

João Vitor Murta

Modelos de Machine Learning para Previsão de Arrecadação de ICMS em Minas Gerais

Brasília

2024

João Vitor Murta

Modelos de Machine Learning para Previsão de Arrecadação de ICMS em Minas Gerais

Dissertação apresentada ao Curso de Mestrado Acadêmico em Economia, Universidade de Brasília, como requisito parcial para a obtenção do título de Mestre em Economia

Universidade de Brasília - UnB

Faculdade de Administração Contabilidade e Economia - FACE

Departamento de Economia - ECO

Programa de Pós-Graduação

Orientador: Prof. Daniel Oliveira Cajueiro, Dr.

Brasília

2024

Resumo

Esta dissertação avalia a capacidade de modelos de machine learning na previsão da arrecadação do Imposto sobre Circulação de Mercadorias e Serviços (ICMS) em Minas Gerais, empregando uma diversidade de abordagens como LASSO, Ridge, Random Forest, Redes Neurais LSTM. A análise foi conduzida em dois períodos distintos: de Janeiro de 2019 a Dezembro de 2019 e de Setembro de 2022 a Agosto de 2023, permitindo uma avaliação abrangente diante de contextos econômicos variados. Os resultados indicam que o modelo LSTM obteve melhor desempenho no primeiro período analisado, enquanto o LASSO mostrou maior eficiência no segundo período, ressaltando a importância da seleção apropriada de modelos em resposta a diferentes cenários econômicos. Este trabalho contribui para o entendimento de como ferramentas avançadas de previsão podem auxiliar na gestão fiscal, enfatizando a necessidade de adaptação e inovação contínua em métodos de previsão tributária.

Palavras-chave: ICMS, Machine Learning, LASSO, Ridge, Random Forest, Redes Neurais, LSTM, Previsão de Arrecadação.

Abstract

This dissertation assesses the performance of machine learning models in forecasting ICMS tax revenue in Minas Gerais state, utilizing a range of approaches including LASSO, Ridge, Random Forest, and LSTM Neural Networks. The analysis was carried out over two distinct periods: from January 2019 to December 2019 and from August 2022 to September 2023, providing a comprehensive evaluation against varied economic backdrops. The findings indicate that the LSTM model excelled in the first period studied, while LASSO proved to be more efficient in the second period, underscoring the significance of selecting appropriate models in response to different economic scenarios. This work contributes to the understanding of how advanced forecasting tools can aid in fiscal management, highlighting the necessity for continuous adaptation and innovation in tax forecasting methodologies.

Keywords: Revenue Forecasting, ICMS, Machine Learning, LASSO, Ridge, Random Forest, Neural Networks, LSTM.

Lista de ilustrações

Figura 1 – LSTM	19
Figura 2 – ICMS	22

Lista de tabelas

Tabela 1 – MAPE (%) por Modelo e Horizonte de Previsão do Período de Teste entre Jan/2019 a Dez/2019.	28
Tabela 2 – MAPE (%) por Modelo e Horizonte de Previsão do Período de Teste entre Ago/2022 a Set/2023.	29
Tabela 3 – Resultados do Teste de Diebold-Mariano (Jan/2019 a Dez/2019)	30
Tabela 4 – Resultados do Teste de Diebold-Mariano (Set/2022 a Ago/2023)	30

Sumário

1	INTRODUÇÃO	11
2	MODELOS	14
2.1	ARIMA	14
2.2	LASSO	14
2.3	Ridge	15
2.4	Random Forest	16
2.5	Redes Neurais	17
2.6	Redes Neurais Recorrentes	18
2.7	LSTM	18
3	REVISÃO EMPÍRICA	20
4	DADOS	22
5	METODOLOGIA	24
6	RESULTADOS	28
7	CONCLUSÃO	31
8	REFERÊNCIAS BIBLIOGRÁFICAS	33

1 Introdução

Diversos estudos têm sido realizados com o objetivo de desenvolver modelos mais precisos e confiáveis para previsão de arrecadação, utilizando diferentes técnicas e metodologias. Previsões precisas auxiliam os agentes econômicos na tomada de decisões, pois os ajudam a estabelecer expectativas mais bem fundamentadas. A previsão de arrecadação estadual desempenha um papel fundamental no contexto da gestão fiscal dos estados brasileiros, sendo respaldada por importantes legislações.

A Lei de Responsabilidade Fiscal (Lei Complementar nº 101/2000) estabelece diretrizes para o equilíbrio das contas públicas, exigindo uma gestão responsável e transparente dos recursos arrecadados. A previsão de arrecadação é uma ferramenta essencial para o cumprimento dos limites e metas estabelecidos pela lei, como a limitação de despesas com pessoal, o equilíbrio entre receitas e despesas, e o estabelecimento de metas de resultado primário. A falta de uma previsão adequada pode comprometer o cumprimento dessas obrigações legais, gerando desequilíbrios financeiros e consequências negativas para a administração pública.

Além das legislações mencionadas, outras normativas complementares também destacam a importância da previsão de arrecadação. O Plano Plurianual (PPA), por exemplo, exige a definição de metas de arrecadação para o período de vigência do plano, orientando as ações e os investimentos do governo estadual. Já a Lei de Diretrizes Orçamentárias (LDO) estabelece as metas fiscais e os parâmetros para a elaboração do orçamento anual, dependendo diretamente de uma previsão confiável da arrecadação para sua correta formulação.

Dessa forma, a previsão de arrecadação estadual, respaldada pela legislação vigente, assume uma importância estratégica na gestão financeira dos estados brasileiros. Ela permite um planejamento mais eficiente, uma alocação adequada dos recursos públicos e a conformidade com as obrigações legais, contribuindo para a transparência, a responsabilidade fiscal e o desenvolvimento sustentável das unidades federativas.

No cenário federal, o recém-aprovado arcabouço fiscal de 2023 tem como pilar central a restrição do crescimento real das despesas, vinculado à variação da receita no período anterior. Em outras palavras, qualquer projeção dos níveis de gastos para os próximos períodos está intrinsecamente ligada à previsão da arrecadação do período imediatamente anterior. Isso enfatiza a vital importância da capacidade de antecipação na previsão da receita tributária, visto que essa antecipação desempenha um papel fundamental na determinação dos limites e diretrizes para as despesas futuras. Essa relação estreita destaca a necessidade de uma abordagem precisa e informada na estimativa de receitas, tornando

a antecipação da arrecadação de impostos um componente crucial para a eficácia do novo arcabouço fiscal.

A situação fiscal de Minas Gerais em 2023 permanece desafiadora, com um cenário de dívida e despesas consideráveis. A dívida do estado com a União está projetada para alcançar aproximadamente R\$ 210 bilhões ao fim de nove anos, o que reflete a magnitude dos desafios fiscais enfrentados pelo estado. Neste contexto, a precisão na previsão de receitas torna-se ainda mais crítica. Com previsões de receita mais precisas, o estado pode planejar de forma mais eficaz para lidar com suas obrigações e buscar formas de aumentar a receita ou reduzir despesas onde possível. A melhoria nas previsões de receita pode contribuir para uma gestão fiscal mais responsável e para a sustentabilidade financeira a longo prazo de Minas Gerais.

