

Ciência de Dados Aplicada à uma Análise do Desemprego no Brasil Antes e Depois da Covid-19: Uma abordagem por Algoritmos de Séries Temporais

Bianca A. Barroso, Danielle P. Moura, Larissa Navarro, Thaina S. Reis
Orientador: Prof. Dr. Arnaldo R. A. Vallim Fo.

Faculdade de Computação e Informática (FCI)
Universidade Presbiteriana Mackenzie
São Paulo, SP - Brasil

{biancaabarroso, daniellepmoura}@gmail.com, {lanavarro_, thaina-santos-reis}@hotmail.com, arnaldo.vallim@mackenzie.br

Abstract. *This document describes the research and in observations of a Final Paper – TCC of an undergraduate course, which aimed to study the effects of the Coronavirus pandemic (SARS-CoV-2) on unemployment in Brazil, through a "before and after" study. The analysis was developed through Data Science techniques, with an approach of time series algorithms in the period between 2012 and 2021, based on statistical models following the Box-Jenkins methodology, the SARIMAX model that presented the best solution for the forecast. The application of the model makes it possible to demonstrate the impact of the pandemic on the behavior of the unemployment time series in the Brazil as a tool of risk management and decision making.*

Resumo. *Este documento descreve as pesquisas e observações de um Trabalho de Conclusão de Curso - TCC, que teve como objetivo estudar os efeitos da pandemia do Coronavírus (SARS-CoV-2) no desemprego do Brasil, por meio de um estudo do tipo “antes e depois”. A análise foi desenvolvida por meio de técnicas de Ciência de Dados, com uma abordagem de algoritmos de séries temporais no período entre 2012 e 2021, baseados em modelos estatísticos seguindo a metodologia Box-Jenkins, no qual o modelo SARIMAX apresentou a melhor solução para a previsão. A aplicação do modelo possibilita demonstrar o impacto da pandemia no comportamento da série temporal do desemprego no país como uma ferramenta de auxílio ao gerenciamento de risco e tomada de decisão.*

1. Introdução

Em março de 2020, a Organização Mundial da Saúde (OMS) declarou que o mundo estava diante de uma pandemia. A rápida propagação do vírus do COVID-19 gerou um problema de saúde pública global e nos obrigou a enfrentar uma crise socioeconômica. Surge então uma nova dinâmica para a economia mundial, devido a exigência do distanciamento social e medidas de restrições, com bloqueios parciais ou totais, realizadas por vários países na tentativa de controlar a transmissão do vírus.

Como consequência a este cenário, houve a exposição da desigualdade social e o aumento das taxas de desemprego nacional, impactando mais severamente as pessoas que vivem na informalidade e as que estão às margens dos cargos dos diversos setores

trabalhistas. É proposto então analisar os dados das taxas de desemprego no Brasil antes e durante a pandemia do COVID-19, gerando séries temporais e avaliar como a situação pandêmica afetou a classe trabalhadora.

O objetivo deste trabalho é responder à pergunta: “Como a pandemia afetou o desemprego no país, e qual seria esse impacto quando medido por modelos de análise e previsão de séries temporais?”.

Para a definição da solução ótima do problema, considera-se que seria estabelecida em função da seguinte hipótese: A pandemia afeta diretamente o nível de emprego no trabalho formal.

A utilização de séries temporais comparando o nível de desemprego real no Brasil e a previsão de como seria sem a pandemia, tem a finalidade de observar ao longo da passagem de tempo os números e o seu comportamento. Assim, busca-se utilizar modelos estatísticos implementados em algoritmos utilizando a linguagem de programação *Python* para elaboração de gráficos demonstrando:

- i) Previsões das taxas de desemprego sem a pandemia;
- ii) Taxas de desemprego reais antes e ao longo da pandemia;
- iii) Apresentar os resultados obtidos por meio de séries temporais.

2. Referencial Teórico

A rápida disseminação do vírus e as medidas de contenção propostas pela Organização Mundial da Saúde (OMS) causaram inúmeros prejuízos à economia mundial. No Brasil, as estatísticas do Ministério da Saúde (MS) contabilizaram até 2 de julho de 2021, mais de 1,5 milhões de casos confirmados.

Latorre & Cardoso (2021) definem séries temporais como “sequências de dados quantitativos relativos a momentos específicos e estudados segundo sua distribuição no tempo”, sendo assim uma maneira de organizar as informações podendo facilitar a busca por previsões e discutir os efeitos da pandemia. Analisando contexto do mercado de trabalho, Barbosa Filho (2017) tem em seu artigo um resumo sociopolítico e econômico do Brasil entre os anos de 2014 a 2017, descrevendo a situação de forte recessão que a economia brasileira se encontrava a partir do segundo trimestre de 2014, durante o governo de Dilma Rousseff, resultado de um conjunto de choques de oferta e demanda, que também segundo o autor foram ocasionados por “erros de políticas públicas que reduziram a capacidade de crescimento da economia brasileira e geraram um custo fiscal elevado”.

No período de pré-pandemia, que antecede março de 2020, o cenário econômico brasileiro era de superação da crise entre 2014 e 2017 e estes fatores provocaram uma redução do crescimento da economia brasileira e incertezas sobre a solvência das finanças públicas (BARBOSA FILHO, 2017).

De fevereiro a março de 2020, houve desaceleração da economia chinesa, epicentro do surgimento da doença e, posteriormente, na Europa, sendo o primeiro e terceiro principais parceiros comerciais do Brasil respectivamente (COMEXSTAT, 2020). O país então sofre uma redução de expectativas da parte dos investidores, além da falta de vários suprimentos com a redução na exportação.

Em um segundo período, entre os meses de abril e julho de 2020, ainda devido às políticas restritivas ao redor do mundo, a economia se retrai, gerando a perda de empregos, diminuição da renda, introdução massiva do trabalho remoto dada a exposição a situações de risco de contágio, e sendo esses os desafios em maior destaque.

De acordo com o SEBRAE (2020a) o setor empresarial brasileiro é composto por aproximadamente 99% de micro e pequenas empresas, e em 2020, 60% dos proprietários de pequenos negócios não tiveram acesso a crédito, tendo pedidos negados pelos bancos por falta de comprovação de garantias ao pagamento. Este setor é responsável por 52% dos empregos formais do setor privado (SEBRAE, 2020b) e sem acesso ao crédito se viram obrigadas a reduzir as jornadas de trabalho e salários.

Nesse artigo, será feita uma breve reflexão das consequências para os trabalhadores brasileiros e todo esse período de incertezas reflete nos dados selecionados e nos resultados obtidos após as análises das séries temporais que de acordo com Reis (s.d) são variações irregulares, flutuações que são resultado de fatos fortuitos e inesperados como catástrofes naturais, atentados terroristas – exemplo o atentado de 11 de setembro de 2001 - decisões intempestivas dos governos, e no presente estudo, a pandemia do COVID-19.

