

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
FACULDADE DE CIÊNCIAS ECONÔMICAS  
DEPARTAMENTO DE ECONOMIA E RELAÇÕES INTERNACIONAIS**

**RAIANE PADILHA SILVEIRA**

**TAXA DE DESEMPREGO NO BRASIL: VARIÁVEIS DETERMINANTES PARA  
PREVER O PATAMAR**

**Porto Alegre**

**2022**

**RAIANE PADILHA SILVEIRA**

**TAXA DE DESEMPREGO NO BRASIL: VARIÁVEIS DETERMINANTES PARA  
PREVER O PATAMAR**

Trabalho de conclusão submetido ao Curso de Graduação em Ciências Econômicas da Faculdade de Ciências Econômicas da UFRGS, como requisito parcial para obtenção do título Bacharel em Economia.

Orientador: Prof. Dr. Mauricio Andrade Weiss

**Porto Alegre**

**2022**

**RAIANE PADILHA SILVEIRA**

**TAXA DE DESEMPREGO NO BRASIL: VARIÁVEIS DETERMINANTES PARA  
PREVER O PATAMAR**

Trabalho de conclusão submetido ao Curso de Graduação em Ciências Econômicas da Faculdade de Ciências Econômicas da UFRGS, como requisito parcial para obtenção do título Bacharel em Economia.

Aprovada em: Porto Alegre, 06 de outubro de 2022.

BANCA EXAMINADORA:

---

Prof. Dr. Mauricio Andrade Weiss – Orientador

UFRGS

---

Prof. Dr. Hudson da Silva Torrent

UFRGS

---

Prof. Dr. Rober Iturriet Avila

UFRGS

### CIP - Catalogação na Publicação

Silveira, Raiane Padilha  
Taxa de desemprego no Brasil: variáveis  
determinantes para prever o patamar / Raiane Padilha  
Silveira. -- 2022.  
62 f.  
Orientador: Maurício Andrade Weiss.

Trabalho de conclusão de curso (Graduação) --  
Universidade Federal do Rio Grande do Sul, Faculdade  
de Ciências Econômicas, Curso de Ciências Econômicas,  
Porto Alegre, BR-RS, 2022.

1. Taxa de Desemprego. 2. Séries Temporais. 3.  
Aprendizado de Máquina. 4. Macroeconomia. 5. Seleção  
de Variáveis. I. Weiss, Maurício Andrade, orient. II.  
Título.

## **AGRADECIMENTOS**

Este é meu segundo trabalho de conclusão de curso na UFRGS, e posso dizer que a Raiane que escreve esse segundo agradecimento hoje é muito diferente da Raiane que escreveu o outro há 5 anos atrás. O primeiro trabalho foi cheio de medo, receio e expectativa, o segundo foi calmo, tranquilo e seguro.

De qualquer forma, eu não existiria sem as pessoas que são meu alicerce, apoio e alavancagem emocional, e neste espaço deixo mais uma vez meus agradecimentos a minha mãe, meu pai e irmã que mesmo que não tenham acompanhado tão de perto dessa vez, ou até mesmo tenham ficado libertos da responsabilidade financeira, deram sempre todo o apoio e confiança necessários para que eu seguisse em frente. Agradeço principalmente pelas infinitas demonstrações de orgulho.

Ao meu agora noivo, que no outro trabalho de conclusão ainda era namorado, agradecimentos pela companhia, por ser meu motorista e por também sempre demonstrar tanto orgulho de mim.

Às minhas melhores amigas: Letícia, Fernanda, Taiani, Iana, Ana Júlia e Renata, por serem fuga dessa vida tão louca e corrida que eu mantenho durante tantos anos. Por sempre estarem por perto e tornarem a minha vida tão mais fácil.

E, por último, mas não menos importante, ao meu orientador, que mais do que apoiar na construção desse trabalho, luta por educação gratuita e de qualidade todos os dias, travando a cada dia uma nova batalha de resistência. Se não for pedir demais, continue Mauricio.

Diferente da outra graduação essa aqui passou voando, mais uma vez agradeço a oportunidade de ter usufruído dessa instituição de ensino que mesmo sendo deteriorada pelo governo a cada ano segue figurando entre as melhores universidades do país – o que faz com que eu tenha ainda mais orgulho desse segundo diploma.

“Investir em conhecimento rende sempre os melhores juros.” - Benjamin Franklin

## RESUMO

Neste trabalho foram utilizadas técnicas clássicas de séries temporais e algoritmos de aprendizado de máquina para prever o patamar da taxa de desemprego no Brasil. Para isso, foram utilizados dados macroeconômicos de 2000 a 2021, coletados de fontes como IBGE, IPEA, Banco Central, entre outras. As variáveis foram selecionadas através da revisão bibliográfica presente neste trabalho e para redução de dimensionalidade foram aplicadas as técnicas de componentes principais e correlação de Spearman. Os resultados mostram que os modelos de aprendizado de máquina apresentam menor erro absoluto e menor raiz do erro quadrático médio quando comparado aos modelos clássicos de séries temporais. As variáveis finais constantes nos modelos de aprendizagem de máquina como relevantes para previsão são: Dívida Pública, Indicador de Custo de Crédito, Índice de Gini, Número de Famílias Atendidas pelo Bolsa Família, Salário Mínimo, Taxa de Pobreza e Taxa Selic. Este estudo é uma abordagem da previsão do patamar da taxa de desemprego no Brasil em relação aos fatores macroeconômicos.

**Palavras-chave:** Taxa de Desemprego. Séries Temporais. Aprendizado de Máquina. Macroeconomia. Seleção de Variáveis.

## **ABSTRACT**

In this work, classical time series techniques and machine learning algorithms were used to predict the level of the unemployment rate in Brazil. For this, macroeconomic data from 2000 to 2021 were used, collected from sources such as IBGE, IPEA, Central Bank, among others. The variables were selected through the literature review present in this work and for dimensionality reduction, the principal components and Spearman correlation techniques were applied. The results show that machine learning models have lower MAE and lower RSME when compared to classical time series models. The final variables in the machine learning models as relevant for prediction are: Public Debt, Credit Cost Indicator, Gini Index, Number of Families Served by Government Aid, Minimum Wage, Poverty Rate and Selic Rate. This study is an approach to forecasting the level of the unemployment rate in Brazil in relation to macroeconomic factors.

**Keywords:** Unemployment Rate. Time Series. Machine Learning. Macroeconomics. Selection of Variables.



## LISTA DE GRÁFICOS

|  |    |
|--|----|
| Gráfico 1 – Percentual de variância explicada: Componentes Principais..... | 44 |
| Gráfico 2 – Série histórica da Taxa de Desemprego anual média.....         | 48 |
| Gráfico 3 – ACF E PACF da série.....                                       | 49 |
| Gráfico 4 – Ajuste do modelo AR.....                                       | 50 |
| Gráfico 5 – % de contribuição da variável na redução do erro<br>.....      | 51 |
| Gráfico 6 – Ajuste do modelo Random Forest.....                            | 52 |
| Gráfico 7 – Ajuste do modelo SVM.....                                      | 53 |
| Gráfico 8 – Relacionamento entre as variáveis.....                         | 54 |
| Gráfico 9 – Ajuste do modelo LASSO.....                                    | 56 |

## LISTA DE TABELAS

|   |    |
|---|----|
| Tabela 1 – Variáveis macroeconômicas mapeadas.....                                    | 36 |
| Tabela 2 – Relevância das variáveis nas componentes.....                              | 45 |
| Tabela 3 – Correlação de Spearman das variáveis.....                                  | 46 |
| Tabela 4 – Correlação de Spearman das variáveis.....                                  | 46 |
| Tabela 5 – Correlação de Spearman individual das variáveis vs Taxa de Desemprego..... | 47 |
| Tabela 6 – Resultados do modelo AR.....   | 49 |
| Tabela 7 – Resultados do modelo Random Forest.....                                    | 52 |
| Tabela 8 – Resultados do modelo SVM.....  | 53 |
| Tabela 9 – Resultados do modelo LASSO.....  | 56 |
| Tabela 10 – Comparativos dos ajustes dos modelos.....                                 | 57 |

