Aprendizado Profundo Aplicado na Previsão de Receita Tributária Utilizando Variáveis Endógenas

Deep Learning Applied in Forecasting Tax Revenues Using Endogenous Variables

Priscila F. Silva[1], Karla Figueiredo[1]

[1]Instituto de Matemática e Estatística – Universidade do Estado do Rio de Janeiro (UERJ) – Rio de Janeiro – RJ – Brazil

priscilafranca28@yahoo.com.br, karlafigueiredo@ime.uerj.br

Abstract. *The forecast of tax revenue for the management of a state in economic and financial crisis, such as the case of the state of Rio de Janeiro, has became a fundamental and challenging task for the State's Department of Finance and Planning, since the temporal series are affected by economic and political uncertainties. In this sense, this work starts the project to investigate and use new and more accurate Machine Learning models, as Long Short-Term Memory (LSTM), to predict the tax revenues. The results present a relative error less than 1% to predict ICMS, indicating a performance superior to the predictions made by SEFAZ-RJ and the MLP models, used for comparison purposes used for comparison purposes.*

Resumo. *A previsão de receita tributária para a gestão de um estado em situação de crise econômico-financeira, caso do estado do Rio de Janeiro, tornou-se uma tarefa fundamental e desafiadora para as secretarias de fazenda e planejamento dos Estados, pois há incertezas econômicas e políticas que afetam a série temporal. Nesse sentido, este trabalho inicia o projeto de investigação e uso de modelos mais novos e acurados de Machine Learning, como Long Short-Term Memory* (LSTM), *para prever a receita de tributos. Os resultados apresentam erro relativo menor que 1%, para previsão do ICMS, indicando desempenho superior às previsões realizadas pela SEFAZ-RJ e aos modelos MLP, usados para efeitos de comparação.*

1. Introdução

A Constituição Brasileira[1] impõe aos entes federativos a realização de um planejamento orçamentário, que deve ser apresentado publicamente em forma de leis orçamentárias. Por meio destas, as metas e objetivos do governo vigente são mostrados ao público, e também devem conter, de forma detalhada, os tributos esperados e a maneira como serões utilizados.

Além disso, por meio da Lei de Responsabilidade Fiscal (LRF)[2], o Estado tem o dever de realizar a previsão de receitas orçamentárias para o exercício governamental, a fim de usar como base para elaboração da estratégia de gestão financeira dos cofres públicos, de forma a manter o equilíbrio entre as receitas e despesas públicas.

[1]Constituição da República Federativa do Brasil. Brasília, DF: Senado Federal, 1988. http://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm

[2] Lei de Responsabilidade Fiscal: http://www.planalto.gov.br/ccivil_03/leis/lcp/lcp101.htm

Para a realização do projeto orçamentário é necessário usar como base a previsão das receitas públicas [LDO 2017], de forma a considerar o quanto de saldo positivo o estado terá disponível para efetivar as políticas públicas e sociais. Assim, entende-se que um bom planejamento é fundamental para que o estado consiga realizar seus projetos e cumprir todas suas obrigações com a população, sem ferir a LRF. Nestas são definidas normas que conduzem o processo de gasto das receitas pelo Governo do Estado.

Em virtude do grau de importância estabelecido na gestão dos recursos, é desejável que as estimativas de receitas sejam as melhores possíveis. O fato do estado do Rio de Janeiro se encontrar em Regime de Recuperação Fiscal [Resolução CSRRF nº 33 2020] torna essa atividade ainda mais sensível aos erros.

De acordo com a LDO (2017), o Imposto sobre Operações relativas à Circulação de Mercadorias e Prestação de Serviços de Transporte Interestadual e Intermunicipal e de Comunicação do Estado do Rio de Janeiro (ICMS) é a principal fonte de Receita do Estado do Rio de Janeiro, sendo o imposto com maior arrecadação no Estado.

Atualmente, a previsão de receitas do Estado do Rio de Janeiro é realizada pela Secretaria de Estado de Fazenda do Rio de Janeiro (SEFAZ-RJ) utilizando modelos estatísticos. De acordo com o documento "Lei de Diretrizes Orçamentárias"[3], tem sido aplicado nos últimos anos técnicas econométricas Autorregressivas Integradas de Médias Móveis Sazonais (do inglês *Seasonal Autoregressive Integrated Moving Average* - SARIMA) [Box and Jenkins 1976] e Autorregressão Vetorial (do inglês *Vector Autoregression* - VAR) [Sims 1980].

