

**PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS**  
**NÚCLEO DE EDUCAÇÃO A DISTÂNCIA**  
**Pós-graduação *Lato Sensu* em Ciência de Dados e *Big Data***

**Lucas Rodrigues Amaral**

**TÉCNICAS DE *MACHINE LEARNING* APLICADAS NA PREVISÃO DA  
ARRECADAÇÃO FEDERAL**

Belo Horizonte

2021

**Lucas Rodrigues Amaral**

**TÉCNICAS DE *MACHINE LEARNING* APLICADAS NA PREVISÃO DA  
ARRECADAÇÃO FEDERAL**

Trabalho de Conclusão de Curso apresentado  
ao Curso de Especialização em Ciência de  
Dados e Big Data como requisito parcial à  
obtenção do título de especialista.

Belo Horizonte

2021

## SUMÁRIO

<b>1. Introdução.....</b>	<b>4</b>
<b>1.1. Contextualização .....</b>	<b>4</b>
<b>1.2. O problema proposto .....</b>	<b>4</b>
<b>2. Coleta de Dados .....</b>	<b>5</b>
<b>3. Processamento/Tratamento de Dados .....</b>	<b>6</b>
<b>4. Análise e Exploração dos Dados .....</b>	<b>6</b>
<b>5. Criação de Modelos de Machine Learning .....</b>	<b>10</b>
<b>6. Apresentação dos Resultados .....</b>	<b>14</b>
<b>7. Links .....</b>	<b>17</b>
<b>ANEXOS .....</b>	<b>17</b>

## 1. Introdução

### 1.1. Contextualização

O mundo foi surpreendido pela pandemia do vírus Covid-19 em 2020 e os impactos econômicos causados precisam ser amenizados com políticas públicas capazes de reduzir as desigualdades e o sofrimento das pessoas mais vulneráveis.

Cabe à Administração Tributária prover o Estado de recursos para implementação das políticas públicas. E diante desse cenário de incertezas, torna-se ainda mais fundamental o desenvolvimento de melhores planejamentos.

*Machine learning* representa o campo de estudo referente ao desenvolvimento de algoritmos computacionais que transformam dados em informação e inteligência<sup>1</sup>. Esse trabalho tem como objetivo tentar enriquecer<sup>2</sup> o planejamento da gestão pública aplicando técnicas de modelagem computacional de *machine learning* na previsão de arrecadação de tributos federais administrados pela Secretaria Especial da Receita Federal (RFB).

### 1.2. O problema proposto

Em tempos de pandemia Covid-19, o planejamento da gestão dos recursos públicos torna-se ainda mais desafiador. E a implementação de políticas públicas depende da arrecadação de tributos, que é influenciada diretamente pelos impactos da pandemia no ambiente econômico.

Uma boa gestão da previsão da arrecadação tributária poderá contribuir para o aprimoramento da gestão dos recursos públicos nesse cenário de incertezas. E esse trabalho tem como propósito apresentar uma técnica complementar aos métodos tradicionais de previsão da arrecadação com técnicas de análises de dados

---

<sup>1</sup> PAWLUS e RODGER em *Hands-On Deep Learning With R*. April 2020. alertam sobre as limitações dessas modelagens computacionais. Embora tenha uma vasta aplicação potencial, esses modelos ainda representam um subconjunto limitado das capacidades do cérebro humano. E oferece pouca flexibilidade de extrapolação além dos parâmetros definidos. Posto isso, recomenda-se muita cautela no entendimento do que o algoritmo aprendeu antes de tentar a sua aplicação no mundo real.

<sup>2</sup> Importante ressaltar que o trabalho não pretende substituir os métodos tradicionais, mas sim contribuir com o aprimoramento desses métodos. Não se tem também a pretensão de esgotamento dos estudos sobre a aplicação de técnicas de *machine learning*, o objetivo é apresentar alternativas de metodologias que podem tentar contribuir com a aprimoramento da previsão da arrecadação pública.

utilizando algoritmos de *machine learning*. Mais especificamente, serão utilizados algoritmos computacionais de redes neurais.

Foram utilizados dados de séries temporais públicas da Secretaria Especial da Receita Federal do Brasil (RFB)<sup>3</sup> publicadas pelo Banco Central do Brasil (BACEN)<sup>4</sup>, pelo Instituto de Pesquisa Econômica Aplicada (IPEA)<sup>5</sup>, pelo Instituto Brasileiro de Geografia e Estatística (IBGE) e pela Eletrobrás. O período trabalhado compreende os dados mensais de janeiro de 1995 a dezembro de 2020.

## 2. Coleta de Dados

Os dados foram coletados por *Application Programming Interface* (APIs) disponíveis em bibliotecas da linguagem de programação R<sup>6</sup> no ambiente de desenvolvimento *open source* do *R Studio*<sup>7</sup>.

Nome da coluna	Tipo	Descrição
Data	Data	Data de referência (mensal).
Arrecadação federal	Numérico	Valor da arrecadação administrada pela RFB.
PIB	Numérico	Valor PIB em R\$ (milhões).
Consumo de energia elétrica – Total	Numérico	Valor do consumo de energia elétrica em GWh.
Consumo de energia elétrica – Comercial	Numérico	Valor do consumo de energia elétrica em GWh.
Consumo de energia elétrica – Residencial	Numérico	Valor do consumo de energia elétrica em GWh.
Consumo de energia elétrica – Industrial	Numérico	Valor do consumo de energia elétrica em GWh.
Consumo de energia elétrica – Outros	Numérico	Valor do consumo de energia elétrica em GWh.
IPCA	Numérico	Índice de preços ao consumidor amplo
Base monetária	Numérico	Base Monetária Ampliada (saldo em final de período)
Dívida Pública	Numérico	Dívida Líquida do Setor Público - Saldos em R\$ milhões - Total - Governo Federal
SELIC	Numérico	Taxa de juros - Selic acumulada no mês
SELIC Anualizada	Numérico	Selic acumulada no mês anualizada base 252
R\$/US\$	Numérico	Taxa de câmbio R\$ por US\$ (Sisbacen PTAX800)

Tabela 1. Dados coletados por API

<sup>3</sup> Relatórios do Resultado da Arrecadação

<https://receita.economia.gov.br/dados/receitadata/arrecadacao/relatorios-do-resultado-da-arrecadacao/arrecadacao-2020/agosto2020/notago20.xlsx>

<sup>4</sup> SGS – Sistema Gerenciador de Séries Temporais do Banco Central do Brasil

<https://www3.bcb.gov.br/sgspub/localizarseries/localizarSeries.do?method=prepararTelaLocalizarSeries>

<sup>5</sup> Instituto de Pesquisa Econômica Aplicada

<https://www.ipea.gov.br/portal/>

<sup>6</sup> As principais bibliotecas utilizadas foram rbc ( <https://github.com/wilsonfreitas/rbcb> ), ipeadatar ( <https://cran.r-project.org/web/packages/ipeadatar/index.html> ).