Entre as técnicas utilizadas para a previsão de arrecadação, destaca-se o uso de modelos econométricos de séries temporais, que buscam relacionar a arrecadação tributária com outras variáveis econômicas relevantes (VAR). Esses modelos têm sido amplamente utilizados pelos governos estaduais e federal no Brasil.

No entanto, a utilização de técnicas de machine learning para a previsão de arrecadação tem se mostrado uma alternativa promissora. Essas técnicas permitem a análise de grandes quantidades de dados históricos de arrecadação e outras variáveis, e são capazes de identificar padrões intrincados e tendências que podem passar despercebidos pelos modelos econométricos tradicionais (ARIMA).

O ARIMA (AutoRegressive Integrated Moving Average) é um modelo amplamente utilizado no contexto de previsão de séries temporais, incluindo a previsão de arrecadação de impostos. Este modelo combina elementos de autorregressão (AR), diferenciação integrada (I) e médias móveis (MA) para modelar e prever padrões temporais em dados. No contexto da arrecadação de impostos, o ARIMA pode ser aplicado para capturar tendências, sazonalidades e padrões cíclicos que influenciam a receita tributária ao longo do tempo. Essa abordagem ajuda governos e entidades tributárias a tomar decisões informadas e planejar estrategicamente.

Técnicas de aprendizado de máquina e, especialmente, algoritmos de aprendizado profundo, trouxeram abordagens inovadoras para problemas de previsão nos quais as interações entre variáveis são modeladas em uma hierarquia profunda e estratificada. Métodos baseados em aprendizado de máquina, como Lasso, Ridge e Florestas Aleatórias (RF), bem como algoritmos de aprendizado profundo, como Redes Neurais Recorrentes (RNN) e Memória de Longo Prazo de Curto Prazo (LSTM), têm sido alvo de grande atenção recentemente, especialmente por suas aplicações em diversas áreas, incluindo os campos econômico e financeiro. As técnicas de aprendizado profundo têm a capacidade de identificar a estrutura e os padrões dos dados, como não linearidades e complexidades, particularmente em previsões de séries temporais. Em destaque, o LSTM tem sido aplicado

na previsão de séries temporais, além de em dados econômicos e financeiros, como na predição da volatilidade do S&P 500.

Estudos recentes têm revelado avanços promissores tanto na previsão de séries temporais de maneira geral, quanto na arrecadação de impostos, graças à aplicação de técnicas inovadoras de machine learning. Na previsão de séries temporais, pesquisadores têm explorado o potencial de algoritmos avançados, como redes neurais recorrentes e modelos baseados em árvores de decisão, para lidar com a complexidade temporal de conjuntos de dados. Essas técnicas têm sido eficazes na previsão de variáveis como demanda de produtos, séries financeiras e eventos climáticos, proporcionando insights valiosos para tomadas de decisão informadas em setores diversos, como comércio, finanças e meteorologia.

Em paralelo, no contexto da arrecadação de impostos, estudos têm destacado a utilidade dos modelos de machine learning para melhorar a precisão nas estimativas tributárias. No Brasil, por exemplo, pesquisadores têm aplicado redes neurais artificiais e algoritmos de aprendizado de máquina na previsão do Imposto sobre Circulação de Mercadorias e Serviços (ICMS), principal tributo estadual responsável por até 90% da arrecadação em algumas unidades da federação. Os resultados têm sido encorajadores, indicando que essas técnicas podem contribuir significativamente para otimizar a gestão fiscal, proporcionando uma administração tributária mais eficiente e adaptável aos desafios econômicos em constante evolução. Essa abordagem inovadora não apenas aprimora a eficiência na arrecadação, mas também fornece insights valiosos para informar políticas públicas e estratégias econômicas.

Diante desse cenário, a presente dissertação busca contribuir com a literatura existente sobre a previsão de arrecadação estadual, avaliando a eficácia de modelos de machine learning, como LASSO, Ridge, Random Forest e redes neurais LSTM, na previsão da arrecadação do ICMS em Minas Gerais, comparando-os ao modelo ARIMA. Espera-se, com isso, identificar melhores práticas e metodologias para a aplicação dessas técnicas na previsão de arrecadação, contribuindo para o desenvolvimento de modelos mais precisos e confiáveis e para a melhoria da gestão fiscal dos estados brasileiros.

2 Modelos

2.1 ARIMA

O modelo Autoregressive Integrated Moving Average (ARIMA) é uma das ferramentas mais utilizadas para análise e previsão de séries temporais. O ARIMA é capaz de modelar uma ampla gama de sequências de tempo através da combinação de três componentes principais: autoregressivo (AR), integrado (I), e médias móveis (MA) (ENDERS, 2014).

O termo autoregressivo (AR) indica que a variável de interesse é regredida em seus próprios valores defasados (lagged). O parâmetro p denota o número de defasagens incluídas no modelo. A componente integrada (I) refere-se à diferenciação da série temporal para torná-la estacionária. O parâmetro d indica o número de diferenças necessárias. O termo médias móveis (MA) modela o erro da previsão como uma combinação linear de termos de erro passados. O parâmetro q indica o número de termos de erro que são utilizados no modelo. A fórmula geral é dada por:

$$y'_t = c + \sum_{i=1}^p \phi_i y'_{t-i} + \epsilon_t + \sum_{i=1}^q \theta_i \epsilon_{t-i}$$

onde y'_t é a série diferenciada (e que pode ser diferenciada mais de uma vez).

O modelo Seasonal ARIMA (SARIMA) estende o ARIMA para incluir padrões sazonais. Ele adiciona termos sazonais AR, I, e MA, notados como SAR, S, e SMA, respectivamente. O modelo SARIMA é representado como $ARIMA(p,d,q)(P,D,Q)_s$, onde P , D , Q são os equivalentes sazonais dos parâmetros p, d e q (s é o tempo de repetição do padrão sazonal):

$$y_t^{(d,D)} = c + \sum_{i=1}^p \phi_i y_{t-i}^{(d,D)} + \sum_{i=1}^q \theta_i \epsilon_{t-i} + \sum_{i=1}^P \Phi_i y_{t-is}^{(d,D)} + \sum_{i=1}^Q \Theta_i \epsilon_{t-is} + \epsilon_t$$

2.2 LASSO

LASSO (Least Absolute Shrinkage and Selection Operator): O LASSO é uma técnica de regularização usada em modelos de regressão linear proposta por Tibshirani (1996). O objetivo do LASSO é introduzir uma penalidade na função de custo do modelo, adicionando a soma absoluta dos coeficientes ao termo de erro. Isso leva à minimização simultânea dos resíduos e da soma dos valores absolutos dos coeficientes, resultando em

uma seleção de variáveis, pois alguns coeficientes podem ser reduzidos a zero. O parâmetro de regularização, controlado por um termo de penalização, permite ajustar o equilíbrio entre ajuste preciso e simplicidade do modelo:

$$p(\boldsymbol{\beta}, \lambda) = \lambda \|\boldsymbol{\beta}\|_1 = \lambda \sum_{i=1}^n |\beta_i|$$

No contexto de uma série temporal univariada, como a arrecadação de impostos, o LASSO é aplicado para modelar a relação entre os valores passados da arrecadação e prever futuros valores. Neste cenário, cada ponto na série temporal é modelado como uma função linear dos pontos anteriores, e o LASSO é incorporado para regularizar os coeficientes dessa função. A penalidade L1 (ℓ_1) introduzida pelo LASSO é crucial para induzir sparsity nos coeficientes, permitindo que apenas um subconjunto das observações passadas tenha um impacto significativo na previsão.