Antunes & Cardoso (2021) utilizam as séries temporais em estudos epidemiológico e mostram a eficácia de se aplicar este modelo para as análises quantitativas, no qual afirmam que para se antever o futuro é primordial conhecer os valores passados, os quais feitos por regressão linear, considerando-se o melhor ajuste dos pontos da série. Já as previsões podem ser feitas com procedimentos estatísticos mais complexos, como o método ARIMA (*autoregressive integrated moving average*), cuja aplicação demanda treinamento especializado adicional.

2.1. Séries Temporais

As séries temporais utilizam um período ao longo do tempo, trata-se do conjunto de observações ordenadas, não necessariamente são espaçadas igualmente, mas é crucial ser de forma sequencial, podendo ser estacionárias ou não estacionárias. Para ser estacionária, o comportamento da série não pode se alterar ao passar do tempo, portanto a média, tendência e sazonalidade deve ser constante.

O objetivo da análise com séries temporais é identificar padrões não aleatórios de uma variável e o comportamento dela pode nos orientar e obter *insights* sobre o futuro para uma possível tomada de decisões.

A maneira tradicional de se fazer uma análise de série temporal é através de seus componentes de tendência, sazonalidade e ciclo. A tendência ocorre quando existe um padrão de crescimento ou decréscimo na série. Os ciclos são as oscilações de subida e queda das séries, ao longo do componente da tendência. A sazonalidade corresponde às flutuações periódicas, ou seja, todo ano, mês, semana ou até mesmo dia a dia. A diferença entre componentes da sazonalidade e ciclos, é que a sazonalidade é algo previsível, enquanto ciclos tendem a ser irregulares.

De acordo com Morettin (2018) dos modelos utilizados, a modelagem ARIMA merece um grande destaque dada a sua ampla divulgação e utilização, dificuldade de aplicação (não automática) e aparente superioridade em várias situações. O presente estudo se baseou principalmente na metodologia *Box-Jenkins*, que afirma que o modelo

ARIMA é recomendado quando o processo estocástico é não estacionário, Box e Jenkins (1976). “A utilidade da aplicação dos operadores de diferença reside no fato de que eles são capazes de deixar as séries estacionárias, o que significa que esses operadores não somente estabilizam a variância, como também removem a tendência que está por trás das séries originais, tornando-as estáveis” (MARGARIDO, 1998, p. 76).

2.1.1 Box-Jenkins ou modelo ARIMA

A metodologia de *Box-Jenkins*, ou como também conhecida como modelo ARIMA (Modelo Autorregressivo Integrado de Médias Móveis), possui características flexíveis, que permite ajustes e adaptações nos parâmetros da série temporal, assim, contribuindo para que diversos tipos de séries temporais sejam representados, assim tendo uma estrutura de correlação linear entre os valores das séries. Os modelos descrevem duas classes de processos:

- i) Processos lineares estacionários: quando a série apresenta ao longo do tempo, média constante ou variando em torno na média.
- ii) Processos lineares não estacionários: no qual há movimentos de tendência, seja ela de crescimento ou queda de qualquer natureza. Para torná-las estacionárias, deve-se aplicar um número determinado de diferenças entre os dados de ordem, assim, para esses processos o modelo é representado por ARIMA.

O número de vezes em que a série original tem de ser referenciada antes de resultar em uma série estacionária é denominada como ordem de homogeneidade. De acordo com Camelo (2018) esses processos são descritos por modelos autorregressivos, integrados e médias móveis de ordem p , d e q que podem ainda ter incluído um operador sazonal, representado por SAMIRA (p,d,q). A fórmula do modelo ARIMA é dada por:

$$\phi(B)(1 - B)^d X_t = \theta(B)\varepsilon_t$$

Onde $\phi(B)$ é o polinômio autorregressivo, $\theta(B)$ é o polinômio de médias móveis, ε_t é um ruído branco com média zero e variância σ^2 e d é a ordem de integração, isto é, o número de diferenças aplicadas à série para que esta se torne estacionária. Quando adicionamos a parte sazonal no modelo ARIMA, passamos a ter o modelo SARIMA.

2.1.2 Modelos SARIMA e SARIMAX

O modelo SARIMA, representado pela fórmula:

$$\Phi(B^s)\phi(B)(1 - B)^d(1 - B^s)^D X_t = \Theta(B^s)\theta(B)\varepsilon_t$$

Onde (p , d , q) são as ordens do modelo referentes à dinâmica ordinal e (P , D , Q) são às ordens da parte sazonal.

O modelo SARIMAX (média móvel integrada auto-regressiva sazonal com fatores exógenos) é uma versão atualizada do modelo ARIMA com adição de uma ou mais variáveis exógenas. Evidenciando X_t na equação, e adicionando os termos relacionados à variável exógena y , que descreve um modelo SARIMAX:

$$X_t = f(y) + \left(\frac{\theta(B^s)\theta(B)\varepsilon_t}{\Phi(B)\phi(B)(1-B)^d(1-B^s)^D} \right)$$

2.1.3 Auto ARIMA

O Auto ARIMA define qual melhor modelo e parâmetros que se adaptam a uma variedade de modelos ARIMA e seleciona automaticamente o que teve o melhor desempenho. Normalmente, no modelo básico do ARIMA, precisamos fornecer os valores de p , d e q que são essenciais. Usamos técnicas estatísticas para gerar esses valores realizando a diferença para eliminar a não estacionaridade, plotando os gráficos ACF (análise de funções de autocorrelação) e PACF (análise de funções de autocorrelação parcial). No Auto ARIMA, o próprio modelo gera os valores ótimos de p , d e q adequados para o conjunto de dados fornecer uma melhor previsão.

3. Metodologia da Pesquisa

A finalidade desta pesquisa é do tipo aplicada, pois estuda um problema atual e busca hipóteses para a aplicação de possíveis soluções e reflexões para o “novo normal”. É realizada uma pesquisa exploratória para identificar os fatos, com métodos hipotético-dedutivo dada uma abordagem quanti-quali, trabalhando com séries temporais e ciência de dados.

Os meios a serem utilizados na pesquisa, serão: uma revisão bibliográfica e a utilização de bases de dados públicas, bem como, técnicas de análise de dados, séries temporais e ferramentas tecnológicas *open source* para execução de *scripts* em *Python*.