<sup>7</sup> *R Studio* <https://rstudio.com/>

### 3. Processamento/Tratamento de Dados

Os dados foram obtidos pelas bibliotecas contidas nos pacotes *ipeadata* e *rbcb*. Cada um desses pacotes retornam os dados de séries temporais da arrecadação federal, do PIB, do consumo de energia elétrica, da SELIC, do câmbio, dentre outras.

Foram obtidos 5.024 registros para todas as séries temporais (arrecadação, PIB, consumo de energia total, comercial, residencial, industrial e outros, empregados PNADC, IPCA, empregos formais, IBOVESPA, câmbio US\$/R\$, base monetária ampliada, dívida líquida do setor público, SELIC, ... ) compreendendo o período de Jan/1995 a Fev/2021.

Nenhum tratamento especial foi necessário ser aplicado aos dados, salvo a junção das séries temporais pela data e a padronização dos valores em intervalos de 0 a 100. Modelos de *machine learning* funcionam melhor quando os dados possuem uma distribuição normal. E para atender a esse requisito, os dados podem ser normalizados ou padronizados<sup>8</sup>. A padronização dos valores na escala de 0 a 100 foram importantes também para se conseguir comparar as séries de dados distintas no mesmo gráfico. Essa padronização se deu pelo cômputo do valor da variável menos o mínimo, dividindo-se então pelo máximo menos o mínimo vezes 100<sup>9</sup>.

### 4. Análise e Exploração dos Dados

A partir dos dados coletados, foram então construídos os gráficos das séries temporais da arrecadação e de variáveis macroeconômicas.

---

<sup>8</sup> LANTZ, Brett. Machine Learning With R. Third Edition April 2019. Packt.

<sup>9</sup> VALOR PADRONIZADO = [ VALOR – MINIMO(VALOR) ] / [ MAXIMO(VALOR) – MINIMO(VALOR) ] x 100

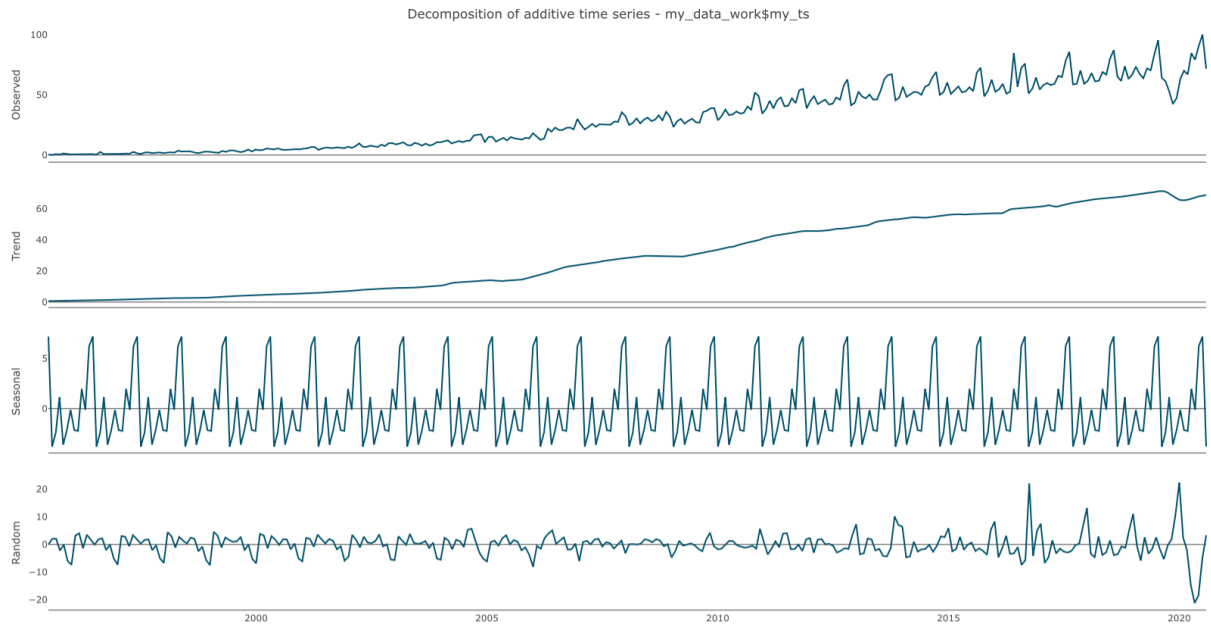


Gráfico 1. Série temporal da arrecadação decomposta.

O gráfico [1] apresenta as curvas da arrecadação, de tendência, sazonalidade e o ruído da série de dados<sup>10</sup>. E o gráfico [2] apresenta as curvas da arrecadação, do consumo de energia elétrica e do PIB.

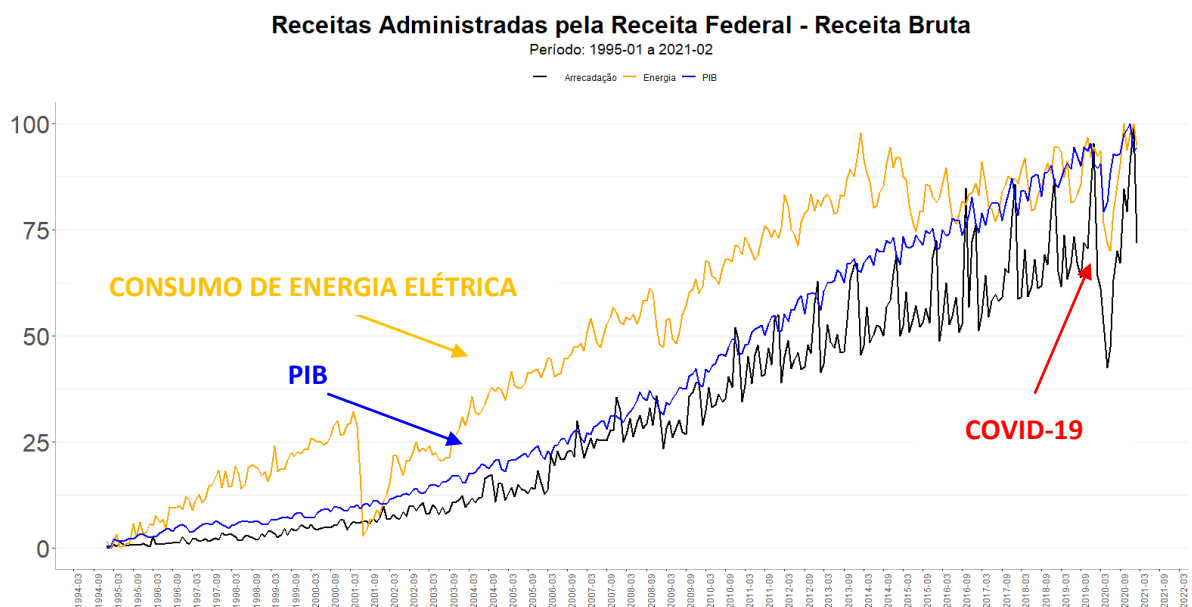


Gráfico 2. Arrecadação Federal vs. Consumo de Energia Elétrica (em amarelo) vs. PIB (em azul)