Assim, dada a grande importância da atual situação econômico-financeira no estado, o presente trabalho visa realizar a previsão mensal (12 meses) *multi-step* do ICMS do Estado do Rio de Janeiro, através do modelo de aprendizado profundo *Long Short-Term Memory* (LSTM) [Hochreiter and Schmidhuber 1997], visando melhorar os resultados em relação aos outros métodos, e principalmente, em relação ao métodos estatísticos adotados pelo Estado. Com isso acredita-se que o uso de modelos de receitas mais acurado contribui para um planejamento mais adequado e justo. Além disso, a série temporal pode ser afetada por maior volatilidade devido aos efeitos causados pela fragilidade econômica e político que estado atravessa.

De forma a analisar a eficácia do modelo LSTM, os seus resultados também serão comparados às previsões produzidas por modelos de redes neurais do tipo *Multi-Layer Perceptron* (MLP) [Haykin 1999], tradicionalmente usados para previsão de séries temporais [Zhang 2012].

Este trabalho está dividido em mais 4 seções. Na próxima seção serão discutidas as técnicas encontradas na literatura para o problema de previsão de series temporais não lineares, as quais apresentam comportamento semelhante à série temporal de arrecadação de tributos. A terceira seção discorre sobre a abordagem realizada para a resolução do problema, contendo as análises de pré-processamento realizadas e os algoritmos utilizados para a previsão de ICMS. Em seguida, na quarta seção¸ são apresentados os resultados obtidos com os modelos avaliados. E por fim, a última seção trata das conclusões e proposições para trabalhos futuros.

[3]Lei de Diretrizes Orçamentárias nº 7652
http://www.fazenda.rj.gov.br/sefaz/content/conn/UCMServer/path/Contribution%20Folders/site_fazenda/Subportais/PortalPlanejamentoOrcamento/2_ppa_ldo_loa/ldo/ldo2018.pdf?lve

2. Fundamentação Teórica

2.1. Rede Neural Artificial

A Rede Neural Artificial (RNA) é um modelo inspirado no funcionamento do cérebro humano, que procura reproduzir computacionalmente o seu comportamento, sendo usada principalmente para aprender e detectar padrões não lineares, produzindo resultados eficazes [Haykin 1999]. São características que a destaca: a sua capacidade de aprender, memorizar e generalizar.

2.2. Rede Neural *Multi-Layer Perceptron (MLP)*

A rede neural do tipo MLP tem sido bastante utilizada em diversas aplicações, tendo vasta literatura em previsão de séries temporais [Zhang 2012]. Sendo considerada como conhecimento básico para a área de *Machine Learning* [Haykin 1999], não será detalhada nesse trabalho.

2.3. Rede *Long Short-Term Memory*

Em [Goodfellow et al. 2016] a rede *Long Short-Term Memory* (LSTM) é apresentada como uma variação das redes neurais recorrentes tradicionais [Haykin 1999]. O modelo inicial foi criado por Sepp Hochreiter e Jürgen Schmidhuber (1997), sendo composto por camadas internas interligadas denominadas células, que são formadas por pequenas redes neurais, e por três portas, conhecidas como *forget gate*, *input gate* e *output gate*, responsáveis pela manipulação das informações a serem processadas.

Resumidamente pode-se mencionar que: o *forget gate* tem a tarefa de descartar as informações que não são úteis e o *input gate* tem a finalidade de determinar quais informações novas serão guardadas. Este é composto por duas camadas com diferentes funções de ativação que trabalham conjuntamente, uma sigmoide e uma tangente hiperbólica, e que, respectivamente, geram novos valores e determinam o quanto desses novos valores será transmitido para a saída. A combinação dessas informações também atualiza um novo atributo desse modelo: estado da célula. O *output gate* determina quais partes do estado da célula serão levados à saída, gerando um novo estado da célula [Schmidhuber 2015]. Assim, as redes LSTMs [Schmidhuber 2015], incorporadas à área de aprendizado profundo, permitem entradas multivariáveis, robustez ao ruído, saída multivariável, extração automática de recursos, modelagem das relações mais complexas nos dados e têm tido ótimos resultados na área de previsão de séries temporais [Schmidhuber 2015].

2.5. Trabalhos Relacionados

Especificamente as RNAs têm ganhado destaque devido a sua eficácia para previsão de séries temporais não lineares [Zhang 2012]. No cenário do mercado financeiro, Souza (2011) observou que para a previsão de índice Ibovespa, os modelos baseados em Redes Neurais mostraram resultados superiores ao de metodologias tradicionais de previsão de séries temporais.

Marangoni (2010) mostrou que as RNAs são capazes de prever de forma efetiva o preço de fechamento de ações de empresas para períodos curtos de tempo, mostrando resultados muito próximos da realidade. Conclui-se que as RNAs podem ser usadas

como método alternativo às analises convencionais não somente para dados absolutos previstos como também estimar novos valores a partir de dados previstos.