A sequência deste projeto está dividida em cinco etapas descritas abaixo no fluxograma abaixo (Figura 1):

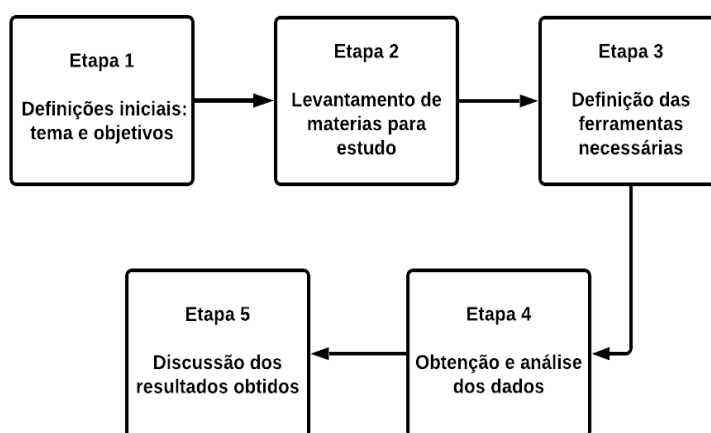


Figura 1. Fluxograma das etapas da metodologia

- i) Etapa 1: Determinação do tema a ser estudado, levantamento de informações para discussão sobre como será aplicado e qual o objetivo geral do estudo.
- ii) Etapa 2: Levantamento e revisão bibliográfica de estudos com a mesma aplicabilidade. Etapa essencial para ter um conhecimento mais aprofundado antes de realizar o desenvolvimento da parte prática.
- iii) Etapa 3: Define-se quais as ferramentas necessárias para fazer o desenvolvimento do estudo, quais as linguagens de programação, softwares mais adequados para o desenvolvimento do projeto.
- iv) Etapa 4: Obtenção dos dados através do IBGE - Instituto Brasileiro de Geografia e Estatística, sobre desemprego no Brasil desde 2012. Nessa etapa também é

realizada a análise desses dados e a implantação do método que se julga o melhor aplicado à situação.

- v) Etapa 5: A última etapa consiste em discutir os resultados obtidos com a análise de séries temporais e analisar a eficácia da metodologia.

3.2. Séries Temporais e a Análise das Taxas de Desemprego Pré e Pós a Pandemia

As bases usadas são da Pesquisa Nacional por Amostra de Domicílios Contínua – “Pessoas de 14 anos ou mais de idade, desocupadas na semana de referência”, no período de 2012 a 2021 (IBGE – Instituto Brasileiro de Geografia e Estatística). O tamanho da série temporal será de 9 anos – 108 meses.

No geral, busca-se realizar uma análise e modelagem de série temporal para verificar características importantes do desemprego durante o COVID-19 e relacioná-las com períodos pré e pós a pandemia.

Foram utilizados modelos estatísticos uni variados, em que se baseiam somente uma única série histórica como fonte.

O modelo escolhido para o desenvolvimento foi o ARIMA com componente sazonal, o SARIMA. A aplicação do modelo segue as seguintes etapas: identificação, estimação, verificação e previsão.

Na identificação do modelo empregam-se procedimentos para identificar a sua estrutura e verificar se as hipóteses foram cumpridas, com isso, é possível separar em duas partes. Primeiro, diferencia-se a série original quantas vezes forem necessárias para termos uma série estacionária. Segundo, a identificação do processo, em que é feita através de autocorrelações. A estimativa de parâmetros usa procedimentos iterativos de estimação de mínimos quadrados não lineares. A verificação busca capturar toda a estrutura de dependência serial presente, assim, os resíduos do modelo não devem apresentar qualquer tipo de dependência no tempo, ou seja, não deve conter autocorrelações. Por fim, a previsão de valores, assim que identificado o processo estocástico que gera a série temporal de interesse e os parâmetros do modelo, pode-se prever os valores futuros ou *forecast*.

4. Desenvolvimento do estudo

4.1. Levantamento e tratamento dos dados

Para o início da análise e posteriormente aplicar o *Box-Jenkins*, primeiramente foi realizado o levantamento de dados de desemprego no Brasil.

O conjunto de dados tem periodicidade trimestral começando em janeiro de 2012 até dezembro de 2021.

Tabela 1. Amostra de dados da PNAD de desemprego no Brasil

Ano	Trimestre	Estimativa	Coefficiente	Varição ao Trimestre (%)	Varição ao Trimestre
2012	jan-fev-mar	7559	1		
	abr-mai-jun	7245	1,1	-4,2	-314
	jul-ago-set	6815	1,2	-5,9	-430
	out-nov-dez	6611	1,1	-3	-204

2013	jan-fev-mar	7704	1,1	16,5	1093
	abr-mai-jun	7225	1,1	-6,2	-479
	jul-ago-set	6753	1,1	-6,5	-472
	out-nov-dez	6013	1,2	-11	-739

As principais colunas da tabela são:

- i) Ano: o ano em que a pesquisa foi realizada.
- ii) Trimestre de coleta e de referência: trimestre referente a coleta dos dados.
- iii) Estimativa (em milhares): Pessoas de 14 anos ou mais de idade, desocupadas na semana de referência que participaram da pesquisa.
- iv) Coeficiente de variação (%): também conhecido como desvio padrão, mostra a extensão da variabilidade em relação à média da população.
- v) Variação em relação ao trimestre anterior (%): valor em porcentagem de quanto a estimativa variou em relação ao trimestre anterior.
- vi) Variação em relação ao trimestre anterior: valor absoluto de quanto a estimativa variou em relação ao trimestre anterior.

Em seguida foi feito um pré-processamento dos dados. A variável alvo, ou seja, a que se deseja prever é a coluna 'estimativaMilhares', por isso só foram mantidas as colunas de data e de quantidade de pessoas desocupadas.

Tabela 2. Amostra dos dados após tratamento

data	estimativaMilhares
2012-02-01	7559
2012-05-01	7245
2012-08-01	6815
2012-11-01	6611

- i) Data: trimestre referente a coleta dos dados, uma data 2020-01-01 refere-se ao trimestre jan-fev-mar, por exemplo.
- ii) Estimativa em milhares (estimativaMilhares): Pessoas de 14 anos ou mais de idade, desocupadas na semana de referência que participaram da pesquisa.

4.2. Análise Exploratória dos dados

A análise exploratória dos dados (AED) tem como objetivo descobrir quais as principais características do conjunto de dados, geralmente feito por métodos visuais, como representações gráficas e tabelas.

4.2.1. Resumo dos Dados

Para iniciar a AED foram analisados um conjunto de estatísticas descritivas da nossa variável alvo, 'estimativaMilhares': a quantidade de valores, a média, a mediana, o desvio padrão, o valor mínimo, os quartis da distribuição e o valor máximo.