A partir da análise dos gráficos, percebeu-se nitidamente o impacto da COVID-19 nas três séries temporais. Observou-se também alta correlação entre as

<sup>10</sup> KRISPIN, Rami. *Hands-On Time Series Analysis with R*. Page 255.

três séries temporais<sup>11</sup>. Para confirmação da correlação, foram realizadas análises e testes de correlação de *Pearson*.

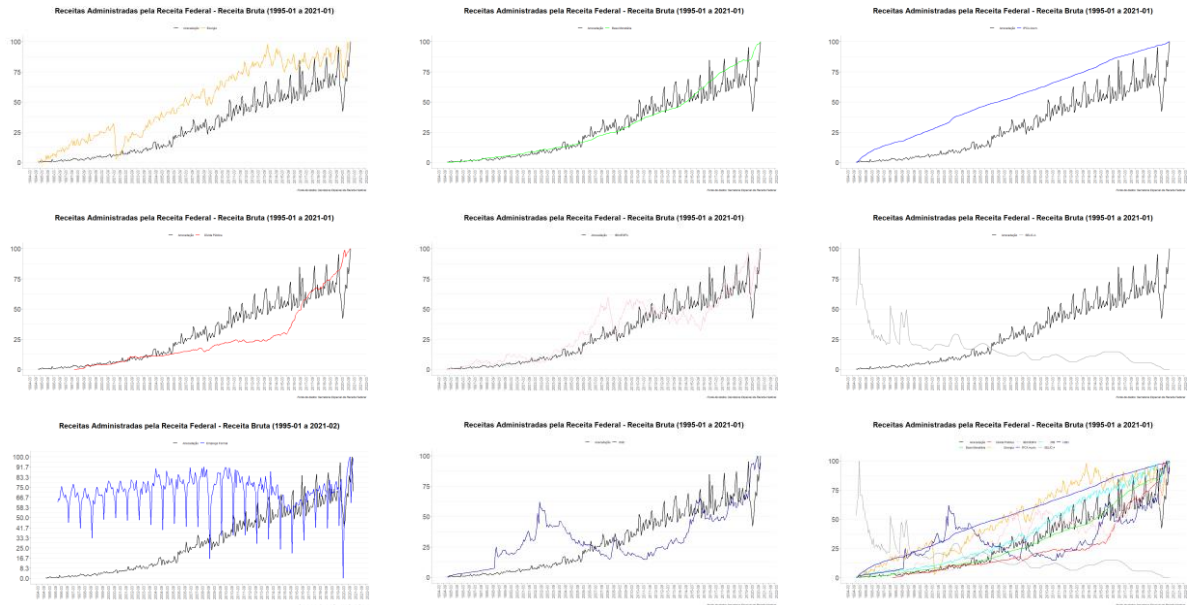


Figura 1. Arrecadação vs. Consumo de Energia, Base Monetária, IPCA, Dívida Pública, IBOVESPA, Empregos Formais, SELIC, PIB e Dólar.

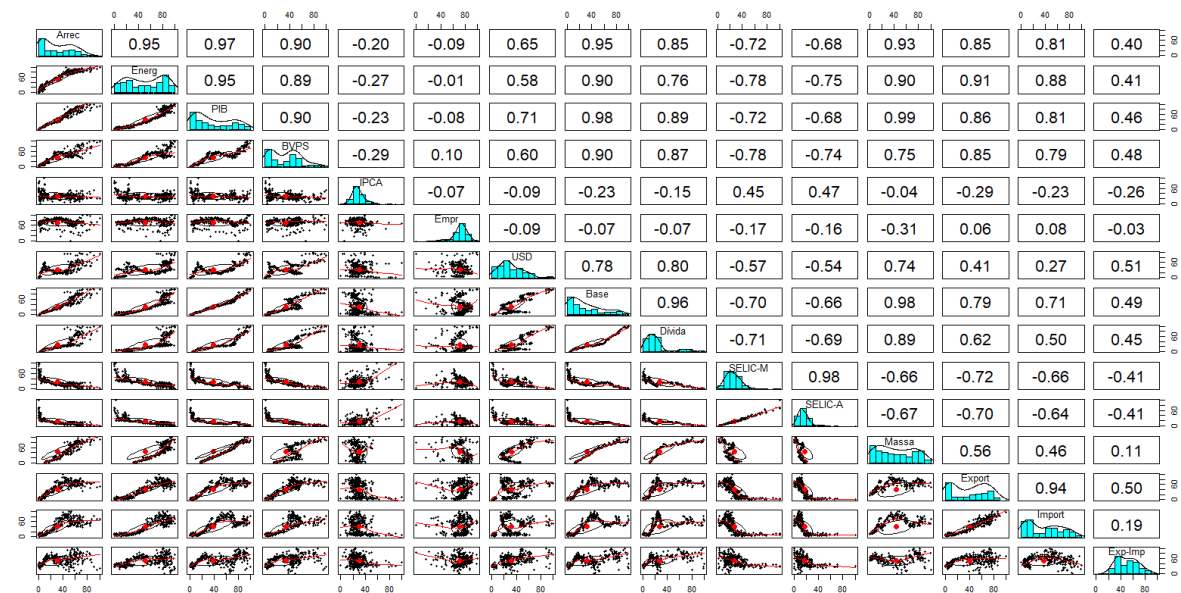


Figura 2. Ilustração gráfica da matriz de correlação de *Pearson* entre variáveis macroeconômicas.

<sup>11</sup> Importante aqui destacar que a modelagem utilizou apenas os dados da série temporal da arrecadação federal. As demais métricas referentes a PIB, energia elétrica, dólar, dívida pública, IPCA, SELIC e emprego serão utilizadas em versões posteriores desse trabalho aplicando-se técnicas multivariadas de redes neurais na previsão da arrecadação.



