries temporais. Isso indica que esses modelos tendem a produzir previsões mais acuradas para determinados grupos de séries, ainda que seu desempenho médio seja menos consistente quando considerado o conjunto completo de dados. Por outro lado, modelos como ARIMA, Chronos e LightGBM local apresentam percentuais de vitória mais modestos, mas ocupam posições superiores no ranking médio, evidenciando maior estabilidade e robustez ao longo das diferentes séries analisadas. Em conjunto, os resultados sugerem que a escolha do modelo envolve um *trade-off* entre consistência global de desempenho e capacidade de obter ganhos expressivos em subconjuntos

específicos de séries temporais.

Tabela 3. Ranking de Desempenho baseado no RMSE: posicionamento médio e % de vitórias por modelo nas séries temporais de receita. Os resultados estão ordenados pelo rank médio.

Abordagem	Modelo	Rank	Vitória (%)
Individual	ARIMA	5,74	7,5
Local	LightGBM	5,81	5,3
Global	Chronos	6,08	9,8
Individual	MLP	6,31	9,8
Local	Ridge	6,40	5,3
Global	Ridge	6,42	3,8
Global	RF	6,57	16,5
Global	MLP	6,91	14,3
Global	LightGBM	7,14	7,5
Individual	Ridge	7,14	5,3
Individual	SVR	7,29	7,5
Local	SVR	9,26	4,5
Local	MLP	9,93	3,0

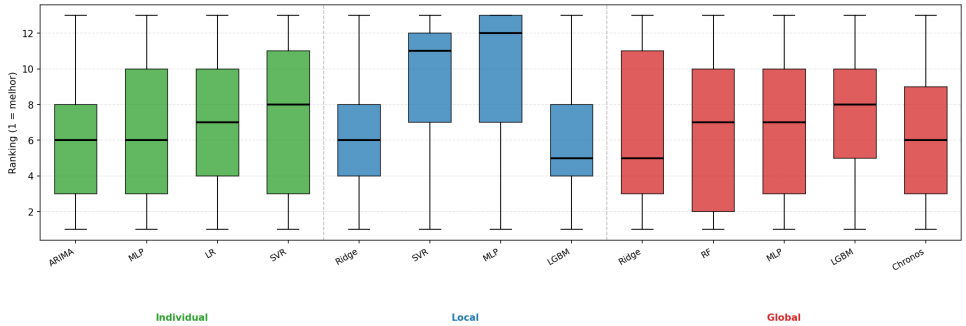


Figura 2. Distribuição do rank baseado em RMSE por modelo, agrupado por abordagem, na previsão das séries de receita.

6.2. Avaliação de desempenho para as séries temporais de despesa

A Tabela 4 apresenta os resultados obtidos pelas abordagens individual, local e global na previsão das séries temporais de despesas. Na abordagem individual, o modelo SVR apresentou o melhor desempenho geral, alcançando os menores RMSE, DP e MAPE, além de ter obtido o maior número de vitórias, sendo o melhor modelo em 7 das 16 séries analisadas (43,8%). Na abordagem local, o SVR também se destacou ao registrar o menor RMSE e a maior frequência de vitórias (31,2%). O modelo Ridge apresentou desempenho próximo, com RMSE de 0,669, mesma porcentagem de vitórias (31,2%) e menor desvio padrão (0,36), enquanto o MLP obteve o menor MAPE da abordagem (49,7%). De modo geral, os resultados indicam que entre os modelos locais, 3 dos 4 apresentaram desempenho relativamente homogêneo nas séries de despesas, com exceção do LightGBM que obteve métricas de erro maiores. Entre os modelos globais, o Chronos alcançou o melhor desempenho em termos de RMSE $(0,447)$ e foi o modelo mais vitorioso da abordagem, obtendo o melhor resultado em 9 das 16 séries (56,2%). Além disso, apresentou o menor MAPE da abordagem global (50,1%). Os demais modelos globais exibiram desempenhos próximos entre si, com destaque para o LightGBM, que registrou o menor desvio padrão (0,38). Na análise do "Melhor" entre as abordagens, o paradigma global apresentou o melhor desempenho geral, alcançando RMSE de $0,352$ e MAPE de 39,8%. Em termos de frequência de vitórias, a abordagem global também apresentou o

maior percentual (43,8%), seguida pela individual (37,5%) e pela local (18,8%). Esses resultados sugerem que os modelos globais foram mais eficazes em capturar padrões compartilhados entre as séries de despesa.

Tabela 4. Resultados em termos de RMSE, MAPE e Desvio Padrão (DP) das abordagens individual, local e global para a previsão de séries temporais de receita. Os melhores desempenhos por abordagem estão em negrito. A coluna "Vitória" indica a contagem e o percentual de séries em que cada modelo obteve o melhor desempenho dentro de sua respectiva abordagem.