Em relação à previsão de arrecadação do tributo ICMS, Contreras and Cribari-Neto (2006) avaliaram a utilização de métodos de previsão baseados em MLPs para estimar a arrecadação do ICMS para os Estados de São Paulo, Rio de Janeiro e Pernambuco, revelando obter previsões mais precisas usando Redes Neurais ao invés de metodologias tradicionais. Destaca-se que o autor usou 10 anos de dados (1994 a 2005) e a previsão apresentada para o ICMS-RJ, para um horizonte de seis meses à frente, obteve um erro MAPE de 7,03%.

Bastos (2010) realizou previsão do ICMS para o estado do Ceará com modelos baseado em redes neurais do tipo MLP e ELMAN para o horizonte de 12 meses e obtendo erros MAPE respectivamente de 8% e 8.9% considerando 15 anos de dados. Em trabalho recente [Matos 2019] o modelo estatístico adotado, ARMA, apresentou MAPE de 4,8% para o imposto ICMS desse mesmo estado.

Silva and Figueiredo (2018) compararam as previsões de arrecadação de ICMS-RJ realizadas pela SEFAZ-RJ e pelo modelo *Multilayer Perceptron* (MLP). O modelo MLP obteve erro relativo menor que 1% em suas previsões, indicando ser uma opção para melhorar as previsões de arrecadação do ICMS-RJ.

Em [Namin and Namin 2018] a rede *Long Short-Term Memory* (LSTM) apresentou superioridade ao ser comparada com metodologias tradicionais (ARIMA) em aplicação para previsão de séries temporais financeiras. Nelson (2017) também analisou o uso da rede LSTM para previsão de series temporais financeiras, obtendo resultados melhores ao comparar redes baseadas em modelo MLP.

Schmidhuber (2015) realiza uma pesquisa histórica sobre as Redes Neurais, destacando trabalhos realizados ao longo do tempo, além de mostrar os diversos trabalhos realizados referentes às redes LSTM em diversos cenários de estudo.

Com isso, fica claro que a previsão de tributos utilizando modelos baseados em aprendizado profundo, ainda não devidamente explorado e carecendo de maiores contribuições, principalmente a partir dos novos modelos baseados em redes recorrentes, que têm tido bastante sucesso em tantas outras áreas de aplicação [Schmidhuber 2015], [Namin and Namin 2018], [Sahoo et al. 2019] e [Khodabakhsh et al. 2020].

3. Metodologia

Essencialmente a metodologia utilizada nesse trabalho envolve dois modelos de rede neurais: MLP e rede recorrente do tipo LSTM, devido à eficácia histórica de seus resultados em relação à previsão de séries temporais não lineares [Marangoni 2010] [Souza 2011] [Nelson 2017] [Namin and Namin 2018] [Silva and Figueiredo 2018] [Sahoo et al. 2019] [Khodabakhsh et al. 2020], e a exploração e extração de informações que pode ser obtida a partir dos dados de entrada.

3.1. Pré Processamento dos Dados

A série temporal usada neste trabalho foi extraída dos dados de arrecadação mensal do tributo ICMS do ERJ, presente no site da SEFAZ-RJ, sendo composta por valores

nominais de arrecadação em reais (R$) e compreende o período de janeiro de 2002 a dezembro de 2019.

A divisão da base de dados foi baseada no conceito de *holdout*, sendo criados três conjuntos distintos: treino, validação (*early stopping*) e teste. O conjunto de teste, validação e treinamento é respectivamente, de janeiro de 2019 a dezembro de 2019, de janeiro de 2017 a dezembro de 2018 e de janeiro de 2002 a dezembro de 2016.

Incialmente foram aplicados os métodos de normalização e diferenciação nos dados presentes no conjunto de entrada. Para avaliação dos resultados construiu-se dois modelos LSTM, considerando dados extraídos e processados de formas diferentes, mas ambos os casos utilizando dados endógenos, ou seja, apenas dados históricos de arrecadação de ICMS, pois nesse primeiro estágio ainda não foi considerada a inclusão de variáveis exógenas tais como consumo de energia elétrica, combustível e telefonia, que representam parcelas significantes na receita do Estado.

O primeiro conjunto de dados de entrada é composto por valores originais da série normalizados e organizados numa janela de 12 meses. O segundo conjunto também se baseia em uma janela temporal de 12 meses sobre os dados originais normalizados, porém são criadas novas variáveis a partir de manipulações matemáticas aplicadas esses dados, conforme realizado no trabalho [Ticona 2013].

As manipulações matemáticas consistem em médias móveis e *lag's*, que são retardos nos valores originais da série. Sendo assim, nesse conjunto são criadas 11 variáveis, conforme pode ser observado na Tabela 1. As quatro primeiras variáveis são as médias móveis, dos últimos 12, seis, três e dois meses, respectivamente. As variáveis numeradas de 5 a 10 são estabelecidas a partir de *lag's* em relação ao mês de saída, e a última é a saída (rótulo) com o valor da receita para o mês a ser previsto.