## SUMÁRIO

|          |  |           |
|----------|--|-----------|
| <b>1</b> | <b>INTRODUÇÃO</b> .....  | <b>15</b> |
| <b>2</b> | <b>PERSPECTIVAS TEÓRICAS SOBRE DETERMINANTES DE DESEMPREGO</b> ... | <b>17</b> |
| <b>3</b> | <b>ANÁLISE ESTATÍSTICA APLICADA AO TRABALHO</b> .....              | <b>26</b> |
|          | 3.1 SELEÇÃO DE VARIÁVEIS.....                                      | 26        |
|          | <b>3.1.1</b> <b>Análise de Componentes Principais</b> .....        | <b>27</b> |
|          | <b>3.1.2</b> <b>Análise de Correlação de Spearman</b> .....        | <b>27</b> |
|          | 3.2 MODELOS CLÁSSICOS DE SÉRIES TEMPORAIS .....                    | 28        |
|          | <b>3.2.1</b> <b>Auto-regressivos</b> .....                         | <b>29</b> |
|          | <b>3.2.2</b> <b>Média móvel</b> .....                              | <b>29</b> |
|          | <b>3.2.3</b> <b>Auto-regressivos de média móvel</b> .....          | <b>30</b> |
|          | <b>3.2.4</b> <b>ARIMA</b> .....                                    | <b>31</b> |
|          | 3.3 MODELOS DE MACHINE LEARNING.....                               | 31        |
|          | <b>3.3.1</b> <b>Random Forest</b> .....                            | <b>32</b> |
|          | <b>3.3.2</b> <b>SVM – (Support Vector Machine)</b> .....           | <b>33</b> |
|          | <b>3.3.3</b> <b>Lasso Regression</b> .....                         | <b>34</b> |
|          | 3.4 BASE DE DADOS .....  | 35        |
| <b>4</b> | <b>RESULTADOS</b> .....  | <b>43</b> |
|          | 4.1 SELEÇÃO DE VARIÁVEIS.....                                      | 43        |
|          | 4.2 PREDIÇÃO.....  | 47        |
|          | <b>4.2.1</b> <b>Modelos clássicos de série temporal</b> .....      | <b>47</b> |
|          | <b>4.2.2</b> <b>Modelos de machine learning</b> .....              | <b>51</b> |
|          | 4.3 COMPARATIVO.....   | 56        |
| <b>5</b> | <b>CONCLUSÃO</b> .....   | <b>59</b> |
|          | <b>REFERÊNCIAS</b> .....   | <b>61</b> |

## 1 INTRODUÇÃO

A situação econômica do Brasil tem sido ao longo de todo o século alvo de estudos. Para esse ano de 2022, com a inflação alta e juros subindo a projeção é ruim em termos de desempenho econômico. O cenário externo é cada vez mais desafiador e no contexto de ser um ano eleitoral ainda prejudica os aspectos de expectativa econômica.

Há diversos componentes que são foco destes estudos, como por exemplo a queda do Produto Interno Bruto (PIB), a alta da taxa de juros, a piora das contas do governo, a alta da inflação e a desvalorização da moeda. Qualquer desses estudos involuntariamente apresentam consequências sobre a taxa de desemprego. O desemprego é uma das principais variáveis econômicas e sociais, uma vez que reflete diretamente em todos os pontos citados acima, em âmbitos de políticas públicas e redistribuição de renda.

Por muito tempo houve um embate sobre o desemprego, onde os economistas clássicos classificavam o desemprego como involuntário e transitório e a escola keynesiana apontava que o desemprego poderia ser persistente e não apenas voluntário, sendo essa segunda escola a que teve mais força considerando o período 1930 até 1970, a despeito de ser predominantemente utilizada a interpretação da síntese neoclássica.

Na literatura recente existem diversos estudos quanto as causas que determinam o nível de emprego. Fraga e Dias (2007) atribuem a escolaridade, Mattos (2001) atribui a componentes estruturais, demografia e a relevância do setor público, Antigo e Araújo (2016) atribuem a qualificação de mão de obra, Becker (2010) aponta que a taxa de desemprego é um componente que pode se explicar ao longo do tempo, entre outros autores e perspectivas.

Há amplos estudos a respeito de possíveis aspectos macro e microeconômicos que impactam essa taxa. Entretanto, não foi levado a exaustão os testes a respeito de mapeamento de variáveis do ponto de vista de significância estatística. A chave deste trabalho é verificar as variáveis macroeconômicas que são determinantes para previsão da taxa de desemprego brasileira, bem como o seu patamar, para atingir esse objetivo serão testadas técnicas estatísticas como componentes principais, correlação de

Spearman, modelos clássicos de séries temporais e modelos de *Machine Learning*. Tem-se como hipótese inicial que variáveis macroeconômicas são determinantes, espera-se que sejam significativamente relevantes as seguintes variáveis: PIB, inflação, dívida pública, taxa selic, salário mínimo e valor do investimento.

Segundo dados do Instituto Brasileiro de Geografia e Estatística (IBGE), do terceiro trimestre de 2021, 13,5 milhões de pessoas estavam desempregadas no Brasil, o que representa uma taxa de desemprego de 12,6%. Segundo levantamento da consultoria Austin Rating de agosto de 2021, o Brasil é o quarto país com maior taxa de desemprego do mundo. Para fins de comparação, os Estados Unidos, país desenvolvido, tem taxa de desemprego inferior a 4%.

Do ponto de vista acadêmico um estudo profundo da atividade econômica de um país requer analisar a dinâmica dos seus mais diversos âmbitos econômicos e sociais. Alguns destes nos fornecem uma posição do estado da economia de um país em termos de distribuição de renda, desenvolvimento e estabilidade econômica. O patamar de taxa de desemprego é uma das principais variáveis econômicas estudadas para compreender e diagnosticar os principais fenômenos econômicos e sociais. Esses dados justificam a necessidade de identificação das variáveis que impactam o desemprego, bem como a capacidade de previsão de um patamar para os próximos anos, e este é o escopo deste trabalho.

Dessa forma, o trabalho a seguir está estruturado no seguinte formato: primeiro está disposta uma revisão bibliográfica, passando pelos clássicos da literatura econômica e atendendo pontos mais atuais, mapeando possíveis variáveis que serão consideradas no decorrer do estudo. Depois é feita uma revisão da metodologia que foi aplicada ao trabalho, dispendo brevemente sobre as técnicas. Em seguida, são descritos os resultados obtidos e considerações finais.

## 2 PERSPECTIVAS TEÓRICAS SOBRE DETERMINANTES DE DESEMPREGO

Existem diversos estudos sobre desemprego, e diferentes perspectivas quanto as causas que determinam o nível de emprego. Para a perspectiva convencional os fatores são principalmente relacionados a custos e rigidez, já para pós-keynesiana são fatores relacionados a demanda. Há ainda autores, com estudos mais específicos, que atribuem a fatores exclusivos, como Marston (1985) com a teoria compensatória. O objetivo dessa revisão é fazer uma breve cobertura sobre os conceitos teóricos que norteiam estudos de nível de emprego e trazer alguns autores com perspectivas mais recentes que irão embasar as variáveis determinantes que serão consideradas no decorrer do trabalho, a ideia é construir o panorama teórico que será trabalhado metodologicamente no decorrer nas próximas seções.

Para iniciar a discussão é necessário apresentar a teoria neoclássica, capitaneada pela maioria dos economistas anteriores a publicação de Keynes da Teoria Geral, em que o desemprego aparece como uma situação anormal e de curto prazo, em que somente fatores exógenos podem explicar o mau funcionamento do mercado de trabalho. Isto porque, em geral nessa teoria, a economia opera em pleno emprego de trabalho e fatores de produção. É a situação em que as curvas de oferta e demanda de mão de obra ficam equilibradas, em que qualquer indivíduo disposto a trabalhar encontrará um emprego.