Abordagem	Modelo	RMSE	DP	MAPE	Vitória
Individual	ARIMA	0,845	1,35	62,9%	2 (12,5%)
	MLP	0,788	0,61	48,0%	4 (25,0%)
	Ridge	0,746	2,00	53,8%	3 (18,8%)
	SVR	**0,691**	**0,59**	**43,3%**	**7 (43,8%)**
	Melhor	0,671	0,45	43,7%	6 (37,5%)
Local	LightGBM	0,893	0,39	59,2%	2 (12,5%)
	MLP	0,689	0,37	**49,7%**	4 (25,0%)
	Ridge	0,669	**0,36**	51,7%	**5 (31,2%)**
	SVR	**0,608**	**0,36**	52,6%	**5 (31,2%)**
	Melhor	0,547	0,34	44,1%	3 (18,8%)
Global	Chronos	**0,447**	0,83	**50,1%**	**9 (56,2%)**
	LightGBM	0,641	**0,38**	58,0%	2 (12,5%)
	MLP	0,593	0,61	62,5%	1 (6,2%)
	RF	0,680	0,40	55,7%	2 (12,5%)
	Ridge	0,589	0,55	71,2%	2 (12,5%)
	Melhor	0,352	0,42	39,8%	7 (43,8%)

Os resultados para as séries de despesa, sintetizados na Tabela 5 e na Figura 3, revelam uma clara superioridade da abordagem Global, com o modelo Chronos liderando tanto o ranking médio (4,81) quanto o percentual de vitórias (25%). Diferente do observado nas séries de receita, a abordagem global demonstrou maior consistência, ocupando quatro das cinco primeiras posições do ranking. O SVR individual destacou-se como o segundo melhor modelo (6,19), apresentando-se como a alternativa mais robusta fora do paradigma global, enquanto o LightGBM local registrou o desempenho médio mais baixo da amostra $(8,88)$. O boxplot confirma essa tendência, evidenciando que a abordagem global, liderada pelo Chronos, obteve a melhor mediana.

Tabela 5. Ranking médio de desempenho baseado no RMSE e % de vitórias por modelo nas séries temporais de despesa (ordenados pelo rank médio).

Abordagem	Modelo	Rank	Vitória (%)
Global	Chronos	4,81	25,0
Individual	SVR	6,19	12,5
Global	RF	6,25	6,2
Global	LightGBM	6,44	6,2
Local	Ridge	6,50	6,2
Global	Ridge	6,75	0,0
Individual	Ridge	6,81	6,2
Individual	MLP	6,81	6,2
Local	MLP	7,12	0,0
Local	SVR	7,38	6,2
Global	MLP	8,38	6,2
Individual	ARIMA	8,69	12,5
Local	LightGBM	8,88	6,2

7. Conclusões

Este trabalho investigou o desempenho dos paradigmas de aprendizado individual, local e global na previsão de receitas e despesas públicas, utilizando uma gama de modelos que

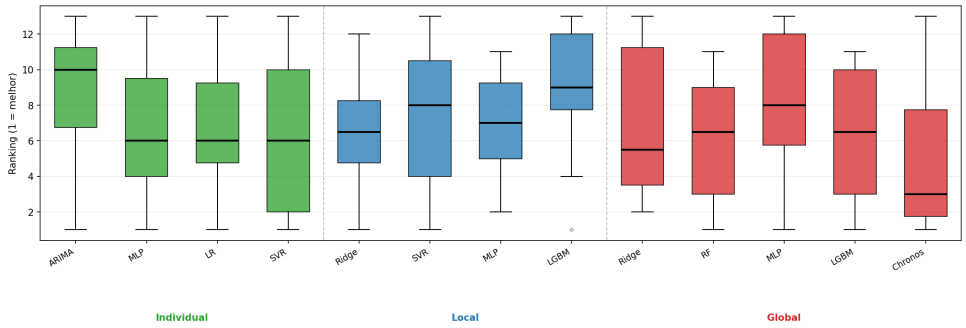


Figura 3. Distribuição do RMSE por modelo, agrupado por abordagem, na previsão das séries de despesa.

incluiu desde abordagens estatísticas clássicas até modelos de fundação. Os resultados demonstraram que não existe uma superioridade universal de um único paradigma, mas sim uma dependência direta das características das séries fiscais. Para as receitas, as abordagens individual e local mostraram-se altamente competitivas, enquanto nas despesas a abordagem global, liderada pelo modelo Chronos, apresentou maior consistência e robustez. Em resumo, os achados evidenciam que a integração de modelos de fundação com treinamentos locais e globais oferece um caminho promissor para lidar com a heterogeneidade das finanças públicas. Como trabalhos futuros, destacam-se a incorporação de variáveis exógenas e o desenvolvimento de combinações que explorem a complementaridade entre modelos de fundação e modelos treinados sobre os próprios dados [Petropoulos et al. 2022].