A principal vantagem do LASSO em séries temporais univariadas é sua capacidade de realizar seleção automática de características temporais relevantes. Isso é particularmente útil quando a arrecadação pode ser influenciada por fatores específicos em determinados períodos. A penalização L1 ajuda a identificar e manter apenas os lags mais informativos, simplificando o modelo e potencialmente melhorando a generalização para previsões futuras.

2.3 Ridge

A Ridge Regression é uma técnica de regularização introduzida por Hoerl e Kennard (1970), semelhante ao LASSO, mas utiliza uma penalidade L2, adicionando a soma dos quadrados dos coeficientes ao termo de erro. Quando aplicada a séries temporais univariadas, como a arrecadação de impostos, a Ridge Regression busca modelar a relação entre os valores passados da variável e prever futuros valores. A penalidade L2 ajuda a evitar overfitting, especialmente quando há multicolinearidade entre os lags, ao restringir os coeficientes sem torná-los exatamente zero:

$$p(\boldsymbol{\beta}, \lambda) = \lambda \|\boldsymbol{\beta}\|_2^2 = \lambda \sum_{i=1}^n \beta_i^2.$$

A Ridge Regression em séries temporais destaca-se por sua capacidade de lidar com multicolinearidade entre lags sucessivos. Isso é crucial em séries temporais, onde os valores passados podem estar correlacionados entre si. Ao suavizar os coeficientes, a Ridge Regression contribui para a estabilidade do modelo, proporcionando uma previsão mais robusta, mesmo em situações em que existem relações complexas entre os lags.

Assim como no LASSO, ao empregar a Ridge Regression em séries temporais, é essencial ajustar o parâmetro de regularização para encontrar o equilíbrio ideal entre ajuste e prevenção de overfitting. A Ridge Regression é particularmente útil quando se espera que vários lags desempenhem papéis importantes na previsão, e a multicolinearidade pode ser uma preocupação. Ao incorporar a Ridge Regression, a modelagem de séries temporais univariadas se beneficia da estabilidade e da capacidade de generalização, proporcionando uma abordagem sólida para a previsão de variáveis temporais como a arrecadação de impostos.

2.4 Random Forest

Random Forest, originalmente proposto por Breiman (2001), é categorizada como um modelo de machine learning baseado em árvores de decisão, destaca-se por construir uma coleção de árvores durante o treinamento, onde cada árvore é instruída com uma amostra aleatória do conjunto de dados e um subconjunto aleatório de características. Durante a fase de predição, as respostas individuais das árvores são combinadas, resultando em uma predição consolidada. Esse método, que envolve a construção de várias árvores e a agregação dos resultados, visa mitigar o overfitting.

$$\hat{y}_{t+h} = \frac{1}{B} \sum_{b=1}^B f_b(\mathbf{x}_t),$$

onde B é o número de árvores, y_{t+h} é a variável objetivo e $f_b(\mathbf{x}_t)$ é o resultado de cada árvore de regressão construída levando em conta uma subamostra bootstrap dos dados de treinamento e a randomização do subconjunto de variáveis consideradas para determinar a melhor divisão em cada nó.

O modelo Random Forest pode ser uma escolha eficaz para prever séries temporais, graças à sua capacidade de lidar com conjuntos de dados complexos e identificar padrões não lineares. As árvores de decisão incorporam dependências temporais, enquanto a introdução de aleatoriedade durante o treinamento ajuda a evitar superestimções. Além disso, a interpretabilidade do modelo permite analisar a importância relativa de cada característica (ou feature) na predição, no caso de um modelo multivariado.

Ao usar a Random Forest para previsões temporais, é crucial otimizar hiperparâmetros, como o número de árvores e a profundidade máxima. A Random Forest, com sua abordagem robusta e versátil, desempenha um papel significativo na exploração de modelos para prever a arrecadação de impostos ao longo do tempo.

2.5 Redes Neurais

As redes neurais em machine learning são modelos inspirados no funcionamento do cérebro humano, projetados para aprender padrões e realizar tarefas específicas. Elas são compostas por unidades chamadas neurônios, organizadas em camadas interconectadas. Cada conexão entre os neurônios tem um peso que é ajustado durante o treinamento do modelo. A primeira camada, chamada de camada de entrada, recebe os dados de entrada, como imagens, texto ou números. As camadas intermediárias, conhecidas como camadas ocultas, processam essas informações por meio de operações matemáticas, ajustando os pesos das conexões. Para cada neurônio em uma camada oculta, a entrada é multiplicada pelo peso associado a essa conexão. Esses produtos ponderados são então somados para obter uma única quantidade, conhecida como a soma ponderada. Após a etapa de soma ponderada, o resultado é passado através de uma função de ativação não linear. Essa função introduz não linearidades no modelo, permitindo que a rede aprenda relações mais complexas nos dados. Exemplos comuns de funções de ativação incluem a função sigmoide, a tangente hiperbólica (tanh) e a unidade linear retificada (ReLU).

Durante o treinamento, a rede neural compara a saída produzida pelas camadas ocultas com a saída desejada e ajusta os pesos das conexões para minimizar a diferença. Isso é feito através de algoritmos de otimização, como o gradiente descendente. O objetivo é encontrar os pesos que minimizam uma função de custo que quantifica a diferença entre as previsões do modelo e os valores reais. O gradiente dessa função de custo em relação aos pesos é calculado, e os pesos são ajustados na direção oposta ao gradiente, repetindo esse processo iterativamente até que a rede aprenda representações úteis para a tarefa específica. A última camada, chamada de camada de saída, produz a resposta final do modelo.

Uma vez treinada, a rede neural pode fazer previsões ou tomar decisões sobre novos dados não vistos. Redes neurais profundas, que têm várias camadas ocultas, são conhecidas como redes neurais profundas ou deep learning, e têm sido especialmente eficazes em tarefas complexas, como reconhecimento de imagem, processamento de linguagem natural e jogos.

O sucesso das redes neurais em machine learning pode ser atribuído à sua capacidade de aprender representações hierárquicas e complexas dos dados, tornando-as poderosas ferramentas em uma variedade de aplicações, desde reconhecimento de voz até diagnóstico médico e automação industrial.

2.6 Redes Neurais Recorrentes

As redes neurais recorrentes (RNNs) são um tipo especializado de arquitetura de rede neural projetada para lidar com dados sequenciais, como séries temporais. A principal característica que diferencia as RNNs de outras arquiteturas é a capacidade de processar informações temporais mantendo uma memória interna, permitindo que elas capturem dependências de longo prazo em sequências de dados.