Com isso, a teoria neoclássica atribui como fator determinante do desemprego a existência de um comportamento irracional por parte dos trabalhadores, ou pela existência de alguma rigidez na função oferta de trabalho, ou ainda, por um problema de informações imperfeitas.

Ainda no âmbito neoclássico, há teorias um pouco mais complexas, que admitem a possibilidade do desemprego involuntário no longo prazo, permanecendo anormal e com causas exógenas. No tocante disso, põe-se a possibilidade de existência de desemprego natural e/ou estrutural, que no longo prazo, pode caracterizar uma situação de equilíbrio.

Portanto, é possível perceber que para os neoclássicos o desemprego é voluntário e causado por fatores exógenos. É importante mencionar também que o salário real é

determinante para este grupo, quanto a determinação da existência de desemprego, sendo responsável por determinar o equilíbrio. Trazendo a questão do capital em plena racionalidade, somente os trabalhadores exigindo salários anormais são capazes de produzir desemprego, sendo assim um problema de oferta de mão de obra, um problema microeconômico.

Em resumo, para os economistas neoclássicos se parte do pressuposto de que dados a tecnologia e o objetivo de maximização de lucro das firmas, o nível de emprego, o salário real e a distribuição de renda são determinados simultaneamente no mercado de trabalho. Um dado nível de emprego de equilíbrio corresponde à quantidade de bens plenamente empregados, que são vendidos integralmente a um determinado nível de preços, dependendo da quantidade de moeda e de sua velocidade de circulação. Se houver distorções que criem rigidez salarial, o nível de emprego determinado pelo mercado de trabalho não corresponderá ao pleno emprego. Logo, nessa teoria, para resolver o desemprego basta eliminar as distorções exógenas.

Feito esse contexto da teoria neoclássica é o momento de trazer para discussão Keynes e a Teoria Geral, que em síntese traz duas críticas a teoria citada anteriormente. A primeira diz respeito à capacidade de o mercado de trabalho determinar, unilateralmente, o nível de salário real, para Keynes o relevante é o salário nominal. E, a segunda, diz respeito à suposição neoclássica de que a demanda agregada é constante, no curto prazo, e independente do mercado de trabalho, trazendo o cerne de Keynes de que o pleno emprego não é uma situação permanente.

Keynes trouxe a perspectiva de que o pleno emprego é gerado pelo comportamento da demanda efetiva e dependente da existência de incertezas e instabilidades da economia monetária, colocando na moeda um papel determinante, o que não era citado anteriormente.

Para adotar a importância da economia monetária como fator determinante, a análise de Keynes aponta para existência de capacidade ociosa e desemprego involuntário como uma situação de equilíbrio, como uma situação normal. Isto é, o desemprego pode existir porque os trabalhadores não conseguem empregos, mesmo se sujeitando a menores salários. Neste ponto, surge a importância das ações governamentais, que norteiam a economia a o pleno emprego.

Dessa forma, no curto prazo, nem a oferta agregada e nem o mercado de trabalho são fundamentais. O emprego é determinado pela demanda efetiva e os trabalhadores não controlam sua própria oferta. É importante citar, que existe o desemprego voluntário, mas este não é relevante, pois a teoria é focada no desemprego involuntário. Sendo assim, o principal componente da demanda efetiva é o investimento, no tocante da taxa de investimentos e como as políticas econômicas podem aumentar essa taxa.

Ainda em relação ao investimento, fator importante na teoria keynesiana, há destaque para a situação de que, no mundo real, os indivíduos que poupam não são necessariamente os mesmos que tomam as decisões de investir. Segundo Keynes, o montante de investimento depende do incentivo para investir, baseado em expectativas futuras, dado o ambiente de incerteza. Conseqüentemente, em alguns cenários, o investimento agregado reduzirá, e dada a propensão a consumir da sociedade cairá a demanda agregada e o nível de emprego. Como indica Furtado (1997 apud Keynes; cap 3)

Não é, portanto, a desutilidade marginal do trabalho, expressa em termos de salários reais, que determina o volume de emprego (...). A propensão a consumir e o nível do novo investimento é que determinam, conjuntamente, o nível de emprego, e é este que, certamente, determina o nível de salários reais - não o inverso. (Furtado, 1997 apud Keynes; 1936, p.46, cap 3)

Em resumo, Keynes sinaliza a relação complexa entre variáveis macroeconômicas e o nível de emprego. Coloca a taxa de desemprego como consequência do nível da demanda efetiva. Fica implícito que o combate ao desemprego é através de política macroeconômica com o objetivo de controlar a demanda agregada.

Uma terceira via corresponde as perspectivas pós-keynesianas, conhecidas também como ressurgimento das ideias neoclássicas na década de 80, se refere aos clássicos buscando explicar a instabilidade financeira como uma característica endógena de economias funcionando de forma liberal.

Foram desenvolvidos vários modelos, cujo principal objetivo era explicar as razões das altas taxas de desemprego em alguns países da Organização para a Cooperação e Desenvolvimento Econômico (OCDE). Basicamente, podendo ser divididos entre rigidez do salário real e conflito entre “*insiders*” e “*outsiders*”.

Quanto a rigidez de salário real, o contexto é o período pós guerra, com o crescimento da filiação sindical. Esse modelo tem como premissas: variações dos



salários nominais não produzem impacto sobre o nível de preços, sendo esse determinado fora do mercado; o sindicato detém o monopólio da oferta de trabalho; o processo de acordo salarial é conduzido pelo sindicato, com o objetivo de maximizar o grau de satisfação.

Esse modelo demonstra que a negociação conjunta pode produzir desemprego. O que conduz ao questionamento de até que ponto os sindicatos podem compensar eventuais quedas do salário real causadas por elevações no nível de preços, causando, assim, uma rigidez na taxa de desemprego? A resposta depende do poder de negociação de revisão salarial ou da introdução de mecanismos eficazes de indexação salarial.

Quanto ao conflito entre “*insiders*” e “*outsiders*” as principais premissas segundo Furtado (1997), são: a força de trabalho é segmentada em dois grupos de trabalhadores: os *insiders*, que possuem um poder de mercado, e os *outsiders*, que não possuem a capacidade de negociar salários; neste tocante, ainda há o poder de negociação salarial dos *insiders* como determinante de custos de rotatividade de mão de obra, criando despesas para as empresas; o poder de negociação causa também uma percepção negativa quando a rotatividade e produtividade; e, neste modelo, as firmas têm o objetivo de tornar máximo seu lucro, sujeitas a uma tecnologia que gera rendimentos marginais decrescentes e a preços determinados em um mercado de competição perfeita.

Neste modelo, segundo Furtado (1997), existe um nível ótimo de salário real, que torna máxima a eficiência da firma em produtividade. Esse nível de salário real está associado a um montante de emprego que, só por mera coincidência, será igual ao número de trabalhadores dispostos a trabalhar àquele dado salário.

Há ainda a teoria Schumpeteriana, que destaca as mudanças estruturais da oferta agregada como motivador potencial do desemprego, uma vez que impactam na demanda por mão de obra.

Para Schumpeter, o que altera a oferta agregada são as inovações tecnológicas que impactam em desenvolvimento econômico. Sendo assim, as mudanças estruturais na oferta agregada podem gerar um desalinhamento entre oferta e demanda de mão de obra, em caráter involuntário.

Todos os textos posteriores se baseiam de alguma forma nessas “teorias mães” a respeito dos determinantes do desemprego, algumas mapeando pontos mais

específicos, como é o objetivo deste trabalho e outras testando as questões anteriores citadas, alguns autores enxergam fatores microeconômicos como pré determinantes do desemprego, enquanto outros apontam para comportamentos macroeconômicos.

Barros, Camargo e Mendonça (1997), em A estrutura do desemprego no Brasil, indicavam que a causa da taxa de desemprego não era a quantidade de postos gerados e sim a qualidade desses postos. É importante contextualizar, que no período de desenvolvimento deste trabalho, a taxa de desemprego do Brasil era relativamente baixa, e os autores queriam evidenciar que mesmo assim isso já desencadeava em outros problemas como grau de pobreza e nível de desigualdade. Desse trabalho foram apontados como fatores determinantes do nível de desemprego: sexo, nível educacional, idade, posição no domicílio, setor de atividade e posição na ocupação.