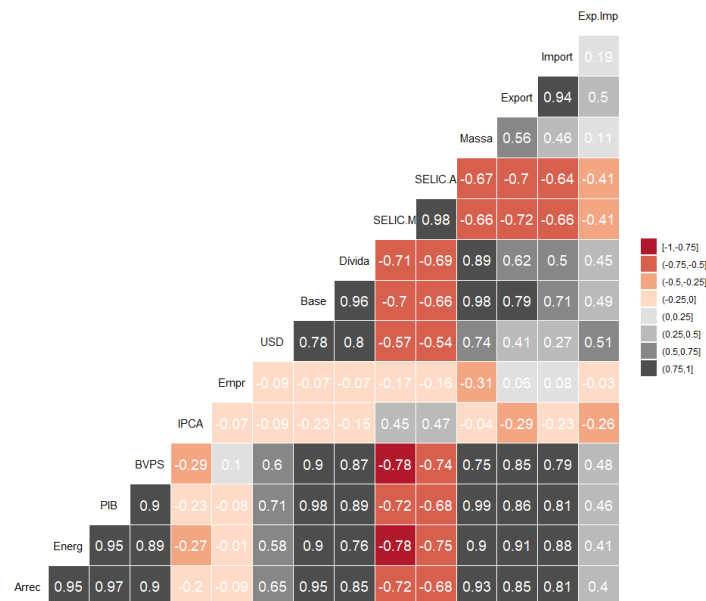


Figura 3. Matriz de Correlação de *Pearson* entre variáveis macroeconômicas.

A figura [1] ilustra graficamente as séries temporais da arrecadação *versus* a série de cada variável econômica. As figuras [2] e [3] apresentam as correlações entre as variáveis macroeconômicas.

	Arrec	Energ	PIB	BVPS	IPCA	Empr	USD	Base	Dívida	SELIC-M	SELIC-A	Massa	Export	Import	Exp-imp
Arrec	1,00	0,95	0,97	0,90	-0,20	-0,09	0,65	0,95	0,85	-0,72	-0,68	0,93	0,85	0,81	0,40
Arrec-1995	1,00	-0,02	0,38	0,13	0,37		0,39	0,21		0,10	-0,26		0,55	0,53	0,05
Arrec-1996	1,00	0,19	0,30	0,29	-0,57	-0,62		0,42		-0,30	-0,36		-0,23	0,44	-0,65
Arrec-1997	1,00	0,23	0,31	0,11	-0,12	-0,68		0,66		0,57	0,41		0,40	0,47	-0,26
Arrec-1998	1,00	-0,24	-0,59	0,07	0,23	-0,44	-0,18	-0,26	-0,09	0,08	0,02		-0,34	-0,30	0,02
Arrec-1999	1,00	0,36	0,53	0,62	0,35	-0,65	0,06	0,71	0,42	-0,27	-0,32		0,39	0,25	0,31
Arrec-2000	1,00	0,51	0,41	-0,26	0,03	-0,96	0,65	0,48	0,59	-0,51	-0,48		-0,08	0,26	-0,57
Arrec-2001	1,00	-0,47	0,53	-0,21	0,46	-0,46	0,32	0,50	0,51	0,67	0,54		0,16	-0,05	0,27
Arrec-2002	1,00	0,21	0,39	-0,56	0,49	-0,46	0,68	0,64	0,66	0,74	0,44		0,61	0,25	0,63
Arrec-2003	1,00	0,70	0,42	0,60	0,08	-0,52	-0,21	0,52	-0,08	-0,42	-0,62		0,24	0,42	0,03
Arrec-2004	1,00	0,65	0,48	0,42	-0,16	-0,18	-0,35	0,59	0,29	0,28	0,14	0,62	0,56	0,54	0,46
Arrec-2005	1,00	0,49	0,55	0,29	-0,17	-0,63	-0,22	0,51	0,59	0,19	-0,34	0,61	0,28	0,19	0,29
Arrec-2006	1,00	0,48	0,76	0,46	-0,26	-0,53	-0,24	0,81	0,75	-0,71	-0,85	0,59	0,54	0,37	0,58
Arrec-2007	1,00	0,64	0,69	0,71	0,60	-0,77	-0,68	0,67	0,83	-0,40	-0,63	0,84	0,49	0,46	0,09
Arrec-2008	1,00	-0,02	0,42	-0,58	-0,31	-0,65	0,53	0,48	-0,32	0,73	0,50	0,58	0,00	0,00	0,00
Arrec-2009	1,00	0,75	0,74	0,63	-0,04	-0,35	-0,53	0,60	0,57	-0,25	-0,40	0,60	0,37	0,67	-0,31
Arrec-2010	1,00	0,56	0,58	0,36	0,24	-0,85	-0,56	0,57	0,46	0,51	0,43	0,67	0,46	0,20	0,58
Arrec-2011	1,00	0,20	0,42	-0,31	-0,28	-0,65	0,27	0,42	0,02	-0,12	-0,27	0,51	0,04	-0,05	0,17
Arrec-2012	1,00	0,06	0,11	0,17	0,57	-0,65	-0,01	0,18	0,04	0,00	-0,07	0,36	-0,38	-0,11	-0,43
Arrec-2013	1,00	0,66	0,49	0,09	0,63	-0,68	0,28	0,34	-0,22	0,52	0,51	0,59	0,05	0,08	-0,01
Arrec-2014	1,00	0,16	0,36	-0,14	0,06	-0,60	0,64	0,33	0,12	0,40	0,17	0,48	-0,35	-0,07	-0,41
Arrec-2015	1,00	0,24	0,18	-0,30	0,23	-0,61	-0,11	0,05	0,46	0,13	-0,16	0,51	-0,03	-0,14	0,07
Arrec-2016	1,00	0,05	0,17	0,31	0,02	-0,35	-0,12	0,25	0,23	-0,30	-0,42	0,43	-0,51	-0,29	-0,51
Arrec-2017	1,00	0,35	0,42	0,43	0,40	-0,62	0,34	0,22	0,24	-0,28	-0,32	0,73	-0,44	0,10	-0,55
Arrec-2018	1,00	0,28	0,28	0,45	-0,14	-0,42	-0,24	0,04	0,36	0,33	0,35	0,68	-0,25	-0,13	-0,11
Arrec-2019	1,00	0,19	0,21	0,39	0,40	-0,48	-0,28	-0,24	0,17	-0,07	-0,32	0,66	-0,05	0,17	-0,27
Arrec-2020	1,00	0,85	0,76	0,69	0,72	0,51	-0,33	0,36	0,10	-0,07	-0,14	0,53	-0,27	0,68	-0,67
Arrec-2021	1,00	1,00	-1,00	1,00	-1,00	-1,00	-1,00	-1,00	-1,00	1,00			-1,00	1,00	-1,00

Tabela 2. Correlação de *Pearson* entre a Arrecadação e variáveis macroeconômicas ano a ano.

E por fim, a tabela [2] apresenta a correlação entre a arrecadação e essas variáveis ano por ano.

Posto isso, tentou-se verificar o grau de acerto de previsões aplicando-se modelagem de redes neurais sobre a série temporal da arrecadação federal.