A estrutura de uma RNN inclui unidades de processamento, chamadas de células, que mantêm uma "memória" interna ou estado oculto. A cada passo de tempo, a entrada atual e o estado oculto anterior são combinados para calcular a saída da célula e o novo estado oculto. Essa arquitetura permite que informações anteriores influenciem as previsões futuras, o que é crucial para a modelagem de séries temporais, onde padrões podem se estender por períodos variados. No entanto, as RNNs padrão podem enfrentar dificuldades em manter e utilizar informações relevantes ao longo de sequências temporais muito longas devido a um problema conhecido como "desvanecimento do gradiente" ou "explosão do gradiente".

A retropropagação (backpropagation) é um algoritmo de otimização amplamente utilizado em treinamento de redes neurais, incluindo Redes Neurais Recorrentes (RNNs). Durante o treinamento, a rede neural faz uma previsão, e a diferença entre essa previsão e o resultado real (o erro) é calculada. O algoritmo propaga esse erro de volta pela rede, atribuindo contribuições de erro a cada peso. Os pesos são então ajustados proporcionalmente a essas contribuições de erro usando gradientes descendentes. No contexto de LSTMs, o processo de backpropagation é estendido através do tempo, permitindo que a rede aprenda a manter e atualizar a "memória" de longo prazo de maneira eficaz, o que é crucial para a modelagem de sequências temporais complexas.

2.7 LSTM

A estrutura de Long Short-Term Memory (LSTM) é uma arquitetura de célula projetada para superar as limitações de dependências temporais de longo prazo em Redes Neurais Recorrentes (RNNs). Cada unidade LSTM possui uma arquitetura mais complexa do que uma unidade RNN padrão, incorporando três portas fundamentais: porta de entrada, porta de esquecimento, e porta de saída. Essas portas controlam o fluxo de informações ao longo do tempo, permitindo que as LSTMs armazenem e atualizem estados de longo prazo de maneira eficaz (HOCHREITER;SCHMIDHUBER, 1997).

A porta de entrada determina quais novas informações da entrada atual devem ser adicionadas ao estado da célula. Isso é feito através de uma operação semelhante à da porta de esquecimento, mas utilizando uma função de ativação diferente, geralmente a

tangente hiperbólica. O resultado é uma atualização proporcional ao valor de esquecimento calculado pela porta de esquecimento e à nova informação que deve ser adicionada.

A porta de esquecimento decide quais informações no estado atual da célula devem ser esquecidas ou mantidas. Essa decisão é baseada na entrada atual e na saída da porta de entrada. O processo envolve a aplicação de uma função sigmoide para a entrada atual e a multiplicação elemento a elemento com a saída da porta de entrada. Os valores resultantes são então usados para atualizar o estado da célula. Valores próximos de 0 indicam que a informação associada ao estado atual deve ser esquecida, enquanto valores próximos de 1 indicam que a informação deve ser mantida.

A porta de saída calcula a saída com base no estado atual da célula. Essa estrutura complexa permite que as LSTMs capturem dependências temporais de longo prazo, tornando-as especialmente eficazes para modelar séries temporais voláteis (HAYKIN, 1998).

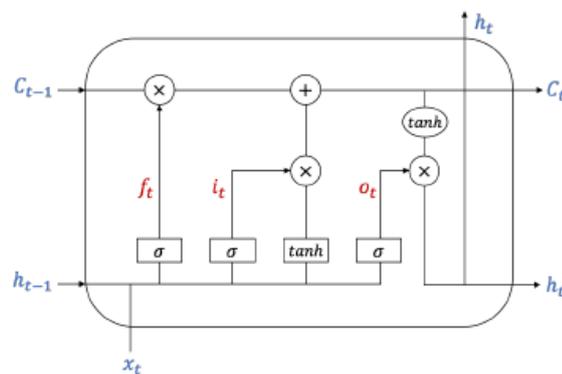


Figura 1 – LSTM

Esses mecanismos permitem que a LSTM ajuste os pesos associados a diferentes partes do estado da célula, decidindo quais informações são mais importantes para manter e quais devem ser esquecidas. O uso de funções de ativação, como a sigmoide e a tangente hiperbólica, permite que a LSTM aprenda esses pesos durante o treinamento, adaptando-se às características específicas dos dados e da tarefa em questão. Essa capacidade de aprender a esquecer ou reter informações é uma característica chave das LSTMs, tornando-as eficazes na modelagem de dependências temporais complexas em dados sequenciais.

3 Revisão Empírica

Entre os estudos que têm se dedicado à aplicação de técnicas de machine learning na previsão de arrecadação no Brasil, destaca-se o trabalho de Figueiredo e Silva (2020), que utilizou redes neurais artificiais para prever a arrecadação de ICMS no estado do Rio de Janeiro. Os autores compararam os resultados da abordagem de redes neurais com os modelos econométricos tradicionais de séries temporais e concluíram que as redes neurais apresentaram resultados superiores, com erro relativo de cerca de 1% utilizando o método Long Short-Term Memory (LSTM). Com o objetivo de comparar a eficácia da LSTM, também foi desenvolvido outro modelo de rede neural menos sofisticado, Multilayer Perceptron (MLP), que apresentou resultados superiores à previsão da SEFAZ-RJ, porém inferiores às previsões utilizando LSTM.

Em Dornelas et al (2022), os autores também utilizam o modelo LSTM para previsão de arrecadação do ICMS no estado do Rio de Janeiro, porém comparando os resultados utilizando a abordagem multi e univariada. A última realiza previsões valendo-se apenas da variável endógena, enquanto na primeira variáveis exógenas de alta correlação com o ICMS foram introduzidas no modelo. Em ambos os casos, os resultados foram mais acurados em relação às metodologias tradicionais, sendo a abordagem multivariada aquela que apresentou os menores erros de previsão.

Namini e Namin (2018) comparam duas técnicas de previsão de séries temporais de diversas variáveis econômicas e financeiras: ARIMA e LSTM. O estudo emprega dados variados, incluindo índices de mercado e indicadores econômicos, para avaliar a precisão e eficácia de ambos os métodos em previsões. Utilizando validação do tipo walk-forward, os resultados revelam que o modelo LSTM supera o ARIMA em termos de acurácia na previsão de séries temporais financeiras e econômicas em quase todas as variáveis testadas. O primeiro melhorou a previsão em 85%, em média, na comparação com o ARIMA. Isso sugere que o LSTM pode ser uma abordagem mais eficaz para modelar e prever séries temporais com comportamentos complexos e não lineares, como aquelas encontradas nas variáveis econômicas.

Este estudo é reconhecido e frequentemente citado na comunidade acadêmica, destacando-se por seu impacto na compreensão e aplicação de métodos de aprendizado profundo em previsões financeiras e econômicas. Utilizando validação cruzada do tipo walk-forward ele ressalta a importância da escolha do modelo apropriado para diferentes tipos de dados de séries temporais e abre caminho para futuras pesquisas nessa área. O artigo também aborda o impacto do número de iterações no ajuste do modelo (epochs), concluindo que variar muito este parâmetro não aumenta a precisão das previsões.