No texto de Costa e Teixeira (2008), relacionado ao período de 1995-2002, um período em que o desemprego já pairava como um problema generalizado em vários países, incluindo o Brasil, há o objetivo de evidenciar algumas características que possam explicar a diferença na probabilidade de desemprego por gênero.

O período de recorte escolhido pelos autores é interessante, dada a lenta, mas real evolução da mudança de perfil e dinâmica do mercado de trabalho brasileiro. Os autores destacam que a década de 1990 é um marco em relação ao desemprego como problema social, muito ligado a gestão política de Fernando Henrique Cardoso (FHC) da segunda metade da década e os rearranjos macroeconômicos, o índice de desemprego nesse período atingiu um patamar médio de 8,7%. Christo (2013) corrobora com essa informação trazida pelos autores, em seu trabalho em que faz o estudo dos índices de desemprego dos governos Lula e FHC a autora também atribui às políticas econômicas o patamar de desemprego.

Deste trabalho, foram apontados como fator determinante além do sexo, o grau de escolaridade, idade e região. Os autores atrelaram esses pontos às políticas macroeconômicas e oportunidades que a população recebeu no período.

Há ainda autores que trabalham em nichos mais específicos de questões de desemprego, como Cunha, Araújo e Lima (2011) que também discutem sobre os determinantes do desemprego no Brasil, mas com foco em jovens e regiões metropolitanas.

O interesse dos autores neste público como foco é porque, o desemprego no Brasil atingia principalmente pessoas entre 15 e 29 anos e moradores de regiões metropolitanas, com base nos dados do Instituto de Pesquisa Econômica Aplicada (IPEA) de 2008. Neste período a taxa de desemprego dos jovens era três vezes superior ao dos adultos.

Nesse trabalho foram testadas as hipóteses de a razão fundamental do desemprego ser a incerteza a respeito da qualidade do trabalhador - diretamente relacionada a baixa experiência. E, a falta de interesse dos jovens em trabalhar por estarem dedicados aos estudos. Foram obtidas como conclusão de fatores determinantes do desemprego: sexo, a raça, o grau de escolaridade e a renda familiar.

Outro autor que trabalhou com foco no público jovem foi Cadiñanos (2019), em seu texto que aborda o período de 1991 a 2014. A seleção do público também é explicada diante da vulnerabilidade à perda de emprego desse público. As hipóteses iniciais postas são de que o crescimento econômico influencia inversamente o desemprego juvenil e a suposição de que os gastos em educação e nível educacional influenciam também de maneira inversa o desemprego juvenil. Diferente de Cunha, Araújo e Lima, Cadiñanos (2011) nos apresenta fatores macroeconômicos na discussão.

Antigo e Araújo (2016) tem como objetivo analisar o desemprego no Brasil através da perspectiva da qualificação da mão de obra entre 2002 e 2011. Trazendo como principais determinantes as variáveis de sexo, condição do domicílio, idade e escolaridade. Quanto a qualificação do trabalhador, os resultados desse estudo mostram que quanto mais escolarizado, menor o efeito adverso das características pessoais de um indivíduo, e que o grupo dos semiquualificados (aqueles que tem entre 4 e 10 anos de estudo) são os mais sensíveis a tais efeitos no sucesso pela busca do emprego.

Há ainda autores que caracterizam a taxa de desemprego como sendo um componente que pode se explicar ao longo do tempo, Becker (2010) trabalha o desemprego com modelos de séries temporais a fim de obter previsões para suas taxas futuras, basicamente baseado em médias móveis com componente de sazonalidade.

Proni (2015), em um debate contemporâneo das teorias do desemprego destaca fatores macroeconômicos como preditores da taxa de desemprego, por exemplo, aponta que as oscilações na taxa de desemprego acompanham as flutuações da atividade

econômica. Isto é, em períodos de recessão, desemprego se torna uma questão social de maior proporção.

Neste contexto de autores que elucidam fatores macroeconômicos como preditores de desemprego é necessário mencionar o conceito da Curva de Phillips que era a principal teoria adotada até os anos 1970, no período conhecido como choques do petróleo, que culminou em altas taxas de inflação e desemprego. Ela rege a interação entre desemprego e inflação, indicando uma relação inversa. Ou seja, se há aumento de um, há redução de outro.

Essa teoria é considerada válida para o curto prazo, em períodos em que a economia varia entre expansão ou recessão. Em períodos de expansão, existe diminuição do emprego e aumento da inflação, e o contrário em períodos de recessão.

Com o aumento simultâneo do desemprego e dos preços, cenário conhecido como estagflação, essa teoria passou a ser questionada, principalmente para o longo prazo. Há diversos estudos que questionam a linearidade dessa curva, como Correa (2009), uma vez que a linearidade implica que a taxa de desemprego não varia com a agressividade da política de desinflação ou com o estado de aquecimento ou recessão da economia.

É possível perceber que todos os textos mais recentes apontam para causas e determinantes do desemprego no país, o que discorda, portanto, que o desemprego seja natural como diziam os neoclássicos.

A questão do salário mínimo aparece como um tema latente na literatura. Em alguns estudos, a questão aparece como impactante em relação ao mercado de trabalho, pois afeta o custo da mão-de-obra, os lucros das empresas e o nível de preço. Entretanto, é alvo de controvérsias, segundo Lucas (2006), se, por um lado, há os modelos competitivos que apontam uma relação inversa entre o nível de emprego do setor formal e o valor do salário mínimo, há, de outro, os modelos não competitivos, em que as firmas podem determinar o preço do fator trabalho e como resultado o impacto do salário mínimo sobre o nível de emprego pode ser nulo ou até positivo.

Diante desses debates pode-se concluir que é um tema complexo, dividido principalmente em duas vertentes: a primeira vertente é a de inspiração neoclássica, que coloca no mercado de trabalho o peso de determinação dos níveis de salário real e de

emprego na economia, desconsiderando quaisquer outros fatores de impacto na demanda agregada. Impondo que o desemprego surge por rigidez salarial e, esta por sua vez, é provocada pelo comportamento dos seus agentes.

A segunda vertente, e que dá vazão a todos estudos posteriores citados, é a vertente de Keynes, que mostra que o nível de emprego depende principalmente da demanda agregada da economia, que tem direta relação com propensão da sociedade a consumir e o montante de investimentos. Keynes também pontua que as negociações entre trabalhadores e empregadores só tem poder de determinar o nível de salário nominal.

A maioria dos estudos macroeconômicos mais recentes parte das seguintes hipóteses: de fato há uma taxa natural de desemprego e compatível com a inflação estável, em função do desemprego friccional, falhas de informações sobre oportunidades de emprego e problemas de oferta e demanda; os reajustes de salário dependem da inflação esperada e das expectativas; a inflação efetiva é uma função da taxa de reajuste dos salários, descontada a taxa possível de crescimento dos salários reais.

Em trecho retirado de Teoria do Desemprego: um guia de estudo do Instituto de Economia da Unicamp há uma excelente síntese que será posta a seguir:

Empiricamente, as oscilações na taxa de desemprego acompanham as flutuações da atividade econômica: em períodos de depressão econômica o desemprego se torna uma questão social de maior proporção; em períodos de crescimento do nível de atividade o problema tende a diminuir. Mas, a causa do problema pode variar de acordo com o perfil dos desempregados, sendo necessário verificar se a privação de trabalho é generalizada ou atinge só os grupos mais vulneráveis e medir o tempo médio que permanecem desempregados. (Proni, 2015, p. 2)

Com o que foi posto aqui, pode-se perceber que existem diferentes vertentes e diferentes formas de entender o que define desemprego. Atualmente, o desemprego se manifesta de forma particular em cada país, em contexto econômico e social. O que nos leva a questionar se é possível teorizar de forma única que o desemprego de fato não se modifica no tempo e no espaço, se as teorias são validas para qualquer conjuntura.