## 5. Criação de Modelos de *Machine Learning*

Para a criação do modelo de *machine learning* foi utilizada a linguagem R, o **R Studio** (<https://rstudio.com/>) e as funções `forecast::nnetar()`, `forecast::forecast()` e `forecast::accuracy()` da biblioteca do pacote **forecast**<sup>12 13</sup>. A figura [1] ilustra o código fonte no ambiente do R Studio<sup>14</sup>.

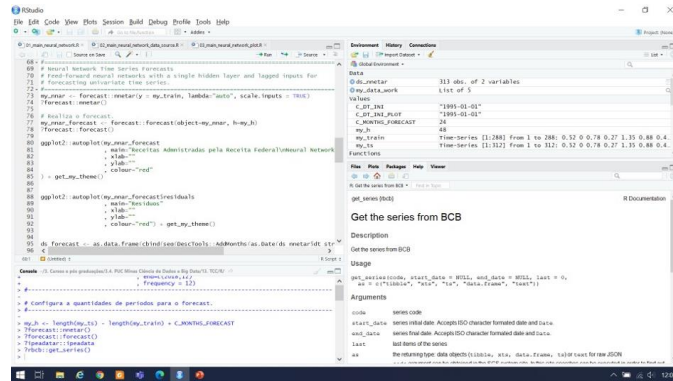


Figura 4. Ilustração do código fonte no ambiente R Studio.

A primeira função, ***nnetar()***, foi utilizada para a modelagem da rede neural:

*Neural Network Time Series Forecasts: Feed-forward neural networks with a single hidden layer and lagged inputs for forecasting univariate time series.*

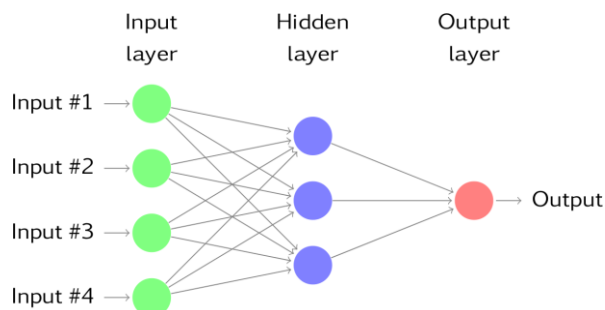


Figura 4. Ilustração de uma rede neural com quatro variáveis de entrada e uma camada interna com três neurônios.<sup>15</sup>

<sup>12</sup> *Forecasting: Principles and Practice*. Rob J Hyndman and George Athanasopoulos. Monash University, Australia. <https://otexts.com/fpp2/>

<sup>13</sup> *forecast: Forecasting Functions for Time Series and Linear Models. Methods and tools for displaying and analysing univariate time series forecasts including exponential smoothing via state space models and automatic ARIMA modelling*. CRAN - Package *forecast* ([r-project.org](https://cran.r-project.org/web/packages/forecast/index.html)) <https://cran.r-project.org/web/packages/forecast/index.html>

<sup>14</sup> O código fonte completo encontra-se nos anexos a esse documento.

<sup>15</sup> [11.3 Neural network models | Forecasting: Principles and Practice \(2nd ed\) \(otexts.com\)](https://otexts.com/fpp2/11.3-Neural-network-models/)

A segunda função, **forecast()**, foi utilizada para realizar a previsão da arrecadação a partir da modelagem obtida pela rede neural:

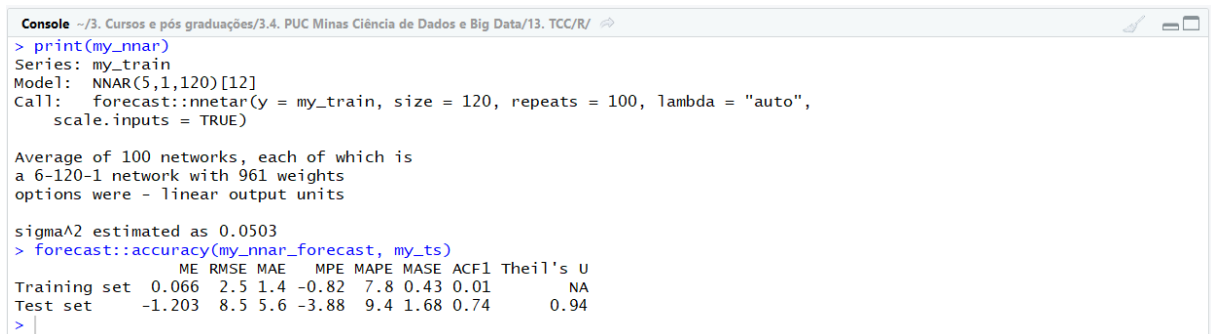
*Forecasting time series: forecast is a generic function for forecasting from time series or time series models. The function invokes particular methods which depend on the class of the first argument.*

E a terceira função foi utilizada para medir o grau de acuracidade do modelo.

*Accuracy measures for a forecast model: Returns range of summary measures of the forecast accuracy. If x is provided, the function measures test set forecast accuracy based on x-f. If x is not provided, the function only produces training set accuracy measures of the forecasts based on f["x"]-fitted(f). All measures are defined and discussed in Hyndman and Koehler (2006).*

O período da série temporal da arrecadação federal foi de Jan/1995 a Fev/2021. O treinamento da rede neural foi até Dez/2018 e os testes de validação de Jan/2019 a Fev/2021. As previsões foram realizadas até Mai/2021.

A figura [3] e o gráfico [4] demonstram a modelagem obtida e o comportamento das previsões realizadas utilizando a rede neural:



```

Console ~/3. Cursos e pós graduações/3.4. PUC Minas Ciência de Dados e Big Data/13. TCC/R/
> print(my_nnar)
Series: my_train
Model: NNAR(5,1,120)[12]
Call: forecast::nnetar(y = my_train, size = 120, repeats = 100, lambda = "auto",
  scale.inputs = TRUE)

Average of 100 networks, each of which is
a 6-120-1 network with 961 weights
options were - linear output units

sigma^2 estimated as 0.0503
> forecast::accuracy(my_nnar_forecast, my_ts)
      ME RMSE MAE  MPE MAPE MASE ACF1 Theil's U
Training set  0.066  2.5 1.4 -0.82  7.8 0.43 0.01      NA
Test set     -1.203  8.5 5.6 -3.88  9.4 1.68 0.74     0.94
>

```

Figura 5. Rede Neural obtida a partir da série temporal da arrecadação federal.