Tabela 1. Cálculos realizados sobre os valores da série.

Número da Variável	Variável de Entrada	Cálculos Realizados Sobre os Valores da Série (V)
1	MM12	$\frac{V_{n-1} + V_{n-2} + \cdots + V_{n-11} + V_{n-12}}{12}$
2	MM6	$\frac{V_{n-1} + V_{n-2} + \cdots + V_{n-5} + V_{n-6}}{6}$
3	MM3	$\frac{V_{n-1} + V_{n-2} + V_{n-3}}{3}$
4	MM2	$\frac{V_{n-1} + V_{n-2}}{2}$
5	L12	V_{n-12}
6	L6	V_{n-6}
7	L4	V_{n-4}
8	L3	V_{n-3}
9	L2	V_{n-2}
10	L1	V_{n-1}
11	rótulo	$V_{n=13}$

Destaca-se que na construção dos modelos foi empregado o método *multi-step* nas etapas de validação e teste, ou seja, para a previsão dos valores sucessivos para os um horizonte de 12 meses de previsão. Assim, para se realizar a primeira previsão do

conjunto de validação ou teste são considerados os 12 últimos valores da série normalizada (com ou sem manipulação matemática). Já para a previsão do segundo mês, a janela de dados se desloca um mês à frente, perdendo o primeiro elemento, e incorpora o valor recém-previsto. Esse processo segue ao longo do restante dos dados de validação e teste para prever os meses seguintes. Dessa forma, para prever os meses a partir do décimo segundo, todos os dados de entrada que alimentam as redes são valores previstos nos passos anteriores. Isso cria uma dificuldade maior para o modelo, mas o aproxima do uso real.

3.2. Estrutura da Rede LSTM / MLP

A estrutura interna da rede é formada por uma camada oculta e por uma camada *dense*. A camada de entrada está preparada para receber doze variáveis de entrada e gerar um único registro de saída, que representa o valor estimado para o mês seguinte em relação ao conjunto de entrada.

3.3. Análise e Avaliação

Inicialmente a rede passa por ciclos de treinamento e validação, que consistem em atualizar os pesos da rede, a partir do erro observado com as previsões dos dados de treinamento, e em seguida avaliar o desempenho do modelo treinado por meio de erros tais como erro médio quadrático (do inglês *Mean Squared Error* - MSE) (equação 1), raiz quadrada do erro médio quadrático (do inglês, *Root Mean Square Error* - RMSE) e o erro percentual médio absoluto (do inglês, *Mean Absolute Percentage Error* - MAPE) (equação 2), com os dados do penúltimo e antepenúltimo anos (2017 e 2018) da série.

$$MSE = \frac{\sum_{i=1}^{n}(real - previsto)^2}{n} \qquad (1)$$

$$MAPE = \frac{\sum_{i=1}^{n}(|real - previsto|)}{real}\ x\ 100 \qquad (2)$$

onde i: é o número de registros da base de dados.

Para cada topologia avaliada, as etapas de treinamento e validação foram realizadas consecutivamente e múltiplas vezes executadas, sendo interrompido de acordo o método de parada antecipada.

Dessa forma, para escolher a melhor arquitetura de ambos os modelos de redes neurais utilizados neste trabalho (MLP e LSTM) foi calculado o erro médio de validação utilizando o erro médio quadrático (RMSE), com diferentes configurações dos parâmetros: número de neurônios, número de camadas ocultas e função de ativação, a partir de 10 instâncias para cada combinação de parâmetros. O mesmo procedimento foi realizado para os experimentos que identificaram a arquitetura ideal para o modelo MLP.

4. Experimentos

Nesta seção são apresentados os experimentos realizados e os resultados obtidos, a partir das investigações de parâmetros envolvidos em cada modelo, visando a definição, entre os valores buscados, a topologia para prever a arrecadação do tributo ICMS. Também são analisados os tipos de rede neural que, entre os modelos LSTM e MLP,

apresentam melhor acurácia, além da escolha do melhor conjunto de dados de entrada a ser utilizado.

Para elaboração dos estudos de caso do modelo LSTM, consideraram-se as duas formas descritas para compor o conjunto de entrada, além da variação dos seguintes parâmetros da rede: função de ativação, número de neurônios na camada intermediária, e número de camadas. Os experimentos serão realizados de acordo com os cenários descritos na Tabela 2, ressaltando-se que foram feitos 10 experimentos para cada arquitetura avaliada. Para os casos de estudo do modelo MLP, foram alterados apenas o número de neurônios na primeira camada escondida.

Tabela 2. Parametrização para as redes LSTM.