Mais recentemente, Noor, Sarlan e Aziz (2023) investigam a capacidade dos métodos de ML para realizar previsões da arrecadação federal do governo da Malásia. Baseados em uma revisão da literatura empírica de previsão de variáveis financeiras, os autores utilizaram redes neurais do tipo Feed Forward Neural Network (FFNN). Os resultados foram favoráveis ao método FFNN, gerando uma previsão mais acurada.

Medeiros e Vasconcelos (2016) mostram que, modelos high-dimensional utilizando um grande conjunto de preditores (131 variáveis macroeconômicas do Estados Unidos) geraram menores erros de previsão para algumas variáveis. Uma versão do método de Least Absolute Shrinkage and Selection Operator (LASSO) foi o que apresentou o melhor desempenho.

Entretanto, não parece ser consenso na literatura que metodologias de redes neurais recorrentes, ou mesmo outros métodos de machine learning, são a melhor opção para previsão de variáveis econômicas e financeiras. Em Krauss et al (2017), por exemplo, deep learning, gradient-boosting trees e random forests são usados para modelar o índice S&P 500. Os resultados apontaram que a modelagem baseada em deep learning apresentou um desempenho inferior. Em Chung, Williams (2022) os métodos econométricos tradicionais de séries temporais, como por exemplo Simple Exponential Smoothing (SES) e variações do ARIMA, foram capazes de realizar melhores previsões em relação à outros modelos de ML usados.

Em 2022, a SEFAZ-RS divulgou em seu portal um estudo conduzido por (DORNELLES; SCHWARTZER; BRAATZ, 2022) sobre a aplicação de redes neurais LSTM na projeção da arrecadação do ICMS no estado do Rio Grande do Sul. Os pesquisadores empregaram uma abordagem univariada, incorporando manipulações de lags e médias móveis conforme relatado por (SILVA; FIGUEIREDO, 2020), utilizando dados do ICMS de janeiro de 2009 a dezembro de 2020 para o treinamento e validação do modelo. As previsões abrangeram os meses de janeiro a agosto de 2021, sendo realizadas mensalmente, com o modelo recalibrado a cada mês utilizando dados reais até o mês anterior a cada previsão.

A arquitetura de rede adotada consistiu em duas camadas, cada uma com 80 neurônios LSTM, aplicando um dropout de 5%. O tamanho do lote (batch) utilizado foi 5. O otimizador escolhido foi o Adam, com uma taxa de aprendizado padrão de 0.001 e a função de ativação padrão - Tangente Hiperbólica. Quanto aos resultados, os autores apresentaram um erro acumulado de -2.33% ao longo dos 8 meses de previsão. Esse desempenho foi comparado a um erro acumulado de -8.24% observado nos métodos de previsão atualmente utilizados pela SEFAZ-RS, os quais não foram detalhados no artigo.

A Sessão 4 descreve os dados usados na estimação dos modelos.

4 Dados

A série temporal usada neste trabalho foi extraída dos dados de arrecadação mensal do tributo ICMS de Minas Gerais, presente no site da SEFAZ-MG, sendo composta por valores nominais de arrecadação em reais (R\$) e compreende o período de janeiro de 1997 a agosto de 2023.

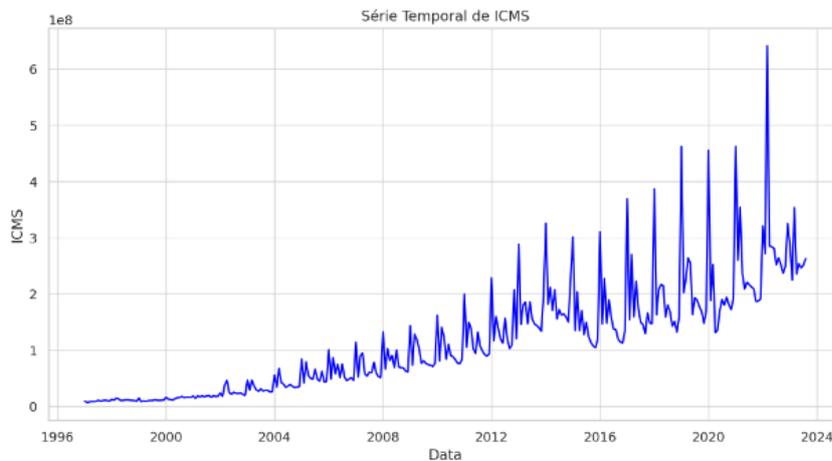


Figura 2 – ICMS

A análise dos dados da série apresenta um padrão claro de sazonalidade, complementado por uma tendência geral ascendente ao longo do tempo, considerando que os valores estão em termos correntes. No entanto, essa trajetória de crescimento é interrompida periodicamente por episódios de choques significativos, claramente identificáveis nos dados.

Esses choques, ou "ruídos", nos dados são particularmente evidentes durante períodos de instabilidade econômica e eventos extraordinários. Por exemplo, a crise financeira global de 2008/2009 emerge claramente como um período de perturbação, onde a arrecadação foi impactada negativamente pelas consequências econômicas globais. Da mesma forma, a crise interna do Brasil, que englobou uma forte recessão em 2015/2016, também se manifesta nos dados como um período de contração na arrecadação.

O impacto da pandemia de COVID-19 em 2020 é outro evento crítico identificável nos dados. Neste caso, observa-se inicialmente uma queda na arrecadação, reflexo das medidas de restrição e da desaceleração econômica. Seguiu-se, no entanto, uma recuperação na atividade econômica, juntamente com a forte alta dos preços das commodities em 2021/2022. Este último fenômeno contribuiu para um aumento na arrecadação, destacando como eventos externos e oscilações no mercado global podem ter um impacto direto e imediato nas finanças públicas.

Importante destacar, neste contexto, a redução da alíquota do ICMS sobre com-

bustíveis em julho de 2022, conforme estabelecido pela Lei Complementar nº 194. Esta mudança legislativa, combinada com uma queda nos preços globais do petróleo, teve um impacto significativo na arrecadação de ICMS no estado. Registrou-se, no segundo semestre de 2022, uma queda real de aproximadamente 10% na arrecadação de ICMS em comparação com o mesmo período de 2021. Essa redução reflete não apenas a diminuição da carga tributária sobre os combustíveis, mas também os efeitos combinados da volatilidade dos preços do petróleo no mercado internacional.

O aumento nos preços das commodities exerce um impacto significativo na arrecadação de ICMS em Minas Gerais, um fenômeno que é particularmente evidente nos setores de combustíveis e mineração, com um efeito cascata na economia local, impactando não apenas as receitas das empresas de mineração, mas também o valor arrecadado através do ICMS. No caso dos combustíveis, o preço do petróleo no mercado internacional influencia diretamente os custos de combustíveis como a gasolina e o diesel.

Além disso, o aumento nos preços dessas commodities pode estimular investimentos e expansão nos respectivos setores, criando um ciclo de crescimento econômico. Isso pode resultar em mais empregos, aumento na demanda por serviços e produtos locais, e conseqüentemente, um aumento na arrecadação de ICMS proveniente de outras fontes além dos setores diretamente impactados.

Considerando que o período de 2020 a 2023 foi marcado por diversas turbulências, incluindo a pandemia de COVID-19 e suas conseqüências econômicas, bem como outras instabilidades globais e locais, poderia ser natural esperar uma redução na capacidade preditiva dos modelos econômicos. Essas irregularidades podem obscurecer padrões e tendências anteriormente confiáveis, desafiando assim as metodologias de previsão.