Como alguns dos autores citados mencionaram, a dinâmica demográfica, o ordenamento institucional e as estratégias mudam o funcionamento do mercado de trabalho. O que não desconstitui as abordagens clássicas, e sim traz mais fatores para avaliação.

Tendo feita essa exposição, é importante reforçar que o intuito dessa revisão não era de levar a teoria do desemprego a exaustão. E sim, apresentar os autores de maior relevância e outros autores que contribuem para literatura com estudos antecessores a este que propõem algumas variáveis preditoras que serão utilizadas no decorrer deste trabalho. Logicamente, haverá dificuldade de incluir todas as variáveis citadas e algumas matrizes teóricas com diferentes enfoques. Neste contexto, este trabalho buscará mapear as variáveis obtidas através da revisão bibliográfica, focando em variáveis macroeconômicas, e verificar seu nível de significância frente a taxa de desemprego do Brasil, sem nenhuma pretensão de obter uma corrente teórica sobre o assunto. O objetivo não é de questionar ou pôr em voga as teorias, apenas apresentar possíveis relações existentes e encontrar determinantes econômicos que sejam relevantes para taxa de desemprego.

A realidade contemporânea brasileira é de elevada taxa de desemprego, e desemprego de longa duração, atingindo 11,2% em janeiro de 2022, segundo o IBGE, fazendo com que esse tema seja parte do dia a dia de milhões de brasileiros. É de suma importância mapear as principais variáveis preditoras e futuramente conseguir fazer previsões acuradas desse indicador para construir políticas eficientes, e soluções que minimizem essa taxa, bem como identificar os principais públicos afetados com o movimento das variáveis macroeconômicas.

### 3 ANÁLISE ESTATÍSTICA APLICADA AO TRABALHO

Neste capítulo estará disposta uma breve revisão teórica dos métodos escolhidos para construção dos resultados. Passando pelas seguintes etapas:

- a) Seleção de variáveis;
- b) Modelos clássicos de séries temporais; e
- c) Modelos de *Machine Learning*.

A escolha das técnicas de seleção de variáveis foi feita baseada em robustez. Os modelos de *Machine Learning* foram escolhidos por apresentarem melhores resultados quando comparados a outras técnicas, embora tenham sido testados uma grande variedade de algoritmos, estarão dispostos aqui os três de melhor performance. Por fim, quanto aos modelos clássicos de séries temporais, estarão descritos os modelos de uso mais comum na academia, conforme descrito por Becker (2010) a análise de séries temporais permite descrever a dependência dos dados em relação ao tempo e prever o futuro, destacando como um dos principais exemplos a própria taxa de desemprego, objeto de estudo deste trabalho.

#### 3.1 SELEÇÃO DE VARIÁVEIS

Quando se fala em utilizar modelos de *machine learning* ao invés de modelos clássicos de séries temporais, destaca-se como uma das grandes diferenças a expectativa que as previsões dependam de outras variáveis, e essas outras variáveis podem chegar a um volume representativo.

Choi et al. (2002) pontuam que um modelo preditivo pode ser altamente comprometido pela baixa qualidade dos dados. Por isso, a seleção das variáveis é um processo imprescindível.

Há muitos outros pontos que tornam a redução de dimensionalidade dos dados essencial, tais como:

- a. quanto maior a gama de variáveis, maior é o trabalho de manutenção e acompanhamento de qualidade do banco de dados;
- b. análises de mensuração de armazenamento; e

- c. facilitar a visualização e entendimento dos dados.

As próximas seções trarão detalhes das técnicas escolhidas para desenvolver este trabalho.

### 3.1.1 Análise de Componentes Principais

A análise de componentes principais (ACP) é uma das técnicas estatísticas mais utilizadas para redução de dimensionalidade em diferentes áreas. Essa técnica foi descrita por Pearson (1901) e descrita computacionalmente por Hotelling (1933, 1936) que usou com o propósito determinado de analisar as estruturas de correlação.

Segundo Hongyu, Junior e Sandanielo (2015) esta técnica transforma linearmente um conjunto inicial de variáveis, num conjunto menor de variáveis, retirando as variáveis que são altamente correlacionadas, deixando somente o conjunto que contém a maior parte da informação do conjunto inicial.

O objetivo primordial dessa técnica é explicar a estrutura de variância e covariância de um vetor aleatório através de combinações lineares das variáveis originais. Essas combinações são nomeadas componentes principais e não apresentam correlação entre si.

### 3.1.2 Análise de Correlação de Spearman

A análise de correlação de Spearman é sugerida para os casos em que os dados não apresentam um comportamento comum, apresentam pontos bem distantes, e não apresentam relações visuais de crescimento ou decrescimento, ou ainda em caso de dados que não sigam a distribuição normal. É um método não-paramétrico que usa somente os postos, colocando os dados em uma escala ordinal e não sendo impactado pela presença de valores atípicos.

A fórmula para calcular o coeficiente de correlação de Spearman é dada por:

$$\rho = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n^3 - n}$$

Onde:



$\rho$  se refere a correlação de Spearman.

$d_i$  se refere a diferença entre os postos de duas variáveis – se os pontos de uma variável for igual a de outra determinada variável  $\rho$  será igual a 1. Em resumo, é equivalente a diferença entre as ordenações.

$n$  se refere ao número de pares de ordenações.

Neste trabalho, essa análise será feita para complementar a análise de componentes principais e eliminar possíveis variáveis correlacionadas que tenham restado nesta primeira métrica.

Segundo Downie e Heath (1959) este coeficiente varia entre -1 e 1, e quanto mais próximo estiver dos extremos significa maior associação. Quando negativa significa que as variáveis variam em sentido contrário, o que significa que enquanto uma cresce a outra decresce e vice-versa.

### 3.2 MODELOS CLÁSSICOS DE SÉRIES TEMPORAIS

Uma série temporal é definida como sendo um conjunto de dados observados e ordenados por um parâmetro de tempo. Para uma série de dados ser classificada como série temporal precisa cumprir o seguinte requisito: os dados devem apresentar uma dependência serial entre eles.

Segundo Granger (1977) o objetivo inicial das séries temporais é de retirar inferências sobre propriedades, características e tendências de um processo estocástico com a expectativa que haja uma relação causal com o tempo que possam influenciar os dados futuramente. É esta possibilidade de identificação de regularidades observáveis que possibilita a construção de modelos matemáticos que simplifiquem e retratem a realidade.

Para analisar uma série temporal é necessário analisar a decomposição das componentes de tendência, ciclo e sazonalidade. A tendência é um fator de comportamento que indica estabilidade e velocidade de mudanças. Os ciclos se referem a oscilações nas séries, ao longo da componente de tendência. E, a sazonalidade, se refere às oscilações que ocorrem em um determinado tempo  $t$ .

Para a aplicação desses modelos, é necessário o entendimento dos termos “**p**”, “**d**” e “**q**”, pois o ajuste a ser realizado será em função de tais variáveis. As definições são as seguintes:

- a) **p**: Faz referência à ordem autorregressiva do modelo.
- b) **d**: Remete ao grau de diferenciação (o número de vezes em que os dados tiveram valores passados subtraídos).
- c) **q**: Refere-se à ordem de médias móveis.

As próximas seções abordarão a respeito destes modelos.

### 3.2.1 Auto-regressivos

Os modelos auto-regressivos utilizam os valores passados da própria série temporal. Isto é, a regressão é realizada com uma variável sobre ela mesma. Segundo Morettin e Tolo (2006) este modelo pode ser escrito da seguinte forma:

$$Z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \dots + \phi_p z_{t-p} + \epsilon_t$$

Onde:

$z_t$  corresponde as observações históricas da série de interesse;

$\epsilon_t$  é uma variável aleatória com distribuição normal de média zero e variância igual a um, representando o erro da previsão, conhecido como ruído branco;

$\phi_1 \dots \phi_p$  são os parâmetros autorregressivos do modelo;

Este modelo é denotado como modelo AR(p), isto é, modelo auto-regressivo de ordem p. Este modelo tem algumas limitações, pois assume que existe uma relação linear entre os elementos da série e de que a média e o desvio padrão das observações não variam no tempo.