Modelo	Número de Camadas	Função de Ativação	Cenários de Estudo
LSTM com variáveis de entrada originais normalizadas e método Adam	1	relu	1
		sigmoide	2
	2	relu	3
		sigmoide	4
LSTM com variáveis de entrada manipuladas matematicamente e método Adam	1	relu	5
		sigmoide	6
	2	relu	7
		sigmoide	8
MLP com variáveis de entrada manipuladas matematicamente e método SGD	1	tangente hiperbólica	9
	2	tangente hiperbólica	10

Em relação aos parâmetros do modelo LSTM, foi utilizado *mini-batch* igual a 32 e adotado o algoritmo ADAM [Kingma and Ba 2015] para realizar o ajuste dos pesos, considerando a taxa de aprendizado de 0,001. A definição de arquitetura do modelo MLP é determinada a partir do melhor modelo identificado no trabalho de [Silva and Figueiredo 2018].

Para os experimentos de ambos os modelos foram avaliados, considerando os trabalhos relacionados, os seguintes números de neurônios na 1ª camada escondida: 1, 2, 3, 4, 5, 10, 20, 50, 80 e 100, visando identificar a melhor arquitetura. Para os casos de estudo com duas camadas intermediárias, somente é alterado o número de neurônios da 1ª camada (na 2ª camada é usado apenas 1 neurônio). Para interromper o treinamento foi usado com critério o método *early stopping* a partir dos dados de validação. Os dados da arrecadação do ICMS-RJ utilizados em todos os experimentos compreendeu o período entre 2002 a 2016 para o treinamento e os anos do 2017 e 2018 para validação.

Na Tabela 3 são apresentados os erros médios de validação dos melhores modelos LSTM obtidos, tendo o cenário 4 o melhor resultado para o erro RMSE. Portanto, a topologia que obteve a melhor acurácia é a que apresenta a configuração: 2 camadas intermediárias, com 50 neurônios na 1ª camada e 1 neurônio na 2ª camada; função de ativação sigmoide; valor 32 para o tamanho do min-batch; e algoritmo Adam para ajuste dos pesos, e como dados de entrada, o conjunto de os valores reais da série que sofreram apenas o processo de diferenciação e normalização.

Na Tabela 4 são apresentados os erros médios de validação dos melhores modelos MLP avaliados. Destaca-se o cenário 10 como o menor erro RMSE. Portanto, a topologia que obteve a melhor acurácia é a que apresenta a configuração: 2 camadas intermediárias, com 1 neurônio na 1ª camada e 1 neurônio na 2ª camada; função de

ativação tangente hiperbólica; *mini-batch* com 32 registros; algoritmo SGD para ajuste dos pesos, e tendo como dados de entrada o conjunto com variáveis originais manipuladas matematicamente.

Tabela 3. Erro médio MSE, RMSE e MAPE de validação para os melhores modelos LSTM descritos nos cenários (Tabela 2).

Erro Médio de Validação				
Cenário (em relação à Tabela 2)	**Número de Neurônios**	**MSE (milhão de R$)**	**RMSE (milhão de R$)**	**MAPE (%)**
1	1	40144533.41	6.32	0,22
2	1	88011852.95	9.38	0,31
3	5	22361176.95	4.73	0,17
4	**50**	**20892625.99**	**4.57**	**0,16**
5	1	89764603.91	9.07	0,26
6	1	112298932.33	10.60	0,33
7	4	24110273.17	4.91	0,1
8	1	21042166.32	4.59	0,16

Tabela 4. Erro médio MSE, RMSE e MAPE de validação para os melhores modelos MLP descritos nos cenários (Tabela 2).

Erro Médio de Validação				
Cenário (em relação à Tabela 2)	**Número de Neurônios**	**MSE (milhão de R$)**	**RMSE (milhão de R$)**	**MAPE (%)**
9	2	5323486562.93	72.71	1,98
10	**1**	2833353125.65	53.22	**1,52**

Como a previsão feita pelo estado é anual, o que se fez foi somar os valores previstos pelos modelos MLP e LSTM e comparar os resultados. Assim, a Tabela 5 apresenta o erro relativo (equação 3) obtido nas previsões feitas pelos melhores modelos LSTM e MLP identificados e pelos modelos SARIMA e VAR propostos pela SEFAZ-RJ, em relação ao valor real do ICMS-RJ para os anos de 2017 e 2018 (conjunto de validação).

$$Erro\ Relativo\ = \frac{|Previsão - Real|}{Real} \qquad (3)$$

Analisando os resultados na Tabela 5, nota-se que os modelos SARIMA e VAR, utilizados pela SEFAZ-RJ, apresentaram erro relativo superiores aos modelos MLP e LSTM.