Diante dessa hipótese, e considerando também a prática comum em modelagem preditiva de reservar os dados mais recentes para testes de acurácia, duas séries temporais distintas foram construídas como input para os modelos. A primeira série abrange o período de janeiro de 1997 até dezembro de 2018, com o ano de 2019 selecionado como um período de teste, visando evitar a influência pelos eventos atípicos que se seguiram.

A segunda série estende-se de janeiro de 1997 até agosto de 2022 para treinamento, separando os 12 meses subsequentes (de setembro de 2022 até agosto de 2023) para o teste. Essa separação de dados tem o objetivo de testar a robustez e a adaptabilidade dos modelos em face de variações e choques significativos, proporcionando uma compreensão mais aprofundada de sua capacidade de previsão em condições variáveis e, muitas vezes, imprevisíveis.

5 Metodologia

A validação cruzada é essencial no contexto de previsão de séries temporais, onde a sequência e a dependência entre as observações são críticas, tornando os métodos tradicionais de validação de modelos menos eficazes. Uma técnica particularmente ajustada para séries temporais é a validação cruzada do tipo "sliding window". Neste método, o conjunto de dados é segmentado em janelas temporais consecutivas, onde o modelo é treinado em uma janela de dados de tamanho fixo e testado em uma subsequente janela mais recente. Esse procedimento permite uma avaliação contínua do desempenho do modelo à medida que novos dados são incorporados, simulando a aplicação do modelo em cenários de previsão real. No contexto de previsão da arrecadação, onde as condições econômicas e os padrões tributários podem variar com o tempo, a validação cruzada do tipo "sliding window" proporciona uma análise dinâmica da capacidade do modelo de se ajustar a essas variações.

Durante o processo de validação cruzada do tipo "sliding window", em cada etapa, uma nova janela de dados é formada avançando no tempo, mantendo o tamanho da janela constante e movendo o conjunto de teste para o período imediatamente subsequente. Esse método permite que os hiperparâmetros do modelo sejam ajustados e validados de maneira iterativa, garantindo que o modelo seja avaliado em diferentes fases da série temporal.

A metodologia empregada inclui a adoção de uma janela de 60 períodos, que se desloca ao longo do conjunto de dados. Os modelos de previsão são então estimados com base nos dados contidos nesta janela de 60 períodos, sendo a estimativa dos modelos realizada para cada combinação de hiperparâmetros, permitindo a consideração de diferentes configurações do modelo.

Para otimização dos hiperparâmetros dos modelos LASSO, Ridge e Random Forest emprega-se a técnica de grid-search, que consiste na avaliação de diversas combinações de valores de hiperparâmetros com o objetivo de identificar aqueles que resultam em melhores previsões. Posteriormente, uma avaliação out-of-sample é conduzida, na qual os modelos estimados com as diferentes combinações de hiperparâmetros são utilizados para realizar as previsões. As previsões obtidas são então comparadas com os valores reais para calcular o erro quadrático médio (MSE).

Para a escolha dos valores do parâmetro de regularização do LASSO e Ridge, utilizamos uma busca por grid em uma escala logarítmica, abrangendo uma ampla gama de valores entre 0,01 e aumentando até 10, cobrindo várias ordens de magnitude. Durante a busca por grid, foi empregada a validação cruzada com a janela deslizante para avaliar o desempenho do modelo em diferentes conjuntos de dados, selecionando o valor de que

minimiza o erro de previsão médio e evita o overfitting aos dados de treinamento. O erro de previsão é medido pelo RMSE.

Para a Regressão Ridge, o procedimento é semelhante ao Lasso. A diferença é que, na primeira, todos os coeficientes podem ser reduzidos, enquanto a Regressão Lasso pode eliminar alguns deles completamente.

O número de árvores e a profundidade de cada árvore são hiperparâmetros cruciais em modelos de Random Forest. Poucas árvores podem resultar em alta variância e previsões imprecisas, enquanto muitas árvores elevam o custo computacional sem grandes ganhos de desempenho após um certo limite. Normalmente, inicia-se com poucas árvores, aumentando gradualmente e avaliando o desempenho via validação cruzada com janela deslizante. Quanto à profundidade, árvores muito profundas podem sobreajustar e falhar na generalização, ao passo que árvores rasas podem não capturar bem a estrutura dos dados. Como regra, mantemos a profundidade e o número de árvores o menor possível.

A principal série de dados utilizada como input para o modelo LSTM é a variação do acumulado em 12 meses, o que torna a série estacionária e livre dos efeitos de sazonalidade. Outras versões da série com diferenciação simples, diferentes lags e médias também foram testadas como input.

A normalização dos dados é outro passo crucial antes de executar um modelo de Rede Neural Long Short-Term Memory (LSTM), especialmente em tarefas de séries temporais. Este processo é importante porque as LSTMs são sensíveis à escala dos dados de entrada. A normalização ajuda a garantir que as características da série temporal tenham a mesma escala, permitindo que o modelo aprenda de forma mais eficiente. Sem normalização, as variáveis com escalas maiores podem dominar o processo de aprendizagem, levando a um treinamento ineficiente e a um desempenho de previsão subótimo.

Diversas configurações de arquitetura foram exploradas, variando o número de camadas e neurônios. Adicionalmente, uma ampla gama de hiperparâmetros foi avaliada, incluindo funções de ativação, tamanho de lote (batch), algoritmos para ajuste de pesos, taxa de aprendizado, entre outros. O tamanho do lote variou entre 4 e 16 em cada experimento. O número de épocas foi fixado em um valor alto, 1000 épocas. Entretanto, com o objetivo de evitar sobreajuste, foi usada a técnica early stopping. O erro quadrático médio (MSE) e o "ADAM" são usados como a função de perda e o algoritmo de otimização, respectivamente.

O algoritmo estabelece uma função denominada "fit lstm" que treina e constrói o modelo LSTM. A função recebe o conjunto de dados de treinamento, o número de épocas, isto é, a quantidade de vezes que um determinado conjunto de dados é ajustado ao modelo, e o número de neurônios.

Na elaboração do modelo preditivo ARIMA para esta pesquisa, empregou-se a bibli-

oteka pmdarima do Python, uma ferramenta sofisticada projetada para otimizar a seleção de parâmetros em modelos ARIMA. A funcionalidade central desta biblioteca, conhecida como autoarima, automatiza o processo de escolha dos parâmetros p (autoregressivo), d (diferenciação) e q (média móvel), percorrendo sistematicamente várias combinações possíveis. O algoritmo autoarima avalia cada modelo potencial com base em critérios estatísticos, incluindo o Critério de Informação de Akaike (AIC), o Critério de Informação Bayesiano (BIC) e o Critério de Informação de Hannan-Quinn (HQIC), para determinar a configuração mais eficaz. Esta metodologia automatizada permite uma seleção de modelo mais precisa e eficiente, eliminando a necessidade de uma abordagem de tentativa e erro manual extensiva.