8. Agradecimentos

Os autores agradecem a Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco (FACEPE) pelo apoio financeiro e concessão de bolsa (APQ-0018-1.03/25). O autor Paulo Salgado agradece ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela bolsa de Produtividade em Pesquisa (processo 306934/2025-6).

Referências

Aghabozorgi, S., Seyed Shirkhorshidi, A., and Ying Wah, T. (2015). Time-series clustering — a decade review. *Information Systems*, 53:16–38.

Aksu, T., Woo, G., Liu, J., Liu, X., Liu, C., Savarese, S., Xiong, C., and Sahoo, D. (2024). Gift-eval: A benchmark for general time series forecasting model evaluation. *arXiv preprint arXiv:2410.10393*.

Ansari, A. F., Stella, L., Turkmen, C., Zhang, X., Mercado, P., Shen, H., Shchur, O., Rangapuram, S. S., Arango, S. P., Kapoor, S., et al. (2024). Chronos: Learning the language of time series. *arXiv preprint arXiv:2403.07815*.

Bandara, K., Bergmeir, C., and Smyl, S. (2020). Forecasting across time series databases using recurrent neural networks on groups of similar series: a clustering approach. *Expert Systems with Applications*, 140:112896.

Barchi, T. M., Dos Santos, J. L. F., Alves, T. A., de Mattos Neto, P. S., Stevan Jr, S. L., Corrêa, F. C., Medina, M. C. C., Siqueira, H. V., and de Oliveira, J. F. L. (2026). A genetic algorithm-based ensemble framework for wind speed forecasting. *Scientific Reports*.

Batóg, B. and Batóg, J. (2021). Regional government revenue forecasting: Risk factors and accuracy. *Risks*, 9(12).

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., and Ljung, G. M. (2015). *Time Series Analysis: Forecasting and Control*. John Wiley & Sons, Hoboken, NJ, 5 edition.

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1):5–32.

Buettner, T. and Kauder, B. (2010). Revenue forecasting practices: differences across countries and consequences for forecasting performance. *Fiscal Studies*, 31(3):313–340.

Congressional Budget Office (2024). An update to the budget and economic outlook: 2024 to 2034. Technical report, Congressional Budget Office. Accessed: 2026-06-22.

Duarte, F. C., Neto, P. S., and Firmino, P. R. (2025). Improving accuracy in short mortality rate series: Exploring multi-step forecasting approaches in hybrid systems. *arXiv preprint arXiv:2509.22395*.

Hewamalage, H., Ackermann, K., and Bergmeir, C. (2023). Forecast evaluation for data scientists: common pitfalls and best practices. *Data Mining and Knowledge Discovery*, 37(2):788–832.

Jonung, L. and Larch, M. (2006). Improving fiscal policy in the EU: the case for independent forecasts. *Economic Policy*, 21(47):491–534.

Makridakis, S., Spiliotis, E., and Assimakopoulos, V. (2018). The M4 competition: Results, findings, conclusion and way forward. *International Journal of Forecasting*, 34(4):802–808.

Makridakis, S., Spiliotis, E., and Assimakopoulos, V. (2022). The M5 accuracy competition: Results, findings, and conclusions. *International Journal of Forecasting*, 38(4):1346–1364.

Masini, R. P., Medeiros, M. C., and Mendes, E. F. (2023). Machine learning advances for time series forecasting. *Journal of Economic Surveys*, 37(1):76–111.

Montero-Manso, P. and Hyndman, R. J. (2021). Principles and algorithms for forecasting groups of time series: Locality and globality. *International Journal of Forecasting*, 37(4):1632–1653.

Padilha, G. A. G., Jung, J. J., and de Mattos Neto, P. S. (2025). Memetic algorithm-based optimization of hybrid forecasting systems for multivariate time series. *Neural Computing and Applications*, 37(17):10689–10706.

Petropoulos, F., Apiletti, D., Assimakopoulos, V., Babai, M. Z., Barrow, D. K., Ben Taieb, S., Bergmeir, C., Bessa, R. J., Bijak, J., Boylan, J. E., et al. (2022). Forecasting: theory and practice. *International Journal of Forecasting*, 38(3):705–871.

Safari, M. J. S., Arashloo, S. R., and Vaheddoost, B. (2025). Signature kernel ridge regression time series model: A novel approach for hydrological drought modeling using multi-station meteorological drought information. *Applied Soft Computing*, 178:113343.