Tabela 5. Previsão do ICMS-RJ realizado pelas Redes LSTM e MLP e pelos modelos estatísticos SARIMA e VAR

	ICMS-RJ Ano 2017		**ICMS-RJ Ano 2018**	
	Valor (milhão de R$)	**Erro Relativo (%)**	**Valor (milhão de R$)**	**Erro Relativo (%)**
Valor Real	32.362,50	-	35.836,06	-
SARIMA e VAR (SEFAZ-RJ)	35.287,45	9,04	33.746,76	5,83
Rede LSTM	32.408,22	0,14	35.881,78	0,13
Rede MLP	32.660,58	0,92	36.025, 09	0,53

4.1. Comparação dos Modelos LSTM e MLP

A Tabela 6 apresenta o erro médio RMSE dos dois melhores modelos LSTM e MLP identificados com o conjunto de validação. Analisando os resultados, o modelo LSTM foi o que apresentou o menor erro RMSE, sendo o modelo de rede neural mais indicado a ser empregado na previsão da série do ICMS-RJ.

Tabela 6. Erro médio MSE, RMSE e MAPE de validação para os melhores modelos de LSTM e MLP identificados.

Erro Médio de Validação (anos 2017 e 2018)					
Modelo	**Número de Neurônios por camada**	**Número de Camadas**	**MSE (milhão de R$)**	**RMSE (milhão de R$)**	**MAPE (%)**
LSTM	**50/1**	**2**	**20892625,99**	**4,57**	**0,15%**
MLP	1/1	2	**2833353125,65**	**53,22**	1,52%

A Tabela 7 indica os valores de erro anual relativo para o período de teste (ano 2019) e a Tabela 8 apresenta o erro relativo entre as previsões mensais feitas para o ano de 2019 com os melhores modelos LSTM e MLP, e o valor real do ICMS-RJ. Observa-se que as previsões de arrecadação do ICMS-RJ obtidas por ambos os modelos de redes neurais apresentam valores bem próximos do real. Destaca-se que o LSTM e MLP têm o erro MAPE iguais a 0,12% e 3,06%, respectivamente e a SEFAZ-RJ não disponibiliza os valores de previsões mensais.

Tabela 7. Erro relativo para previsão anual da arrecadação do ICMS-RJ do ano de 2019 (teste) para os melhores modelos LSTM e MLP e SEFAZ-RJ.

	ICMS-RJ Ano 2019	
	Valor (milhão de R$)	**Erro Relativo (%)**
Valor Real	36.362,00	-
SARIMA e VAR (SEFAZ-RJ)	35.821,25	1,49
Rede LSTM	36.407,73	0,13
Rede MLP	36.671,96	0,85

Tabela 8. Erro relativo das previsões mensais de arrecadação do ICMS-RJ do ano de 2019 (teste) para os melhores modelos LSTM e MLP.

ICMS-RJ - Ano 2019					
Mês	**Valor Real ICMS-RJ (milhão de R$)**	**Valor Previsto pela LSTM (milhão de R$)**	**Erro Relativo LSTM (%)**	**Valor Previsto pela MLP (milhão de R$)**	**Erro Relativo MLP (%)**
Janeiro	3534,88	3538,69	0,11	3671,14	3,85
Fevereiro	3143,84	3147,65	0,12	3188,70	1,43
Março	2718,26	2722,07	0,14	2574,49	5,29
Abril	3143,63	3147,44	0,12	3319,62	5,60
Maio	2938,16	2941,97	0,13	2931,78	0,22
Junho	2666,57	2670,38	0,14	2622,62	1,65
Julho	2703,05	2706,86	0,14	2895,35	7,11
Agosto	2879,30	2883,11	0,13	2824,81	1,89
Setembro	3063,19	3067,00	0,12	3023,71	1,29
Outubro	3070,53	3074,34	0,12	3071,05	0,02
Novembro	3066,48	3070,29	0,12	3224,78	5,16
Dezembro	3434,11	3437,92	0,11	3323,91	3,21

A Figura 1 apresenta a comparação entre a série temporal real do ICMS-RJ para o ano de 2019 (conjunto de teste) e os valores previstos mensalmente pelos melhores modelos de LSTM e MLP indicados na Tabela 8. Observa-se pequena divergência entre os valores previstos pelas modelos de redes neurais e os valores reais da série temporal do ICMS. Ao se aplicar os testes de *t-Student* e *Mann-Whitney Wilcoxon* [Fay and Proschan 2010] aos erros relativos apresentados na Tabela 8, obteve-se p-value=0.00024 e p=0.00028, respetivamente, indicando que são distribuições distintas e portanto pode-se considerar que o erro obtido pela rede LSTM é significativamente inferior ao MLP.