Os experimentos foram executados em Jupiter Notebook usando pacotes padrão do Python: Keras; Statsmodels para modelos ARMA; Scikit-Learn para LASSO, Ridge Regression, LSTM e Random Forest.

A Raiz do Erro Quadrático Médio (RMSE) é uma medida que foi frequentemente utilizada para avaliar a precisão das previsões obtidas pelos modelos nos processos de treinamento e validação. Ele mensura as diferenças ou resíduos entre os valores reais e previstos. A métrica compara os erros de previsão de diferentes modelos para um conjunto de dados específico e não entre conjuntos de dados. A fórmula para calcular o RMSE é a seguinte:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

Onde n é o número total de observações, y_i representa o valor real, enquanto \hat{y}_i é o valor previsto. A principal vantagem do uso do RMSE é que ele penaliza erros grandes. Além disso, ele dimensiona as pontuações nas mesmas unidades dos valores previstos (ou seja, por mês para este trabalho).

A fórmula para calcular o MAPE (Erro Percentual Absoluto Médio) é a seguinte:

$$\text{MAPE} = \frac{100}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

Nesta fórmula, n é o número total de observações, representa o valor real e é o valor previsto. A fórmula calcula a média dos erros percentuais absolutos entre os valores reais e os previstos,

Para avaliar o desempenho dos modelos, são conduzidas previsões para diferentes horizontes temporais: o mês subsequente ($t+1$), os próximos seis meses ($t+6$), e os doze meses seguintes ($t+12$). Esta análise é aplicada a dois conjuntos distintos da série temporal. O primeiro conjunto abrange o período de janeiro de 1997 até dezembro de 2018, designando

o ano de 2019 como período de teste. O segundo conjunto estende-se de janeiro de 1997 até agosto de 2022 para o treinamento, separando os 12 meses subsequentes, de setembro de 2022 até agosto de 2023, para a fase de teste.

6 Resultados

Nesta seção, discutimos os resultados obtidos a partir da comparação entre cinco modelos distintos de previsão: ARIMA, Ridge Regression, LASSO, Random Forest, e LSTM. Os modelos foram avaliados com base no Erro Percentual Absoluto Médio (MAPE) para diferentes horizontes de previsão, especificamente para $t + 1$, $t + 6$, e $t + 12$, que correspondem, respectivamente, a previsões para o próximo mês, seis meses à frente, e doze meses à frente. Os dados foram analisados em dois períodos distintos, sendo o primeiro de janeiro a dezembro de 2019, e o segundo de setembro de 2022 a agosto de 2023. A tabela 1 mostra os resultados para o primeiro período de teste:

Modelo	Horizonte de Previsão		
	$t + 1$	$t + 6$	$t + 12$
ARIMA	9.04%	10.26%	10.78%
Ridge Regression	7.57%	8.06%	8.17%
LASSO	8.77%	10.96%	10.03%
Random Forest	11.46%	11.85%	12.62%
LSTM	6.22%	7.71%	6.47%

Tabela 1 – MAPE (%) por Modelo e Horizonte de Previsão do Período de Teste entre Jan/2019 a Dez/2019.

No período de teste de janeiro a dezembro de 2019, o modelo LSTM apresentou o desempenho mais consistente e superior, com os menores valores de MAPE para todos os horizontes de previsão, variando de 6.22% a 7.71%. Em contrapartida, o modelo Random Forest registrou os maiores valores de MAPE, indicando um ajuste menos preciso às variações dos dados. Notavelmente, o modelo Ridge Regression também demonstrou um bom desempenho, superando os modelos ARIMA e LASSO, especialmente para previsões de curto prazo.

Outra conclusão interessante é a maneira como a escolha do período de teste influencia significativamente a precisão da previsão. Isso é particularmente evidente ao analisarmos os anos de 2020 a 2022, que se mostraram atípicos por motivos descritos anteriormente.

Para o período subsequente, de setembro de 2022 a agosto de 2023, além de uma piora generalizada da capacidade de previsão dos modelos, observou-se uma alteração na performance relativa. O LASSO destacou-se com o menor MAPE para todos os horizontes de previsão, demonstrando uma melhoria significativa em comparação ao período anterior. O Random Forest continuou a apresentar o maior MAPE, o que sugere uma menor eficácia desse modelo em adaptar-se às mudanças nas tendências dos dados ao longo do tempo. Por outro lado, o LSTM, apesar de não ter alcançado a liderança como no período anterior,

Tabela 2 – MAPE (%) por Modelo e Horizonte de Previsão do Período de Teste entre Ago/2022 a Set/2023.

Modelo	Horizonte de Previsão		
	t + 1	t + 6	t + 12
ARIMA	12.17%	13.05%	13.14%
Ridge Regression	10.73%	11.24%	11.02%
LASSO	7.42%	8.07%	8.32%
Random Forest	13.95%	14.18%	15.62%
LSTM	9.54%	10.85%	10.02%

manteve um desempenho robusto, validando sua capacidade de generalização em diferentes conjuntos de dados.

A análise da redução percentual do erro de previsão dos modelos em relação ao ARIMA revela insights valiosos sobre o desempenho relativo dessas abordagens de modelagem em dois períodos distintos. Para o primeiro período (janeiro a dezembro de 2019), o modelo LSTM apresentou a maior redução percentual nos erros de previsão em todos os horizontes de previsão, com reduções de 31.19%, 24.85% e 39.98% para t + 1, t + 6, e t + 12, respectivamente. A Regressão Ridge também demonstrou melhorias significativas, com reduções variando de 16.26% a 24.21%. Por outro lado, o Random Forest mostrou um aumento no erro de previsão em comparação ao ARIMA, indicando um desempenho inferior para este período.

No segundo período (setembro de 2022 a agosto de 2023), o modelo LASSO sobressaiu com a maior redução percentual no MAPE em relação ao ARIMA, apresentando reduções de 39.03%, 38.16%, e 36.68% para os três horizontes de previsão, respectivamente. Isso indica uma melhoria substancial na precisão das previsões do LASSO sobre o modelo de benchmark. A LSTM também manteve uma performance sólida com reduções de 21.61%, 16.86%, e 23.74%, respectivamente. Curiosamente, o Random Forest continuou a apresentar um desempenho desfavorável, com um aumento no erro de previsão em relação ao ARIMA, especialmente notável para o horizonte de t + 12, com uma redução de -18.87%.

Estes resultados evidenciam as diferenças na eficácia dos modelos em função do horizonte de previsão e do período analisado. A superioridade do modelo LSTM no primeiro período de teste e a do modelo LASSO no segundo destacam a importância da seleção do modelo apropriado conforme o contexto específico e a variabilidade dos dados. Ademais, a variabilidade dos resultados entre os horizontes de previsão reforça a complexidade inerente às previsões de longo prazo.

As Tabelas 3 e 4 apresentam os resultados do teste de Diebold-Mariano das previsões dos métodos de aprendizado de máquina em comparação com o ARIMA, para ambos os períodos de referência.