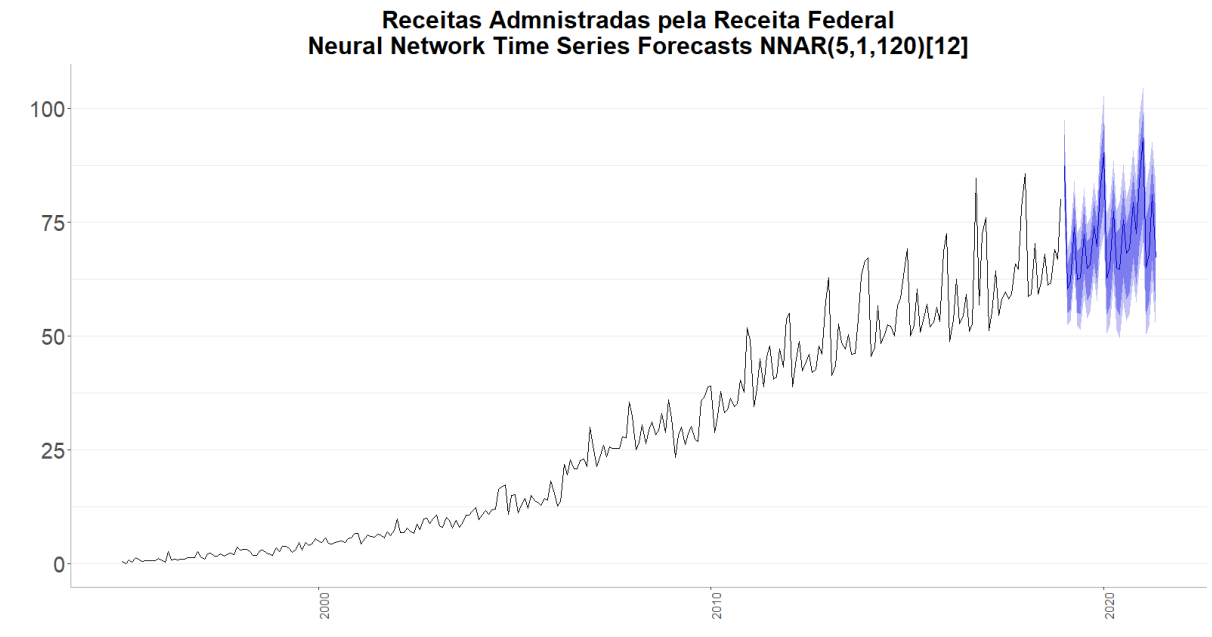


Gráfico 3. Previsão da Arrecadação Federal a partir da aplicação de redes neurais.

E a figura [4] e o gráfico [5] apresentam os resíduos/erros da modelagem. Os parâmetros de input da rede neural foram realizados por testes até que se conseguisse os menores resíduos possíveis.

```
Console ~/3. Cursos e pós graduações/3.4. PUC Minas Ciência de Dados e Big Data/13. TCC/R/ ➤
> forecast::accuracy(my_nnar_forecast, ts(my_ts, start = c(1995, 01), end = c(2021, 02), frequency = 12))
      ME RMSE MAE  MPE MAPE MASE ACF1 Theil's U
Training set  0.066  2.5  1.4 -0.82  7.8  0.43  0.01      NA
Test set     -1.203  8.5  5.6 -3.88  9.4  1.68  0.74     0.94
> |
```

Figura 6. Resíduos da rede neural obtida a partir da série temporal da arrecadação federal.

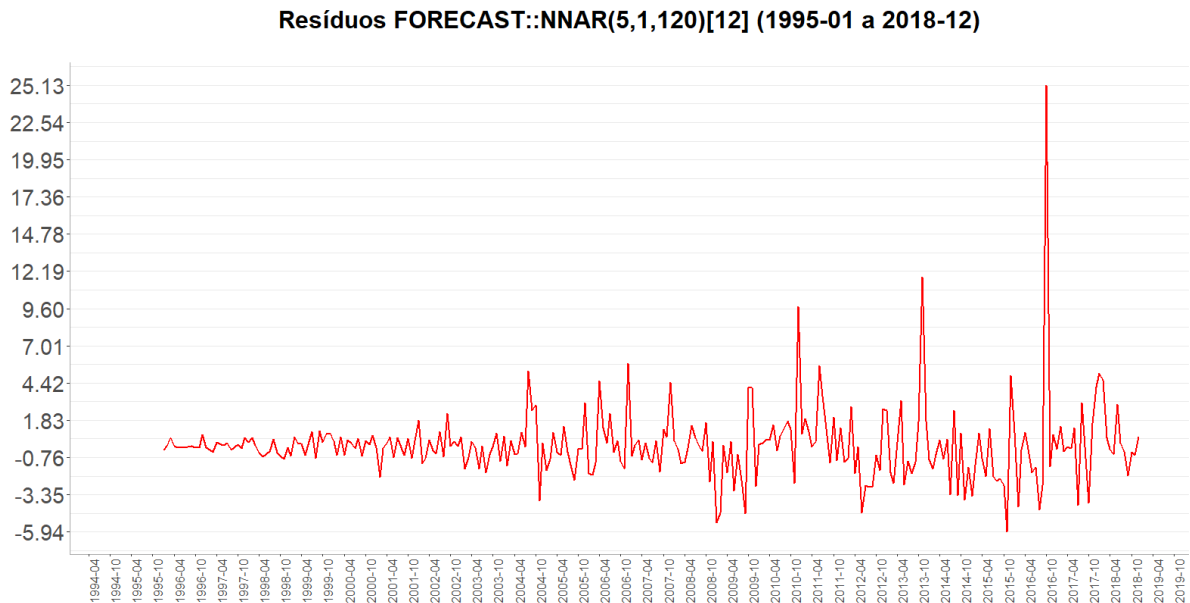


Gráfico 4. Resíduos da previsão da arrecadação Federal a partir da aplicação de redes neurais.

HYNDMAN & ATHANASOPOULOS em *Forecasting: Principles and Practice* (2nd ed) recomendam a utilização de diferentes métodos na mesma série de dados para a melhoria dos resultados. E mencionam que essa abordagem é recomendada também por BATES & GRANGER (1969) e CLEMEN (1989).

*An easy way to improve forecast accuracy is to use several different methods on the same time series, and to average the resulting forecasts. Nearly 50 years ago, John Bates and Clive Granger wrote a famous paper (Bates & Granger, 1969), showing that combining forecasts often leads to better forecast accuracy. Twenty years later, Clemen (1989) wrote:*

*“The results have been virtually unanimous: combining multiple forecasts leads to increased forecast accuracy. In many cases one can make dramatic performance improvements by simply averaging the forecasts.”*

Diante disso, com o objetivo de se tentar diminuir os resíduos, tentou-se também combinar três modelos de séries temporais: ETS, redes neurais e TBATS.