### 3.2.2 Média móvel

Ao contrário do modelo citado anteriormente, em que a regressão era feita com base nos valores passados da série, um modelo de média móvel faz uso dos erros do passado para realizar a regressão. Segundo Morettin e Tolo (2006) este modelo pode ser escrito da seguinte forma:

$$z_t = a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \dots + \theta_q a_{t-q} + \epsilon_t$$

Onde:

$z_t$  corresponde as observações históricas da série de interesse;

$\epsilon_t$  é uma variável aleatória com distribuição normal de média zero e variância igual a um, representando o erro da previsão, conhecido como ruído branco;

$a_{t-1} \dots a_{t-q}$  são choques aleatórios estatisticamente independentes que são assumidos por serem selecionados aleatoriamente a partir da distribuição normal com média zero e variância constante;

$\theta_1 \dots \theta_q$  são os parâmetros de médias móveis do modelo;

Este modelo é denotado como modelo MA(q), isto é, modelo de média móvel de ordem q.

### 3.2.3 Auto-regressivos de média móvel

Estes modelos são uma combinação dos modelos auto-regressivos e de média móvel, também chamados de ARMA, Segundo Morettin e Tolo (2006) este modelo pode ser escrito da seguinte forma:

$$z_t = \varphi_1 z_{t-1} + \varphi_2 z_{t-2} + \dots + \varphi_p z_{t-p} - a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q}$$

Onde:

$z_t$  corresponde as observações históricas da série de interesse;

$\varphi$  e  $\theta$  são respectivamente coeficientes autorregressivo e de médias móveis desconhecidos;

$a_{t-1} \dots a_{t-q}$  são choques aleatórios estatisticamente independentes que são assumidos por serem selecionados aleatoriamente a partir da distribuição normal com média zero e variância constante.

Um modelo ARMA incorpora a autocorrelação e a média móvel em uma regressão. É um modelo auto-regressivo estacionário onde as variáveis independente seguem tendencias estocásticas. As limitações citadas anteriormente são carregadas para este modelo.

### 3.2.4 ARIMA

Esta classe de modelos se refere a combinação de diferenciação com modelo de auto-regressão e modelo de média móvel. A diferenciação é a subtração entre valores consecutivos de uma série temporal.

Para as séries  $z_t$  para um número finito de diferenças,  $d$ , que se convertem em estacionárias, são chamadas não estacionárias homogêneas se obedecerem a seguinte equação:

$$W_t = \Delta^d Z_t$$

Se  $W_t$  é uma diferença de  $Z_t$ , então  $Z_t$  é uma integral de  $W_t$ , portanto pode-se dizer que  $Z_t$  possui um modelo autorregressivo-integrado-médias-móveis, representado da seguinte forma:

$$\varphi(X)\Delta^d Z_t = \theta(x)a_t$$

No modelo, todas as raízes de  $\varphi(x)$  estão fora do círculo unitário. Deve se considerar essa equação equivalente:

$$\xi(X) Z_t = \theta(x)a_t$$

Onde  $\xi(x)$  é um operador autorregressivo não estacionário de ordem  $p+d$ , com  $d$  raízes iguais a um e as restantes fora do círculo unitário, resultando na equação abaixo:

$$\xi(X) = \varphi(X)\Delta^d = \varphi(X)(1 - X)^d$$

Este modelo é denotado como modelo ARIMA( $p,d,q$ ), isto é, modelo autorregressivo de ordem  $p$ , diferenciação de grau  $d$  e média móvel de grau  $q$ .

## 3.3 MODELOS DE MACHINE LEARNING

Cada vez mais se apresenta uma maior gama de dados e maiores desafios em termos de predição, principalmente em termos de economia e mais do que isso, quando se fala na economia brasileira. Ao lidar com esses grandes desafios aparece na literatura mais recente, como alternativa, o uso dos algoritmos de *machine learning*, ou em tradução livre, aprendizado de máquina.

Lantz (2015) descreve o *machine learning* por estudo interessado no desenvolvimento de algoritmos capazes de transformar dados em ação. Géron (2019),

define *machine learning* como a ciência de programar computadores para que eles aprendam por meio de dados.

Em síntese, modelos de aprendizado de máquina são uma subcategoria dos modelos de inteligência artificial, e se refere ao processo de um computador desenvolvendo reconhecimento de padrões automaticamente. O funcionamento de cada um desses modelos depende do algoritmo utilizado, neste trabalho serão utilizadas algumas técnicas e elas serão detalhadas nas seções seguintes.

Contudo, é importante mencionar que dependendo da técnica escolhida os algoritmos funcionam com mais ou menos intervenção humana. Os tipos mais conhecidos são os seguintes:

Aprendizado supervisionado: o computador aprende como o humano faz a tarefa através de dados organizados, essa é a técnica menos complexa e a que será aplicada neste trabalho.

Aprendizado não supervisionado: os dados que o computador recebe não são rotulados, havendo a necessidade de reconhecer padrões a partir do desconhecido (sem intervenção humana).

As próximas seções abordarão alguns desses algoritmos, os que serão utilizados na construção dos resultados deste trabalho.

### **3.3.1 Random Forest**

Quando se menciona o uso de *machine learning* os modelos mais comuns se referem a árvore de decisão, que se apresentam como combinações de resultados de vários classificadores. O objetivo dessas árvores de decisão é desmembrar todos os dados que o sistema pode aprender, e em um segundo momento agrupar os dados em grupos com membros semelhantes entre si e diferentes entre os grupos. A árvore de decisão descreve as diferentes variáveis que podem determinar com maior poder um resultado específico.

Iniciado por Breiman (2001) este algoritmo é um nome geral para métodos de *ensemble* utilizando classificadores do tipo árvore. Métodos *ensemble*, por sua vez, são construídos da mesma forma que outros algoritmos mais conhecidos como regressão

linear e árvore de decisão, onde a grande diferença é a combinação de diferentes modelos para se obter um único resultado, tornando o algoritmo mais robusto.

Normalmente, o que faz o cientista de dados escolher a técnica de *machine learning* é a acurácia obtida, com um método *ensemble* serão criados e otimizados diferentes modelos a partir de um mesmo algoritmo, no caso deste trabalho *Random Forest*.

O algoritmo de *Random Forest*, em síntese, se refere a criação de várias árvores de decisão de maneira aleatória, formando o que se chama de uma floresta, onde cada árvore será utilizada na escolha do resultado, isto é, uma regressão no caso deste trabalho. Neste algoritmo, são construídos uma grande quantidade de árvores de decisão sob o conjunto de dados definido como conjunto de treinamento. Extraíndo casos aleatoriamente a partir do conjunto total construindo árvores específicas, cada árvore classificadora é um componente preditor. A *Random Forest* constrói a decisão em uma espécie de contagem de votos dos componentes preditores, selecionando a classe vencedora em termos de quantidade de votos acumulados. Dessa forma, podemos concluir que existem duas fases: a formação de cada árvore e a votação.

Neste trabalho será apresentado o desempenho do algoritmo de *Random Forest* para fins de comparação com outros algoritmos de *machine learning* e os modelos clássicos de séries temporais na previsão da taxa de desemprego.

### **3.3.2 SVM – (Support Vector Machine)**

Este algoritmo tem como objetivo a determinação de limites de decisão que produzam uma separação ótima entre classes através da minimização dos erros, com o intuito de reconhecimento de padrões.

A teoria foi introduzida por Vapnik (1995), e recentemente, foi utilizado na área de sensoriamento remoto com relativo sucesso (Brown et al., 2000; Melgani e Bruzzone, 2004).

Segundo Nascimento et al. (2009) o processo de treinamento consiste em treinar um classificador de forma que este aprenda a mapear as classes e a máquina seja capaz

de classificar um exemplo, ainda que não siga a mesma distribuição de probabilidade dos exemplos de treinamento.

O diferencial desse método é minimizar a expectativa de erro de classificação, considerando a não necessidade de conhecer a distribuição de probabilidade.

Neste trabalho será apresentado o desempenho do algoritmo para fins de comparação com outros algoritmos de *machine learning* e os modelos clássicos de séries temporais na previsão da taxa de desemprego.