Tabela 3. Estatísticas descritivas dos dados de desemprego (em milhares)

Medidas	estimativaMilhares
Quantidade	40.0
Média	10654.0
Mediana	11984.5
Desvio Padrão	2920.62
Valor Mínimo	6013.0
25%	7480.50
50%	11984.50
75%	12910.25
Valor Máximo	14805.00

4.2.2 Identificação das Características

O conjunto de dados do desemprego foi separado em dois grupos para realização da análise:

- i) Completo: o conjunto de dados completo, de 2012 ao último trimestre de 2021.
- ii) Pré Covid-19: descartando os dados do ano de 2020 e 2021.

A partir dos gráficos da Figura 2 (Completo e Pré Covid-19), que mostra a distribuição do conjunto de dados em relação ao tempo, nota-se alguns pontos importantes:

- i) Houve uma grande tendência de alta do ano de 2015 até 2017.
- ii) Do ano de 2017 até o final de 2021, não há uma tendência evidente na série temporal, olhando o gráfico (Completo).

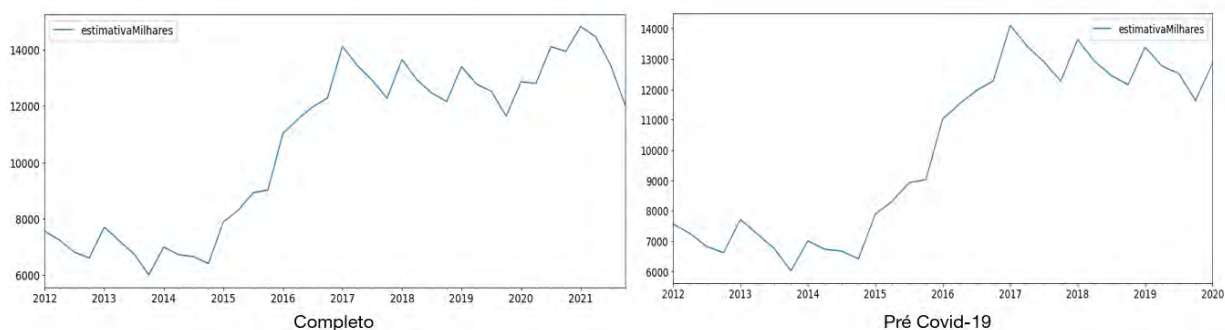


Figura 2. Evolução do desemprego no Brasil de 2012 a 2021 e Pré Covid-19

Ehlers (2009) diz que para validar uma tendência de uma série temporal, deve-se verificar se a sua média móvel não se mantém constante com a variação do tempo.

Na Figura 3, observa-se que a curva da média móvel, em vermelho, possui uma inclinação evidente em relação ao tempo. A grande tendência de alta nos anos de 2015 até 2017, comentada anteriormente, é representada pela curva de desvio padrão, em preto, mostrando uma variação neste período.

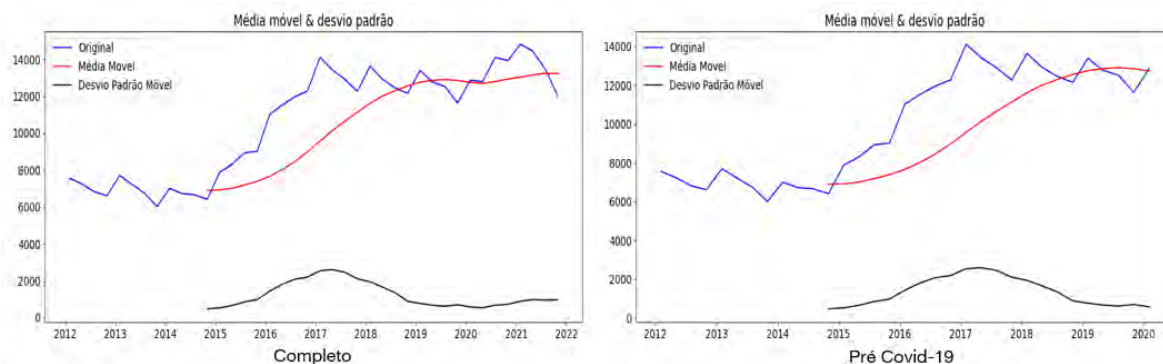


Figura 3. Validação da tendência com a média móvel

Assim, confirma-se a presença de tendência pela curva da média móvel, porém não podemos confirmar a presença de sazonalidade. Portanto, gerou-se outra análise, a decomposição da série temporal.

A partir da decomposição da série temporal (Figura 4) conclui-se:

- i) Houve uma tendência de crescimento até 2017, e a partir desse ano até 2021 não apresentou uma tendência evidente.
- ii) Existe sazonalidade nos dados, aproximadamente a cada 1 ano.
- iii) A partir de 2017, houve um aumento nos resíduos, o mesmo período em que não há tendência evidente.

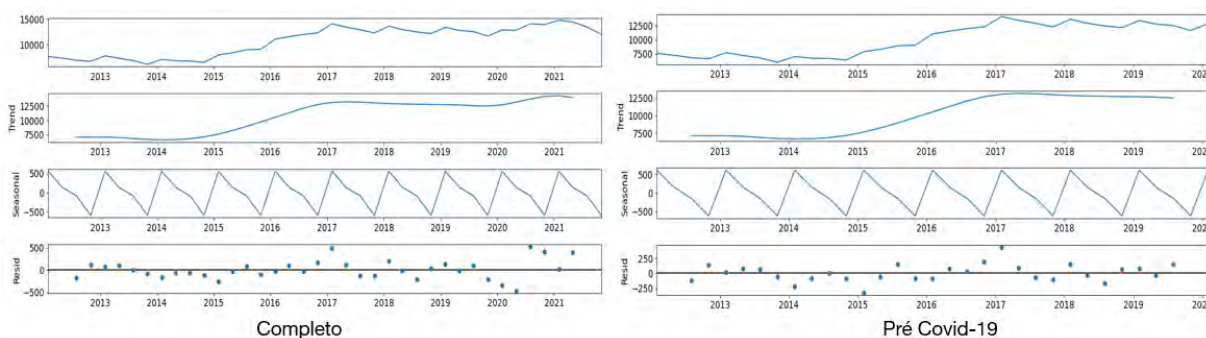


Figura 4. Decomposição da série temporal

4.2.3 Diferenciação

Como a série temporal possui tendência e sazonalidade, ela não é estacionária. Portanto, é necessário o uso de técnicas para torná-la estacionária.

A técnica escolhida para aplicar na série temporal foi a diferenciação e para avaliar a estacionariedade aplicou-se o teste Dickey-Fuller Aumentado (ADF), em que é necessário atingir um valor-p menor que 0.05 para que a série temporal seja considerada estacionária.