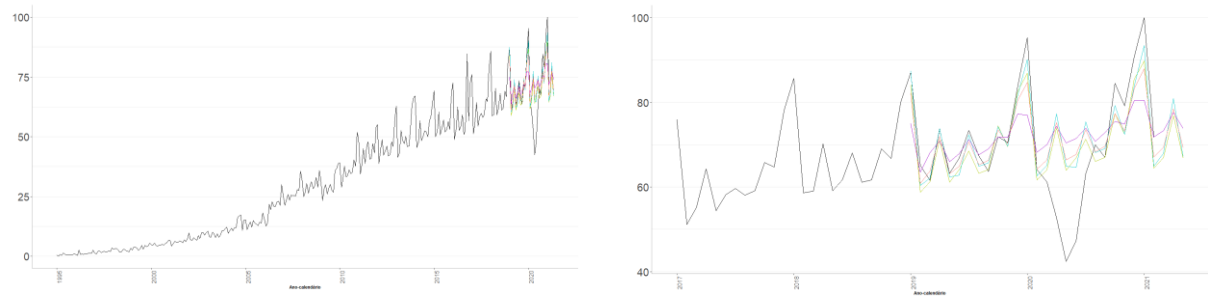


Gráfico 5. Combinação de modelos ETS, NNAR e TABTS

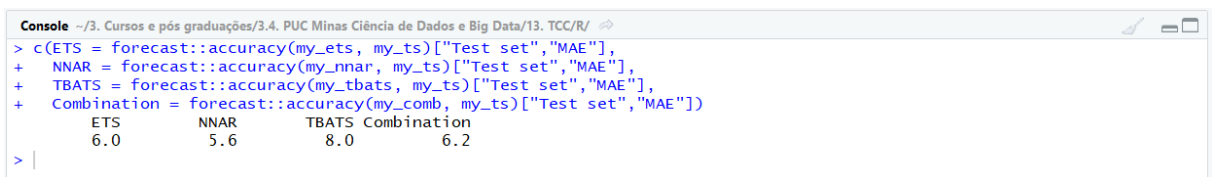


Figura 7. Resíduos da rede neural obtida a partir da combinação de modelos.

O modelo ETS (MAE=6.0) apresentou melhor resultado em relação ao TBATS (MAE=8.0), mas inferior à modelagem com redes neurais (MAE=5.6). A combinação de modelos (MAE=6.2), nesse caso específico, também não superou o algoritmo de rede neural<sup>16</sup>.

## 6. Apresentação dos Resultados

O período da série temporal da arrecadação federal utilizado foi de Jan/1995 a Jan/2021. O treinamento da rede neural foi até Dez/2018 e os testes de validação de Jan/2019 a Fev/2021. As previsões foram realizadas até Mai/2021. Os gráficos seguintes apresentam os resultados obtidos e a comparação entre a série temporal real e a projetada pela rede neural.

<sup>16</sup> [Mean absolute error - Wikipedia https://en.wikipedia.org/wiki/Mean\\_absolute\\_error](https://en.wikipedia.org/wiki/Mean_absolute_error)

*In statistics, mean absolute error (MAE) is a measure of errors between paired observations expressing the same phenomenon. Examples of Y versus X include comparisons of predicted versus observed, subsequent time versus initial time, and one technique of measurement versus an alternative technique of measurement.*

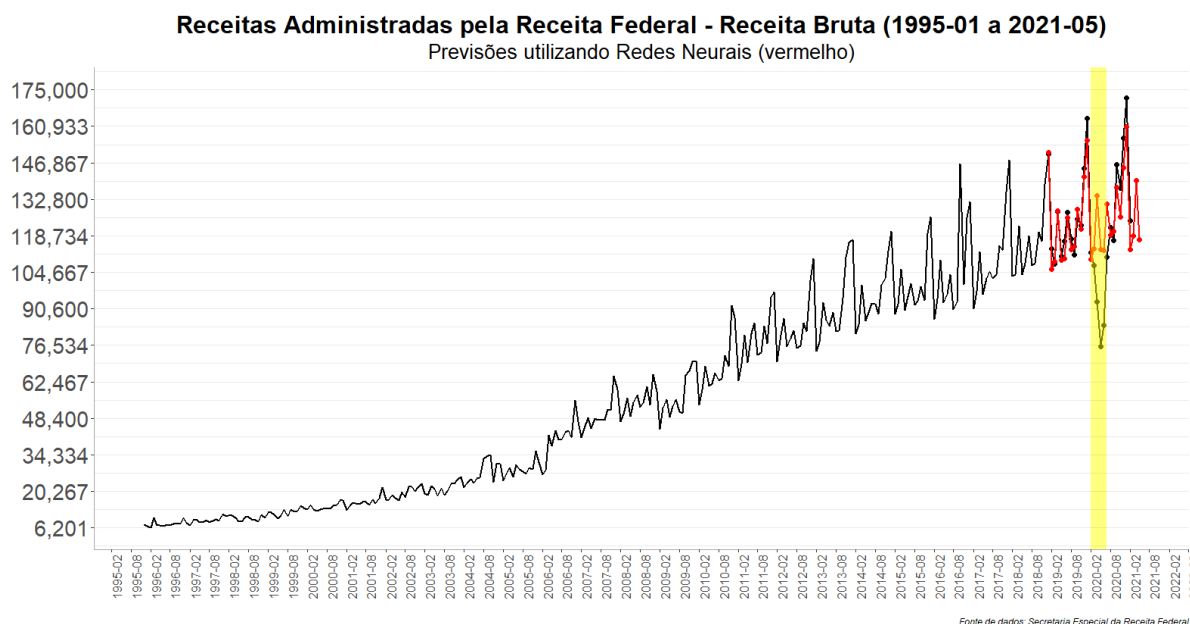


Gráfico 6. Arrecadação Federal e previsão até maio/2021.