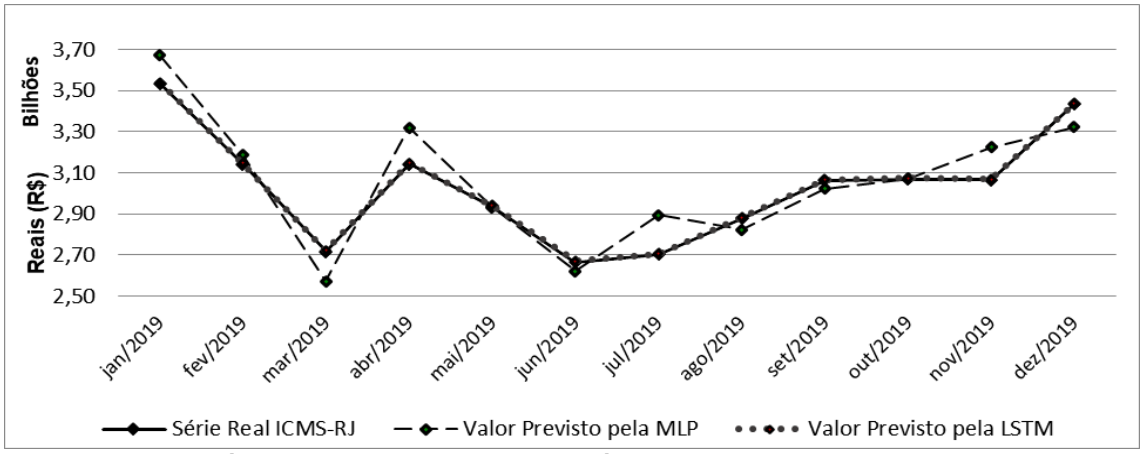

Figura 1. Gráfico de comparações entre série ICMS-RJ e valores previstos pela MLP e LSTM.

5. Conclusão e Trabalhos Futuros

O objetivo principal deste trabalho foi investigar modelos de *Machine Learning* e variáveis de entrada, para melhorar a previsão do mais importante imposto Estadual, que, quando tem previsão mais acurada, aumenta a qualidade do planejamento do orçamentário anual. Destaca-se também que aplicação de *Machine Learning* para previsão de tributos também estava estacionada em modelos tradicionais, não tendo sido identificada na literatura nenhuma iniciativa de uso de redes LSTM para essa finalidade. Assim, foram investigados e analisados modelos de redes neurais LSTM, baseado em aprendizado profundo, e modelos baseado em redes neurais do tipo MLP, para realizar a previsão mensal da série temporal do ICMS do Estado do Rio de Janeiro em um horizonte de 12 meses à frente. A partir dos resultados obtidos e apresentados, entende-se que o uso de modelo LSTM se mostra como uma opção eficiente para melhorar a previsão do ICMS e aumentar a qualidade das informações a serem apresentadas e utilizadas pela administração pública, considerando um histórico de treinamento de 14 anos.

A fim de comparar a eficácia da LSTM também foi desenvolvido outro modelo de rede neural (MLP), já explorado no trabalho realizado por Silva e Figueiredo (2018). As duas redes apresentaram valores previstos bem próximos dos valores reais, contudo o modelo LSTM foi considerado superior ao modelo MLP por pequena margem de diferença entre o erro relativo obtido por ambas, considerando o conjunto de validação e o conjunto de teste como dados de estudo. Também foram avaliados dois tipos de variáveis de entrada, conclui-se que a escolha das variáveis depende da forma de

processamento do modelo, sendo para LSTM o uso de valores originais da série normalizados e variáveis criadas a partir de manipulações matemáticas sobre os dados originais da série para as redes MLP. O que parece bastante coerente, comprovando que as redes LSTM conseguem explorar e extrair relações temporais das séries de forma mais eficiente do que as redes MLP, tendo essa rede melhor desempenho com dados previamente manipulados. Finalmente, o melhor modelo LSTM reduz o erro relativo em aproximadamente 91%, em relação à previsão anual feita pelo Estado para o mesmo período (2019).

Os próximos passos devem considerar horizontes de previsões maiores, em relação ao período de apuração do imposto e investigação de variáveis exógenas ao tributo, mas que guardem correção (linear ou não linear) que possam agregar informação. Devido à brusca redução da arrecadação já sendo observada para 2020, devida à pandemia de COVID-19, deve ser incluir técnicas de *concept drift* [Gama et al. 2014] aos modelos.

Referências

Bastos, F.A.A. (2010) "Previsão de Receitas Tributárias Mediante Redes Neurais Artificiais". 78 f. Dissertação (Mestrado) – Universidade. Estadual do Ceará, http://www.uece.br/mpcomp/index.php/arquivos/doc_download/263-dissertacao82, Março.

Box, G.E.P., Jenkins, g.ma.1976. Time series analysis Forecasting and control, 2nd ed. Holden-Day.San.