Observa-se que com uma diferenciação o conjunto de dados Completo atingiu o necessário para a estacionariedade da série, tendo seu valor-p na Tabela 4 (B) inferior a 0.05. O mesmo processo de diferenciação foi feito para a base de dados Pré Covid-19.

Tabela 4. Resultado do teste ADF antes das diferenciações

Dickey-Fuller Aumentado	Dickey-Fuller Aumentado
Teste Estatístico: -1.4347	Teste Estatístico: -2.9245
Valor-p: 0.5655	Valor-p: 0.0426
Valores Críticos:	Valores Críticos:
1%: -3.6392	1%: -3.6889
5%: -2.9512	5%: -2.9720
10%: -2.6144	10%: -2.6253
(A) Antes das diferenciações	(B) Após uma diferenciação

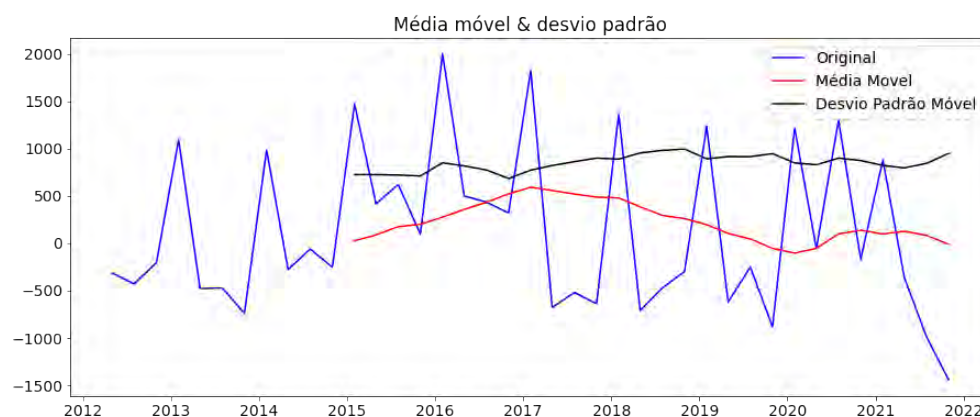


Figura 5. Primeira diferenciação da série temporal

4.3. Modelos de Previsão

Dois cenários foram considerados para a construção dos modelos:

- i) Completo: o conjunto de dados inteiro, com uma previsão dos próximos 4 anos;
- ii) Pré Covid-19: descarta os valores referente ao ano de 2020 e 2021, período de Covid-19, e é feita uma previsão de como seriam os anos de 2020 e 2021 caso a pandemia não tivesse chegado ao Brasil.

4.3.1. Definição do modelo

Para definir qual melhor modelo e parâmetros, foi utilizado o Auto ARIMA, que se adapta a uma variedade de modelos ARIMA e seleciona automaticamente o que teve o melhor desempenho, como mostra nas tabelas 5 e 6.

Tabela 5. Avaliação dos modelos Auto ARIMA - Completo

Modelos	AIC
ARIMA(1,1,0)(1,0,1)[4]	602.901
ARIMA(2,1,0)(1,0,1)[4]	604.844
ARIMA(0,1,1)(1,0,1)[4]	606.392
ARIMA(1,1,0)(2,0,0)[4]	609.829

Tabela 6. Avaliação dos modelos Auto ARIMA - Pré Covid-19

Modelos	AIC
ARIMA(1,1,0)(1,0,1)[4]	480.922
ARIMA(2,1,0)(1,0,1)[4]	483.749
ARIMA(1,1,0)(3,0,0)[4]	484.618
ARIMA(1,1,0)(2,0,0)[4]	484.103

Após algumas observações, o modelo escolhido por apresentar o menor valor para a estatística AIC foi o ARIMA com componente sazonal, SARIMAX (1, 1, 0) x (1, 0, [1], 4) para os dois cenários, como observado na Figura 6.

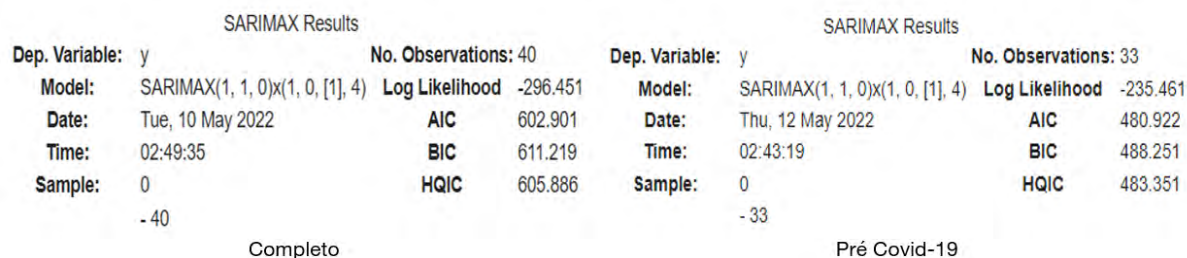


Figura 6. Modelo escolhido no Auto ARIMA

Com o modelo definido e os parâmetros obtidos, foi feita a implementação do modelo de previsão do ARIMA, apresentado no capítulo seguinte de resultados.

5. Resultados de Previsões

A partir dos modelos aplicados a cada cenário, foi possível gerar as previsões, observadas na Figura 7.

O período Pré Covid-19 mostra, na linha azul, que existia uma sazonalidade na série temporal, porém com uma tendência estável. E a previsão, em vermelho, se manteve essa sazonalidade e tendência.

No período Completo observa-se uma sazonalidade e uma tendência estável de 2017 até 2020, porém em 2020 houve a pandemia, apresentando um pico seguido de uma queda em dois anos. A previsão, em vermelho, também manteve a sazonalidade, porém com uma leve tendência de alta.