### 3.3.3 Lasso Regression

Segundo Virgolino (2017), o LASSO (*least absolute shrinkage and selection operator*), é um método de redução do conjunto de coeficientes e tem como objetivo estimar um modelo que determine o conjunto de preditores que melhor expliquem a variável resposta, e produza previsões com mínima variância. Dessa forma, pode-se dizer que este algoritmo realiza a seleção de recursos automaticamente, gerando vários coeficientes com peso zero, que são ignorados pelo modelo, o que facilita a interpretação dos resultados.

O estimador via LASSO no contexto da regressão linear é dado pela resolução da equação:

$$\hat{\beta} = \arg \min_{\beta} \|Y - X\beta\|_2^2 + \lambda \sum_{j=1}^P |\beta_j|$$

Onde,

$\beta$  é um vetor  $n \times 1$ ;

$Y$  se refere a variável resposta;

$X$  é uma matriz  $p \times n$  com as variáveis preditoras; e

$\lambda$  é o parâmetro de redução de dimensionalidade.  $\lambda$  é estabelecido por validação cruzada, esta técnica particiona a amostra original em  $K$  subamostras de tamanho igual, uma das  $K$  subamostras é retirada e o modelo é estimado com  $K-1$  subamostras. se  $\lambda = 0$ , então o estimador via LASSO será igual ao estimador via mínimos quadrados, logo, o estimador de mínimos quadrados de  $\beta$  é um caso particular do estimador via LASSO.

O algoritmo estimador via LASSO é em síntese um problema de minimização dos erros ao quadrado acrescido de uma penalização em relação aos parâmetros do modelo de regressão.

Considerando consistência dos modelos LASSO, é importante verificar se a solução representa bem o modelo, conseguindo selecionar o subconjunto correto de variáveis relevantes e seus parâmetros devem ser assintoticamente como dos estimadores MQO, o que só acontecerá se o subconjunto de variáveis escolhido for o das variáveis relevantes para o modelo.

Neste trabalho, o principal ganho deste algoritmo e dos outros algoritmos de *machine learning* em relação aos modelos clássicos de séries temporais é a possibilidade de analisar outras variáveis que podem ser consideradas importantes para previsão de desemprego.

### 3.4 BASE DE DADOS

Para condução do estudo e testes de modelos estatísticos faz-se necessária a construção do banco de dados. O mapeamento das variáveis de interesse foi solidificado na revisão bibliográfica, que foi construída com o desejo de passar pelas primeiras escolas que estudaram desemprego e chegar aos dias atuais. Para redigir a revisão bibliográfica foram utilizados como fontes: Lume UFRGS, Google acadêmico, Scielo, Portal de periódicos da Capes, Biblioteca digital de teses e dissertações da USP, e algumas revistas como: Revista de economia e agronegócio, Revista de economia contemporânea, Revista de economia, Revista estudos sociais, Revista de economia política, entre outras.

Abaixo estão listadas as variáveis mapeadas, bem como a fonte e uma breve descrição. Para todas as variáveis foi observado o período de 2000 até 2021.



Tabela 1 – Variáveis macroeconômicas mapeadas

(continua)

| Variável   | Significado   | Fonte                   |
|--|---|-------------------------|
| Despesa com Educação (valor nominal)                                 | Se refere ao gasto público com educação.  | Tesouro Nacional        |
| Dívida Pública (valor do último mês do ano em % do PIB)              | Para poder utilizar verbas públicas o Governo Federal trabalha com um orçamento elaborado no ano anterior através da Lei Orçamentaria Anual. Nesta lei há uma estimativa de receitas que deverão ingressar no cofre público no ano seguinte e despesas de cada setor. Quando o governo não consegue arrecadar verba suficiente para cobrir as despesas há um déficit orçamentário, originando a necessidade de captar recursos, o que gera a dívida pública. É importante destacar que todos os países têm dívida, os componentes que são relevantes para avaliar a qualidade são: prazo, composição e custo. | Ipea Data               |
| Dólar comercial (média anual)  | Valor da moeda utilizada no pagamento ou recebimento de recursos das importações e exportações de bens e serviços do Brasil com o mundo.  | Ipea Data               |
| Exportação (Valor em U\$\$)  | Se refere ao valor total de venda, envio ou doação de bens e serviços do Brasil para outros países.   | Ministério da Economia  |
| Famílias atendidas pelo bolsa família (valor em milhões)             | Se refere ao número total de famílias atendidas pelo benefício social.  | Portal da transparência |
| ICC - indicador de custo de crédito (média anual)                    | O ICC tem a finalidade de estimar o custo médio, em um mês, que onera as famílias e as empresas que tenham tomado crédito no sistema financeiro nacional.   | Ipea Data               |
| ICEI - indicador de confiança do empresário industrial (média anual) | Um indicador antecedente do desempenho industrial, sinaliza as mudanças de tendência da produção industrial.  | Ipea Data               |
| IDH (índice de desenvolvimento humano)                               | O índice de desenvolvimento humano compara indicadores de países nos itens riqueza, alfabetização, educação, esperança de vida, natalidade e outros, com o intuito de avaliar o bem-estar de uma população, especialmente das crianças, variando entre 0 e 1.   | Country Economy         |
| IEC - indicador de expansão do comércio (média anual)                | Representa interesse absoluto na expansão dos negócios.   | Ipea Data               |
| Importação (Valor em U\$\$)  | Se refere ao valor total de venda, envio ou doação de bens e serviços de outros países para o Brasil  | Ministério da Economia  |

**Tabela 1 – Variáveis macroeconômicas mapeadas**

(conclusão)

| <b>Variável</b>   | <b>Significado</b>   | <b>Fonte</b>            |
|---|--|-------------------------|
| Índice de Gini  | Instrumento utilizado para medir o grau de concentração de renda em determinado grupo, quanto maior melhor o indicador.  | FGV e PNAD              |
| Inflação  | É o aumento de preços dos bens e serviços que culmina em redução do poder de compra. O Brasil tem vários índices de preços, o Índice Nacional de Preços ao Consumidor Amplo (IPCA) é o índice utilizado no sistema de metas para a inflação.   | Inflation               |
| PIB   | O Produto Interno Bruto de um país é a soma de todos os bens e serviços finais produzidos por um país, em um determinado período.  | IBGE                    |
| Salário Mínimo (último valor do ano)  | Se refere ao menor salário que uma empresa pode pagar para um funcionário, sendo estabelecido por lei e revisado anualmente com base no custo de vida, objetivando garantir o mínimo para sobrevivência.   | Ipea Data               |
| Taxa de Desemprego (média anual)  | Se refere às pessoas com idade para trabalhar (acima de 14 anos, segundo o IBGE) que não estão trabalhando, mas estão disponíveis e tentando encontrar emprego. Não sendo consideradas como desempregadas as pessoas que dedicam seu tempo somente aos estudos, donas(os) de casa que optam por não trabalhar fora e empreendedores dos seus próprios negócios.          | IBGE                    |
| Taxa de Não Alfabetizados   | Se refere a quantidade de famílias com renda familiar mensal per capita de até meio salário-mínimo relativizada pela quantidade total de famílias na população.  | IBGE e PNAD             |
| Taxa de Pobreza   | Se refere a quantidade de famílias com renda familiar mensal per capita de até meio salário-mínimo relativizada pela quantidade total de famílias na população.  | FGV e Cepal             |
| Taxa Selic (taxa média diário de juros anualizada com base em 252 dias úteis) | É a taxa básica de juros da economia, o principal instrumento de política monetária utilizada pelo Banco Central para controlar a inflação, influenciando todas as taxas de juros do país. A taxa Selic refere-se à taxa de juros apurada nas operações de empréstimos de um dia entre as instituições financeiras que utilizam títulos públicos federais como garantia. | Banco Central do Brasil |
| Valor de Investimentos (média trimestral do ano em % do PIB)                  | Se refere a taxa pessoas com 15 anos ou mais não alfabetizadas em relação a população total.   | Ipea Data               |

Fonte: Elaborada pela autora (2022).