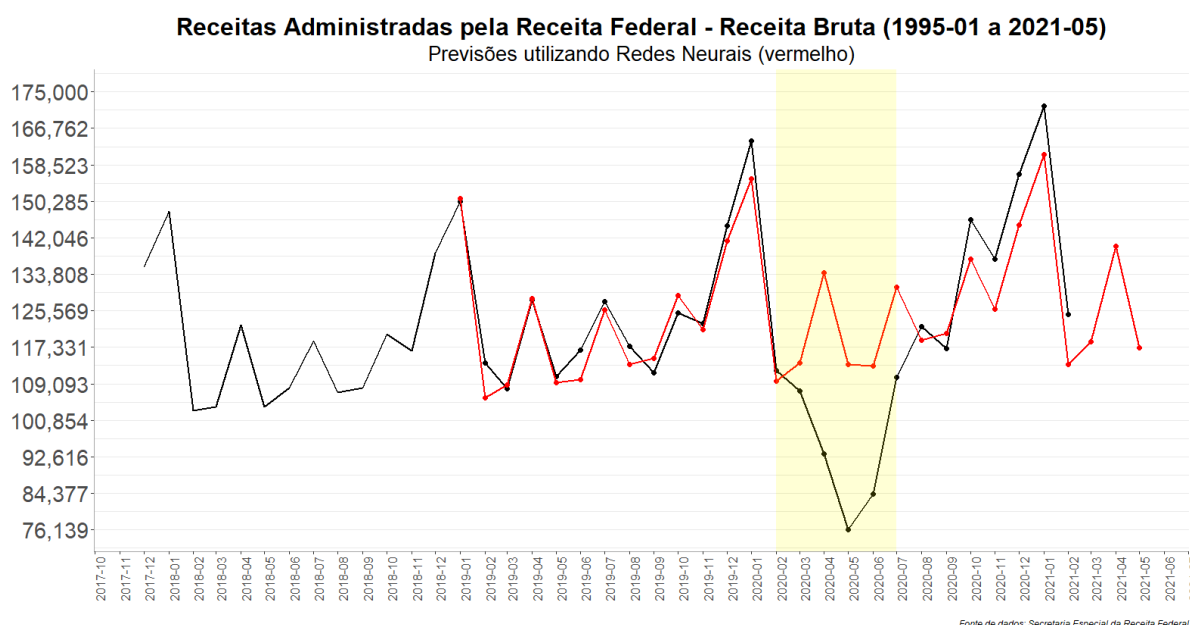


Gráfico 7. Arrecadação Federal e previsão até maio/2021 (zoom do gráfico anterior).

Observa-se, a partir dos gráficos [5] e [6] que, salvo os períodos de Abril/2020 a Julho/2020, a curva dos valores previstos (em vermelho) ficou próxima dos valores reais. A figura [5] destaca os percentuais de erro nesses quatro meses: períodos em que a arrecadação federal foi mais impactada pela pandemia do Covid-19.

RStudio Source Editor

ds\_resultado

Filter

	dt_str	value	value_forecast	value_orig	value_norm	value_orig_forecast	desvio_perc	desvio_perc_orig
286	2018-10-01	120310	NA	120310	69.01	NA	NA	NA
287	2018-11-01	116587	NA	116587	66.77	NA	NA	NA
288	2018-12-01	138615	NA	138615	80.03	NA	NA	NA
289	2019-01-01	150298	87	150298	87.06	150861	-0.39	-0.37
290	2019-02-01	113804	60	113804	65.09	105971	7.24	6.88
291	2019-03-01	107912	62	107912	61.55	108807	-0.88	-0.83
292	2019-04-01	127999	74	127999	73.64	128289	-0.24	-0.23
293	2019-05-01	110753	62	110753	63.26	109282	1.40	1.33
294	2019-06-01	116729	63	116729	66.85	110001	6.06	5.76
295	2019-07-01	127637	72	127637	73.42	125810	1.50	1.43
296	2019-08-01	117533	65	117533	67.34	113474	3.63	3.45
297	2019-09-01	111523	66	111523	63.72	114750	-3.05	-2.89
298	2019-10-01	125169	74	125169	71.93	128995	-3.20	-3.06
299	2019-11-01	122741	70	122741	70.47	121353	1.19	1.13
300	2019-12-01	144817	82	144817	83.76	141431	2.43	2.34
301	2020-01-01	163948	90	163948	95.27	155379	5.41	5.23
302	2020-02-01	112141	63	112141	64.09	109749	2.25	2.13
303	2020-03-01	107390	65	107390	61.23	113861	-6.36	-6.03
304	2020-04-01	93332	77	93332	52.77	134182	-46.59	-43.77
305	2020-05-01	76139	65	76139	42.42	113534	-53.06	-49.11
306	2020-06-01	84267	65	84267	47.32	113032	-36.59	-34.13
307	2020-07-01	110540	75	110540	63.13	130969	-19.48	-18.48
308	2020-08-01	121993	68	121993	70.02	118891	2.67	2.54
309	2020-09-01	117022	69	117022	67.03	120417	-3.05	-2.90
310	2020-10-01	146081	79	146081	84.52	137322	6.24	6.00
311	2020-11-01	137180	72	137180	79.16	126018	8.49	8.14
312	2020-12-01	156369	84	156369	90.71	144907	7.60	7.33
313	2021-01-01	171798	93	171798	100.00	160904	6.56	6.34
314	2021-02-01	124693	65	124693	71.65	113533	9.38	8.95
315	2021-03-01	NA	68	NA	NA	118560	NA	NA
316	2021-04-01	NA	81	NA	NA	140138	NA	NA
317	2021-05-01	NA	67	NA	NA	117252	NA	NA

Showing 285 to 317 of 317 entries

Figura 8. Erros mensais percentuais da rede neural.

Por fim, recomenda-se que novas tentativas sejam realizadas utilizando-se como *input* da rede neural novas variáveis, como por exemplo energia elétrica, PIB, taxas de juros para se tentar o aperfeiçoamento e a previsibilidade da rede neural em períodos atípicos. Recomenda-se também que sejam realizados novos estudos com outros algoritmos que implementem as redes neurais: *Recurrent Neural Networks (RNN) Long Short Term Model (LSTM)*. Esses tipos de redes neurais são boas opções para previsões de séries temporais, pois se valem também do tempo no aprendizado. As LSTM utilizam aprendizados atuais para retroalimentar o processo de aprendizagem<sup>17</sup>.

<sup>17</sup> PAWLUS, Michael. RODGER, Devine. *Hands-On Deep Learning With R. April 2020. Packt. Chapter 8.*



## REFERÊNCIAS

- HYNDMAN, Rob J. ATHANASOPOULOS, George. *Forecasting: Principles and Practice* (2nd ed). Monash University, Australia. <https://otexts.com/fpp2/>
- LANTZ, Brett. *Machine Learning With R. Third Edition* April 2019. Packt.
- PAWLUS, Michael. DEVINE, Rodger. *Hands-On Deep Learning With R. April 2020. Packt.*
- KRISPIN, Rami. *Hands-On Time Series Analysis with R.*
- *Time Series Forecasting using LSTM in R*  
<http://rwanjohi.rbind.io/2018/04/05/time-series-forecasting-using-lstm-in-r/>
- *Deep Neural Network in R*  
<https://www.r-bloggers.com/2021/04/deep-neural-network-in-r/>

## 7. Links

Os arquivos com o código fonte e o vídeo de apresentação encontram-se como anexos desse relatório.

## ANEXOS

Anexo I: Código fonte R. Arquivos estão anexados e também disponíveis no endereço [https://github.com/amarallr/ml\\_arrecadacao](https://github.com/amarallr/ml_arrecadacao)

Anexo II: Vídeo de apresentação.