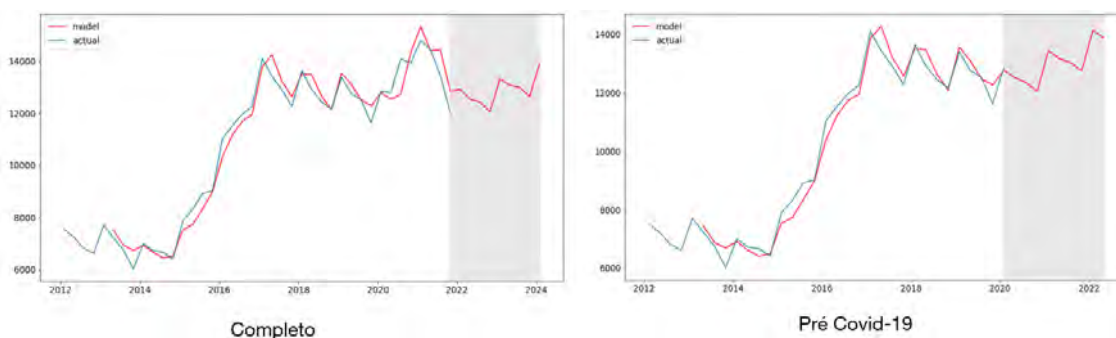


Figura 7. Aplicação do modelo ARIMA para previsão na série temporal

As figuras 8 e 9 mostram como os modelos estão com um ajuste satisfatório, por garantirem que os resíduos do modelo são não correlacionados e normalmente distribuídos com média zero. Isso pode ser observado:

- i) No gráfico de resíduos ao longo do tempo (superior esquerdo) que não mostra uma sazonalidade óbvia.
- ii) O gráfico superior direito, a KDE (linha laranja) segue de perto a linha verde $N(0,1)$, a $N(0,1)$ indica distribuição normal com média 0 e desvio padrão de 1. Portanto, indica que os resíduos são normalmente distribuídos.
- iii) O gráfico de resíduos Q-Q normal (inferior esquerdo), que mostra que os dados estão alinhados a linha vermelha, indicando uma distribuição normal. Se os pontos tivessem se desviado drasticamente da linha, teria que ajustar o modelo.
- iv) E o gráfico de autocorrelação (inferior direito) confirma que os resíduos da série temporal têm baixa correlação com as versões defasadas de si mesmo.

Portanto, pode-se seguir com os parâmetros definidos pelo Auto ARIMA para prever valores futuros.

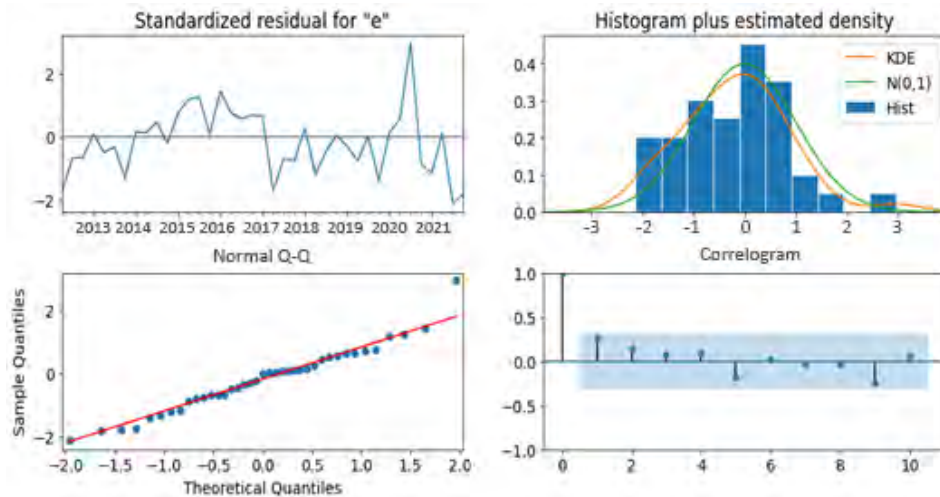


Figura 8. Diagnósticos do modelo SARIMAX – Completo

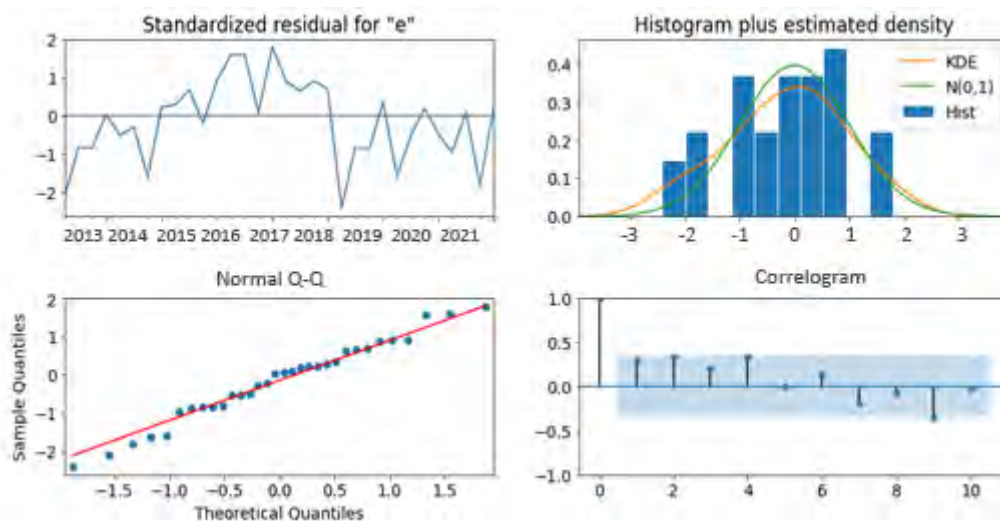


Figura 9. Diagnósticos do modelo SARIMAX – Pré Covid-19

Na figura 10, apresenta-se uma comparação entre o que de fato aconteceu entre 2020 e 2021 (linha laranja) e o que o modelo para a previsão do que poderia acontecer se não tivéssemos a pandemia do Covid-19 (linha verde).

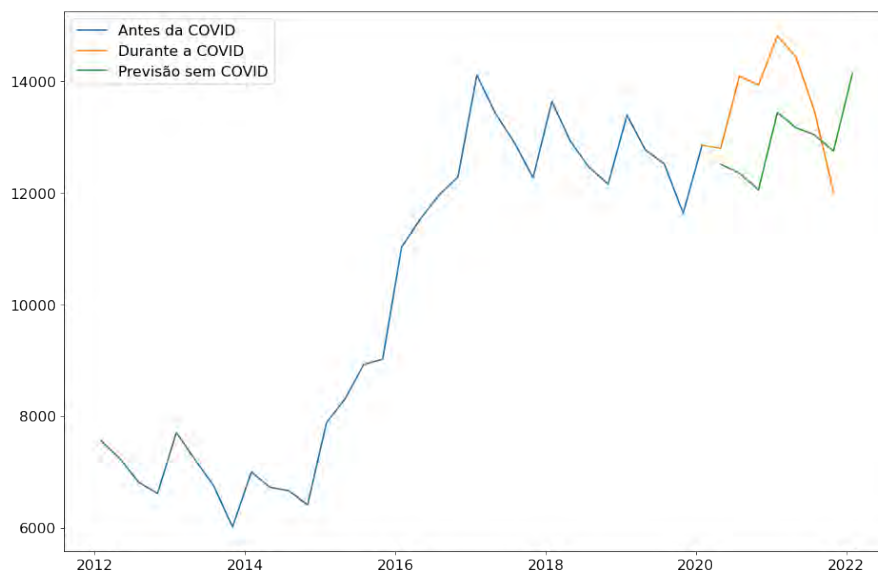


Figura 10. Comparação da previsão Pré Covid com a série temporal real

6. Conclusões

Mintzberg, Ahlstrand e Lampel (2000) diz que o objetivo de modelar previsões de dados não é acertar exatamente os valores, mas ter uma noção de como será o comportamento das observações no futuro, dado que não existe precisão em previsões a longo prazo.