Em ambas os resultados, os modelos LASSO e LSTM demonstram uma perfor-

Tabela 3 – Resultados do Teste de Diebold-Mariano (Jan/2019 a Dez/2019)

Modelo	t + 1 (DM, p-valor)	t + 6 (DM, p-valor)	t + 12 (DM, p-valor)
Ridge Regression	(-2.45, 0.014)	(-3.12, 0.002)	(-3.58, <0.001)
LASSO	(-1.10, 0.271)	(1.35, 0.177)	(-1.29, 0.197)
Random Forest	(2.58, 0.010)	(3.01, 0.003)	(3.46, <0.001)
LSTM	(-4.76, <0.001)	(-4.89, <0.001)	(-5.32, <0.001)

Tabela 4 – Resultados do Teste de Diebold-Mariano (Set/2022 a Ago/2023)

Modelo	t + 1 (DM, p-valor)	t + 6 (DM, p-valor)	t + 12 (DM, p-valor)
Ridge Regression	(-3.14, 0.002)	(-4.21, <0.001)	(-4.56, <0.001)
LASSO	(-8.75, <0.001)	(-9.82, <0.001)	(-9.06, <0.001)
Random Forest	(2.34, 0.019)	(2.87, 0.004)	(4.38, <0.001)
LSTM	(-5.61, <0.001)	(-4.98, <0.001)	(-5.92, <0.001)

mance superior ao ARIMA, como evidenciado por estatísticas DM negativas e valores-p significativamente baixos, indicando diferenças de desempenho estatisticamente significativas em vários horizontes de previsão. Esta consistência na performance superior desses modelos sugere sua eficácia em capturar padrões nos dados de forma mais eficiente do que o tradicional modelo ARIMA, independentemente do período de análise. Por outro lado, o Random Forest exibe uma performance variável, em alguns casos inferior ao ARIMA, destacando a importância do contexto e das características dos dados na escolha do modelo de previsão. A Regressão Ridge também mostra melhorias significativas sobre o ARIMA, reforçando a ideia de que técnicas modernas de aprendizado de máquina e regularização podem oferecer vantagens em termos de precisão de previsão.

7 Conclusão

Nesta dissertação, exploramos a aplicabilidade e eficácia de modelos de machine learning na previsão da arrecadação do ICMS no estado de Minas Gerais. Investigamos a eficácia de uma variedade de modelos, incluindo LASSO, Ridge, Random Forest, redes neurais e LSTM, na previsão da arrecadação do ICMS no estado de Minas Gerais, tendo como referência de comparação o modelo ARIMA. Realizamos uma análise comparativa abrangente ao longo de dois períodos distintos: Janeiro de 2019 a Dezembro de 2019 e Setembro de 2022 a Agosto de 2023.. Os resultados revelaram que, enquanto o modelo LSTM demonstrou um desempenho superior no primeiro período, evidenciando sua capacidade de adaptação a variações normais nos dados, o modelo LASSO destacou-se no segundo período, que incluiu episódios de choques econômicos significativos, sugerindo uma maior eficiência em adaptar-se a períodos atípicos.

A análise comparativa forneceu insights valiosos sobre a adaptabilidade e robustez dos modelos frente a diferentes conjunturas econômicas, realçando a importância da seleção de modelo baseada no contexto específico e na natureza dos dados. Ademais, o estudo sublinhou a limitação imposta pela disponibilidade de séries temporais extensas e variadas no contexto brasileiro, um desafio que compromete a exploração total do potencial dessas técnicas avançadas de previsão.

Como direção para trabalhos futuros, recomendamos a exploração de modelos multivariados que incorporem variáveis correlatas para enriquecer a análise e potencialmente melhorar a precisão das previsões. Adicionalmente, a comparação de modelos em um contexto internacional, com acesso a séries temporais mais longas, pode oferecer perspectivas adicionais sobre a generalização e aplicabilidade dos modelos de machine learning em diferentes cenários econômicos.

Este trabalho contribui para a literatura existente ao oferecer uma melhor compreensão sobre a aplicabilidade de modelos de machine learning avançados na previsão de receitas fiscais, um aspecto crucial para a gestão fiscal eficiente e planejamento orçamentário nos governos estaduais. As descobertas evidenciam a promessa dessas técnicas em melhorar as estratégias de previsão tributária e, por extensão, a gestão fiscal.

8 Referências Bibliográficas

Box, G. E. P., Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control* (2nd ed.). Holden-Day.

Brownlee, J. (2016). *Time Series Prediction with LSTM Recurrent Neural Networks in Python with Keras*. Retrieved from <https://machinelearningmastery.com/time-series-prediction-lstm-recurrent-neural-networks-python-keras/>

Chung, I. H., Williams, D. W., Do, M. R. (2022). For better or worse? Revenue forecasting with machine learning approaches. *Public Performance Management Review*, 45(5), 1133–1154. <https://doi.org/10.1080/15309576.2022.2073551>

Dornelas, A. F., Campos, L. D., Figueiredo, K. (2022). Modelos para Previsão Tributária Utilizando Redes Neurais LSTM.

ENDERS, W. *Applied Econometric Time Series*. Wiley, 2014. (Wiley Series in Probability and Statistics). ISBN 9781118808566. Disponível em: <<https://books.google.com.br/books?id>

Fischer, T., Krauss, C. (2017). Deep learning with long short-term memory networks for financial market predictions. *FAU Discussion Papers in Economics*, No. 11/2017. Friedrich-Alexander University Erlangen-Nuremberg, Institute for Economics.

Gers, F. A., Schmidhuber, J., Cummins, F. (2000). Learning to Forget: Continual Prediction with LSTM. *Neural Computation*, 12(10), 2451-2471.

HAMILTON, J. D. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, [Wiley, Econometric Society], v. 57, n. 2, p.357–384, 1989. ISSN 00129682, 14680262.

Hastie, T., Tibshirani, R., Friedman, J. (Data de publicação não fornecida). *The Elements of Statistical Learning*.

Haykin, S. (1999). *Neural Networks: A Comprehensive Foundation* (2nd ed.). Prentice-Hall.

Hochreiter, S., Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8), 1735-1780.

Hoerl, A.; Kennard, R. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, v. 12, n. 1, p. 55–67, 1970.

MASINI, R. P.; MEDEIROS, M. C.; MENDES, E. F. Machine learning advances for time series forecasting. *Journal of economic surveys*, Wiley Online Library, v. 37, n. 1, p.76–111, 2023.

Medeiros, M. C., Vasconcelos, G. F. R. (2016). Forecasting macroeconomic variables in data-rich environments. *Economics Letters*, 138(C), 50-52.

Namini, S. S., Namin, A. S. (2018). Forecasting economic and financial time series: ARIMA vs LSTM. Retrieved from <https://arxiv.org/ftp/arxiv/papers/1803/1803.06386.pdf>

Patterson, J. (2017). *Deep Learning: A Practitioner's Approach*. O'Reilly Media.

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, 61, 85-117.

Silva, P. F., Figueiredo, K. (2020). *Aprendizado Profundo Aplicado na Previsão de Receita Tributária Utilizando Variáveis Endógenas*. ENIAC.

Zhang, G. P. (2012). *Neural Networks for Time-Series Forecasting*. In G. Rozenberg, T. Bäck, J. N. Kok (Eds.), *Handbook of Natural Computing*. Springer, Berlin, Heidelberg.