Contreras, J.C.S., Cribari-Neto, F. (2006) "Previsão da Arrecadação do ICMS no Brasil através de redes neurais". In: Revista Brasileira de Estatística, Rio de Janeiro, v. 67, n. 227, p. 7-43.

Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M. and Bouchachia, A., (2014) A survey on concept drift adaptation. ACM computing surveys (CSUR), 46(4), p.44.

Goodfellow, I., Bengio, Y. and Courville, A. (2016) Deep Learning. [s.l.] MIT Press.

Fay, M., Proschan, M. (2010). "Wilcoxon–Mann–Whitney or t-test? On assumptions for hypothesis tests and multiple interpretations of decision rules". Statistics Surveys. 4: 1–39. doi:10.1214/09-SS051

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

Hochreiter, S., Schmidhuber, J. (1997) "Long Short-Term Memory. Neural Computation".

Khodabakhsh A., Ari I., Bakır M., Alagoz S.M. (2020) Forecasting Multivariate Time-Series Data Using LSTM and Mini-Batches. In: Bohlouli M., Sadeghi Bigham B., Narimani Z., Vasighi M., Ansari E. (eds) Data Science: From Research to Application. CiDaS 2019. Lecture Notes on Data Engineering and Communications Technologies, vol 45. Springer, Cham.

Kingma, P.D., Ba, J. (2015) "Adam: A Method for Stochastic Optimization", In: International Conference for Learning Representations, 3., 2015, San Diego, https://arxiv.org/pdf/1412.6980.pdf, Julho.

Marangoni, P.H. (2010) “Redes Neurais Artificiais Para Previsão de Séries Temporais no Mercado Acionário”. Dissertação (Graduação em Ciências Econômicas) - Universidade Federal de Santa Catarina, Florianópolis.

Matos M.A.C.S. (2019) Avaliando o Forecast Content e Forecast Horizon de Modelos ARMA para os principais Agregados de Arrecadação de ICMS do Estado do Ceará, Dissertação - Curso de Mestrado em Economia, da Universidade Federal do Ceará.

Namin, S.S., Namin, A.S. (2018) “Forecasting economic and financial time series: ARIMA vs LSTM”, https://arxiv.org/ftp/arxiv/papers/1803/1803.06386.pdf, Julho.

Nelson, D.M.Q. (2017) “Uso de Redes Neurais Recorrentes Para Previsão de Séries Temporais”. Dissertação - Instituto de Ciências Exatas, Universidade Federal de Minas Gerais, Belo Horizonte.

Rio de Janeiro. (2017) Lei de Diretrizes Orçamentárias nº 7652, de 19 de julho de 2017**.** Dispõe sobre as diretrizes para elaboração da Lei de Orçamento Anual de 2018 e dá outras providências. http://www.fazenda.rj.gov.br/sefaz/content/conn/UCMServer/path/Contribution%20Folders/site_fazenda/Subportais/PortalPlanejamentoOrcamento/2_ppa_ldo_loa/ldo/ldo2018.pdf?lve, Março.

Rio de Janeiro. (2020) Resolução nº 33 do Conselho de Supervisão do Regime de Recuperação Fiscal do Estado do Rio de Janeiro, de 05 de agosto julho de 2020**.** http://www.fazenda.rj.gov.br/sefaz/content/conn/UCMServer/path/Contribution%20Folders/transparencia/RecuperacaoFiscal/docs/item%207/Resolucoes/SEI_ME%20-%209673492%20-%20Resoluc%cc%a7a%cc%83o%2033-2020.pdf?lve, Agosto.

Sahoo, B. B., Jha, R., Singh, A., Kumar, D. (2019) Long short-term memory (LSTM) recurrent neural network for low-flow hydrological time series forecasting. Acta Geophys. 67, 1471–1481.

Schmidhuber, J. (2015) “Deep Learning in neural networks: An overview Neural Networks”, 2015.

Silva, H., Figueiredo, T. (2018) “Previsão da Arrecadação do ICMS por Redes Neurais Utilizando Variáveis Endógenas”. Trabalho de Conclusão de Curso (Graduação em Ciência da Computação) – Instituto de Matemática e Estatística, UERJ.

Sims, C.A. (1980) Macroeconomics and reality. Econometrica, 48, 1-48

Souza, R. L. R. (2011) “Previsão do Índice Bovespa por Meio de Redes Neurais Artificiais: uma Análise Comparada aos Métodos Tradicionais de Séries de Tempo”. Dissertação – Universidade Federal do Rio Grande do Norte, Natal, https://repositorio.ufrn.br/jspui/handle/123456789/12197, Março.

Ticona, W. M. (2013) “Estudo de Métodos de Mineração de Dados Aplicados à Gestão Fazendária de Municípios”. Dissertação (Mestrado), Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, http://www.dbd.puc-rio.br/pergamum/tesesabertas/1012165_2013_completo.pdf, Julho.

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