Observando o gráfico da Figura 10 é possível constatar a diferença média entre as duas curvas, que é de 6,99%, podendo ser considerada como o impacto da pandemia na série temporal de desemprego no país. A comparação da previsão do período Pré Covid-19 com a série temporal real, também na Figura 10, nota-se que houve uma alta no desemprego já esperada, mas com a nova realidade, houve uma antecipação de 1 a 2 anos do pico do desemprego, que foi atingido de forma rápida pela crise econômica e sanitária gerada pela pandemia.

Diversos fatores influenciam as crescentes taxas de desemprego observadas. As primeiras demissões que ocorreram foram entre aqueles que vivem do trabalho precário, trabalhadores que são privados das condições básicas de trabalho e proteção social, como terceirizados, balconistas, garçons, diaristas etc. (COSTA, 2020).

O número de trabalhadores informais no Brasil sempre foi alto, com 36,8 milhões de pessoas, cerca de 39,9% da população brasileira, mas a Agência Brasil relata que em 2020 houve um crescimento na informalidade dos empregos, e de diferentes formas de trabalhos informais surgiram, que vão desde motoristas de aplicativos (transporte ou entrega) a trabalhadores sem carteira e empreendedores sem CNPJ. O crescente número foi observado devido a ampliação da pobreza e desigualdade.

Em contrapartida houve crescente número de pessoas trabalhando remotamente, entre maio e julho de 2020, nos diversos setores trabalhistas do mercado nacional que conseguiram se manter em atividade, se tornando uma alternativa a todos os

desdobramentos causados as exigências das restrições exigidas. Porém grande parte dessas pessoas são profissionais de ciências, intelectuais ou pessoas com cargos mais altos, como gerentes e diretores, que somam cerca de 22,8% dos que trabalham remotamente (GOÉS, 201).

É possível observar na Figura 7 (Completo), que a previsão para os próximos 4 anos será de um leve aumento no desemprego, porém sem grandes flutuações. Mudanças não previstas, como uma pandemia, podem alterar consideravelmente a análise, o que implica em um risco de erro inesperado. Além disso é difícil reunir todas as variáveis independentes que impactam na dependente.

Desta forma, não é possível afirmar que o aumento do desemprego está diretamente associado à pandemia exclusivamente, apenas afirmar que essas reflexões são capazes de sustentar que a classe mais baixa foi diretamente afetada pela crise sanitária. E como consequência ao aumento do desemprego temos a elevação da informalização do trabalho, dos terceirizados, subcontratados, flexibilizados, trabalhadores em tempo parcial e do subproletariado.

Observa-se também que há uma queda acentuada nos últimos dados registrados, provavelmente pelo período de vacinação, pois em agosto de 2020 a início de 2021, temos uma retomada das atividades, mesmo com medidas restritivas ainda vigentes, e houve a ampliação na execução de testes e liberação de recursos para pesquisa de vacinas ou medicamentos.

Para futuros trabalhos existe a possibilidade de explorar novos modelos de previsão e fazer uma comparação entre esses modelos, vendo qual tem uma maior acurácia para que seja usado em previsões futuras. Há a possibilidade também de atualizar os dados do trabalho atual com novos dados após a pandemia, e ser realizada uma comparação dos dados previstos com o cenário real, como um mecanismo para o auxílio ao gerenciamento de risco e tomada de decisão.

Referências

- ANTUNES, José L.F.A; CARDOSO, M.R.A. Uso da análise de séries temporais em estudos epidemiológicos. *Aplicações da Epidemiologia. Epidemiologia e Serviços de Saúde* v.24, 2015. Brasília. Disponível em: <<https://www.scielo.br/j/ress/a/zzG7bfRbP7xSmqgWX7FfGZL/abstract/?lang=pt>>. Acesso em: 06 dez. 2021.
- BARBOSA FILHO, Fernando de Holanda. A crise econômica de 2014/2017. *Estudos Avançados*, v.31, n.89, 2017. Disponível em: <<https://www.scielo.br/j/ea/a/B4Nt6NXVr9y4v8tqZLJnDt/?format=pdf&lang=pt>>. Acesso em: 02 dez. 2021.
- BRIDI, M. A; BOHLER, F. B; ZANONI, A. P; BRAUNERT, M.B; BERNARDO, K. A. S; MAIA, F. L; FREIBERGER, Z; BEZERRA, G. U. O trabalho remoto/home-office no contexto da pandemia COVID-19. Instituto de Economia da Unicamp, São Paulo, 2020. Disponível em: <https://www.eco.unicamp.br/remir/images/Artigos_2020/ARTIGO_REMIR.pdf>. Acesso em: 01 mai. 2021.
- BOX, G.; JENKINS, G. *Time series analysis forecasting and control*. Oakland: Holden Day Inc.,1976.
- CADENAS, Erasmo; RIVERA, Wilfrido. Short term wind speed forecasting in La Venta, Oaxaca, México, using artificial neural networks. *Renewable Energy*, 2009.

- Disponível em: <<https://www.sciencedirect.com/journal/renewable-energy>>. Acesso em: 06 dez. 2021.
- CAMELO, H. N. et.al. Modelagem Box-Jenkins aplicada a previsão de velocidade do vento em regiões do nordeste brasileiro para fins de geração eólica. Santa Maria, 2018. Disponível em: <<http://www.repositorio.ufc.br/handle/riufc/40999>>. Acesso em: 02 dez. 2021.
- COMEXSTAT. Brasil: informações gerais. s.d. Disponível em: <<http://comexstat.mdic.gov.br/pt/comex-vis>>. Acesso em: 08 dez. 2021.
- COSTA, Simone da Silva. Pandemia e desemprego no Brasil. *Revista de Administração Pública* [online]. 2020, v. 54, n. 4. Disponível em: <<https://doi.org/10.1590/0034-761220200170>>. Acesso em: 02 dez. 2021.
- DALIAKOPOULOS, Ioannis N.; COULIBALY, Paulin; TSANIS, Ioannis K. Groundwater level forecasting using artificial neural networks. *Journal of hydrology*, v. 309, n. 1-4, p. 229-240, 2005. Disponível em: <<https://www.sciencedirect.com/science/article/abs/pii/S0022169404005840>>. Acesso em: 06 dez. 2021.
- EHLERS, R.S. Análise de Séries Temporais. Departamento de Estatística, UFPR. Paraná, 2007. Disponível em: <<http://leg.est.ufpr.br/~ehlers/notas>>. Acesso em: 06 dez. 2021.
- EKONOMOU, L. Greek long-term energy consumption prediction using artificial neural networks. *Energy*, 35(2), pp. 512-517, 2010. Acesso em: 06 dez. 2021.
- ERDINÇ, A; M. HAKAN, Esq Satman. Stock Market Forecasting: Artificial Neural Network and Linear Regression Comparison in na Emerging Market, Turkey, 2005. *Journal of Financial Management and Analysis*.
- GHORBANI, Mohammad Ali et al. Sea water level forecasting using genetic programming and comparing the performance with artificial neural networks. *Computers & geosciences*, v. 36, n. 5, p. 620-627, 2010. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0098300410000506>>. Acesso em: 06 dez. 2021.
- HYNDMAN, Rob J. Box-Jenkins modelling. In: *Regional Symposium on Environment and Natural Resources*, Mai. 2001. p. 10-11.
- KROLLNER, Bjoern; VANSTONE, Bruce; FINNIE, Gavin. Financial Time Series Forecasting with Machine Learning Techniques: A Survey. *European Symposium on Artificial Neural Networks - Computational Intelligence and Machine Learning*. Bruges, Belgium, 2010. Disponível em: <https://www.researchgate.net/publication/221165259_Financial_Time_Series_Forecasting_with_Machine_Learning_Techniques_A_Survey>. Acesso em: 06 dez. 2021.
- LATORRE, Maria do R. D. de O; CARDOSO, M.R.A. Análise de séries temporais em epidemiologia: uma introdução sobre os aspectos metodológicos. *Revista Brasileira de Epidemiologia*, 2005. São Paulo. Disponível em: <<https://www.scielo.br/j/rbepid/a/KM9MndgpCGSnjSNDddSydCG/abstract/?lang=pt>>. Acesso em: 12 dez. 2021.
- LODETTI, Lidiane. Aplicação da metodologia Box & Jenkins na previsão das vendas de leite de um laticínio em Guaíra-PR. 2015. Trabalho de Conclusão de Curso (Especialização em Estudos Matemáticos) - Universidade Tecnológica Federal do

- Paraná, Medianeira, 2015. Disponível em: <<http://repositorio.utfpr.edu.br/jspui/handle/1/20480>>. Acesso em: 06 dez. 2021.
- MARGARIDO, Mario A. Transmissão de preços internacionais de suco de laranja para preços ao nível de produtor de laranja no Estado de São Paulo. São Paulo: IEA. 1998. 127p. (Coleção Estudos Agrícolas, 6/98).
- MEHDI, Khashei & Bijari. An artificial neural network (p, d, q) model for timeseries forecasting. Department of Industrial Engineering, Isfahan University of Technology, Isfahan, Iran. 2010. Disponível em: <<https://www.sciencedirect.com/science/article/abs/pii/S0957417409004850>>. Acesso em: 06 dez. 2021.
- MINTZBERG, H.; AHLSTRAND, B.; LAMPEL, J. Safári de Estratégia. [S.l.]: Bookman, 2000.
- MORETTIN, P. A. & TOLOI, C. M.C. Análise de Séries Temporais. 3ª Ed. – São Paulo – Blucher, 2018.
- MORETTIN, P. A. & TOLOI, C. M.C. Modelos para Previsão de Séries Temporais, volume I, 1981. Rio de Janeiro IMPA – CNPq.
- NASSERI, M; ASGHARI, K; ABEDINI, M.J. Iran, 2008. Optimized scenario for rainfall forecasting using genetic algorithm coupled with artificial neural network. Disponível em: <<https://www.sciencedirect.com/science/article/abs/pii/S0957417407003557>>. Acesso em: 06 dez. 2021.
- NCSS Statistical Software. NCSS Documentation: The Box-Jenkins Method, Chapter 470, c2021. Disponível em: <https://www.ncss.com/wp-content/themes/ncss/pdf/Procedures/NCSS/The_Box-Jenkins_Method.pdf>. Acesso em: 06 dez. 2021.
- OLIVEIRA, G.L; RIBEIRO, A.P. Relações de trabalho e a saúde do trabalhador durante e após a pandemia de COVID-19. Cad. Saúde Pública, v.37, n.3. Rio de Janeiro. 2021. Disponível em: <<https://www.scielosp.org/article/csp/2021.v37n3/e00018321/>>. Acesso em: 02 mai. 2021.
- REIS, Marcelo Menezes Reis. Estatística para Administradores I: Notas de Aula – Capítulo 4: Análise de Séries Temporais. s.d. Universidade Federal de Santa Catarina – UFSC. Disponível em: <<https://www.inf.ufsc.br/~marcelo.menezes.reis/Cap4.pdf>>. Acesso em: 12 dez. 2021.
- SERVIÇO BRASILEIRO DE APOIO ÀS MICRO E PEQUENAS - SEBRAE. Pequenos negócios em números. Disponível em: <<https://m.sebrae.com.br/sites/PortalSebrae/ufs/sp/sebraeaz/pequenos-negocios-em-numeros,12e8794363447510VgnVCM1000004c00210aRCRD>>. Acesso em: 06 dez. 2021b.
- SERVIÇO BRASILEIRO DE APOIO ÀS MICRO E PEQUENAS - SEBRAE. Sebrae e Caixa vão ampliar o acesso de pequenos negócios a crédito. Disponível em: <<https://www.sebrae.com.br/sites/PortalSebrae/artigos/sebrae-e-caixa-va-o-ampliar-o-acesso-de-pequenos-negocios-a-credito,9c10d1e079a71710VgnVCM1000004c00210aRCRD>>. Acesso em: 06 dez. 2021a.
- TÁVORA, Marcelus Santana. Metodologia para Agrupamento de Dados - Uma versão da Busca Tabu para Sistema R. 2015. viii, 32 f. Trabalho de Conclusão de Curso (Bacharelado em Estatística) - Universidade de Brasília, Brasília, 2015. Disponível

em:<https://www.bdm.unb.br/bitstream/10483/14108/1/2015_MarcelusSantanaTavora.pdf>. Acesso em: 08 mai. 2021.

ZHANG, Guoqiang; PATUWO, B. Eddy; HU, Michael Y. Forecasting with artificial neural networks: The state of the art. *International journal of forecasting*, v. 14, n. 1, p. 35-62, 1998.