Neste momento cabe realizar um breve contexto sobre a expectativa do relacionamento de cada uma dessas variáveis com a taxa de desemprego.

- a) Despesa com Educação: Espera-se que um aumento nessa variável indique uma redução de taxa de desemprego futura.

De acordo com a Secretaria de Política Econômica do Ministério da Economia (2021), metade dos brasileiros desempregados por muito tempo possuem idade entre 18 e 29 anos e que 80% desses jovens possuem apenas o ensino médio.

Segundo o Estadão (2021), um estudo realizado pelo Insper e Fundação Roberto Marinho, mostrou que o Brasil perde anualmente R\$ 214 bilhões, o equivalente a 3% do PIB, com a evasão escolar e isso afeta diretamente a empregabilidade e a remuneração desses jovens no mercado.

Dessa forma, há indícios de que a redução da taxa de desemprego passa pelo o crescimento do país que precisa investir, entre outras variáveis, em educação.

- b) Dívida Pública em relação ao PIB: Espera-se que um aumento nessa variável indique um crescimento de taxa de desemprego futura. A dívida pública é uma situação normal de todos os países, mesmo os mais desenvolvidos e deve ser sempre analisada em relação aos seus custos, composição e prazo.

Além disso, essa variável está sendo analisada em relação ao PIB, então o crescimento pode ser em função de uma redução do PIB, tendo como consequência o sacrifício de outros âmbitos sociais como: saúde, educação ou previdência, por exemplo.

- c) Dólar comercial: Espera-se que um aumento nessa variável indique um crescimento de taxa de desemprego futura. Aumentos nessa variável são positivos para países que produzem bastante em moeda nacional e vendem em dólar, o que não é o caso do Brasil.

O Brasil faz uso de muitos insumos externos, e um aumento dessa variável faz com que o cenário geral fique mais caro e abre duas possibilidades: repassar ao consumidor ou reduzir pessoas.

Portanto um aumento dessa variável dificulta as relações comerciais e geral um cenário de desconfiança e incerteza.

- d) Exportação: Espera-se que um aumento nessa variável indique uma redução de taxa de desemprego futura. A hipótese *Export-Led Growth* (ELG) afirma que as exportações impulsionam o crescimento econômico. Pelo lado da demanda aumenta o grau de abertura da economia e potencializa a dimensão do mercado. Pelo lado da oferta, promove o crescimento econômico através de produtividade, o que impacta diretamente os níveis de emprego.
- e) Famílias atendidas pelo bolsa família: Espera-se que um aumento nessa variável indique uma redução de taxa de desemprego futura. Este é um indicador social, é natural que com a melhoria das condições sociais haja paralelamente um maior número de pessoas empregadas.
- f) ICC - indicador de custo de crédito: Espera-se que um aumento nessa variável indique um crescimento de taxa de desemprego futura. Quanto mais caro fica o custo de crédito maior estão as taxas de juros, e taxas de juros mais altas são parte de políticas monetárias de períodos recessivos.
- g) ICEI - indicador de confiança do empresário industrial: Espera-se que um aumento nessa variável indique uma redução de taxa de desemprego futura. Quando cresce o índice de confiança, crescem as possibilidades de nascimento de novas empresas e conseqüentemente de novos empregos.
- h) IDH - índice de desenvolvimento humano: Espera-se que um aumento nessa variável indique uma redução de taxa de desemprego futura. Este é um indicador social, é natural que com a melhoria das condições sociais haja paralelamente um maior número de pessoas empregadas.
- i) IEC - indicador de expansão do comércio: Espera-se que um aumento nessa variável indique uma redução de taxa de desemprego futura. Quando há expectativa de expansão do comércio, crescem as possibilidades de nascimento de novas empresas e conseqüentemente de novos empregos.
- j) Importação: Espera-se que um aumento nessa variável indique um crescimento de taxa de desemprego futura.

Importação nos remete a desindustrialização que impacta o nível de emprego, que é uma ligação forte com produtividade. Não à toa no passado recente passamos pelo processo de substituição de importação, um dos traços mais marcantes desse período é o crescimento dos níveis de emprego em 1933.

k) Índice de Gini: Espera-se que um aumento nessa variável indique uma redução de taxa de desemprego futura. Esse indicador está diretamente ligado a renda, se ele está maior significa que o país está cada vez mais igual em termos de distribuição de renda, o que em hipótese está ligado a mais pessoas empregadas.

l) Inflação: Espera-se que um aumento nessa variável indique uma redução de taxa de desemprego futura.

Segundo a teoria da Curva de Phillips essas duas variáveis tem uma relação inversa, quando há mais empregos, os preços costumam subir, provocados por um aumento no poder de compra e conseqüente aumento da demanda. Portanto, a queda do desemprego pode influenciar no aumento da inflação.

m) PIB: Espera-se que um aumento nessa variável indique uma redução de taxa de desemprego futura.

Segundo a Lei de Okun, elaborada em 1960, quando a taxa de desemprego diminui o crescimento econômico aumenta.

Por outro lado, em momentos de crescimento econômico também se espera que cresçam as ofertas de emprego, reduzindo assim os níveis da taxa de desemprego.

n) Salário Mínimo: Espera-se que um aumento nessa variável indique um crescimento de taxa de desemprego futura.

Souza e Targino (2005) apontam que ao pegar o salário mínimo como referência de mão de obra pouco qualificada, mudanças no seu valor aumentam os custos para as empresas. Essas elevações tendem a provocar mudanças nas relações de trabalho, provocando maior grau de informalidade, alterando a estrutura econômica de renda e emprego.

- o) Taxa de Não Alfabetizados: Espera-se que um aumento nessa variável indique um crescimento de taxa de desemprego futura.
- Nos dados do IBGE (2017) é possível verificar a proximidade dos dados de desemprego geral e de quem tem menos de 1 ano de estudo, enquanto o patamar de desemprego geral foi de 11,8%, o patamar entre quem tem menos de 1 ano de estudo foi de 9,2%, indicando semelhança e mesma direção dos indicadores.
- p) Taxa de Pobreza: Espera-se que um aumento nessa variável indique um crescimento de taxa de desemprego futura.
- Intrinsicamente a taxa de pobreza é uma consequência da situação da taxa de desemprego, num país com a desigualdade do patamar do nosso é comum observarmos dados que apontam para a crescimento de taxa de desemprego entre os mais pobres. Como podemos ver na reportagem da Folha de São Paulo (2021), que traz a informação de que durante a pandemia a taxa de desemprego da metade mais pobre dos brasileiros subiu dez pontos percentuais, enquanto entre os 10% mais ricos o aumento foi de 0,27.
- Entretanto, o objetivo dessa variável nesse contexto é avaliar se em períodos mais longos ela também não possa ter associação com a taxa de desemprego.
- q) Taxa Selic: Espera-se que um aumento nessa variável indique um crescimento de taxa de desemprego futura.
- Esta taxa é o principal instrumento de política monetária utilizada pelo Banco Central para controlar a inflação impactando a economia como um todo. Quanto ao desemprego, a taxa Selic está ligada ao custo do dinheiro, se a Selic está baixa há tendência de aumento de consumo, mais demanda e conseqüentemente maior necessidade de produção e o contrário também é válido.
- r) Valor de Investimentos: Espera-se que um aumento nessa variável indique uma redução de taxa de desemprego futura.

O montante dos investimentos constrói o estoque de capital físico necessário para fixar o crescimento da economia, culminando em melhor distribuição de renda e redução da taxa de pobreza, o que dificilmente acontece sem estar alicerçado a melhores índices da taxa de desemprego. Ainda, Siqueira e Orair (2018) apontam que os investimentos públicos são instrumentos de política anticíclica de estímulo da economia em períodos recessivos.

## 4 RESULTADOS

Neste capítulo estarão dispostos os principais resultados obtidos com a aplicação da metodologia explicada anteriormente.

### 4.1 SELEÇÃO DE VARIÁVEIS

O próximo passo metodológico é referente as técnicas estatísticas estudadas para fazer a seleção das variáveis relevantes para prever a taxa de desemprego.

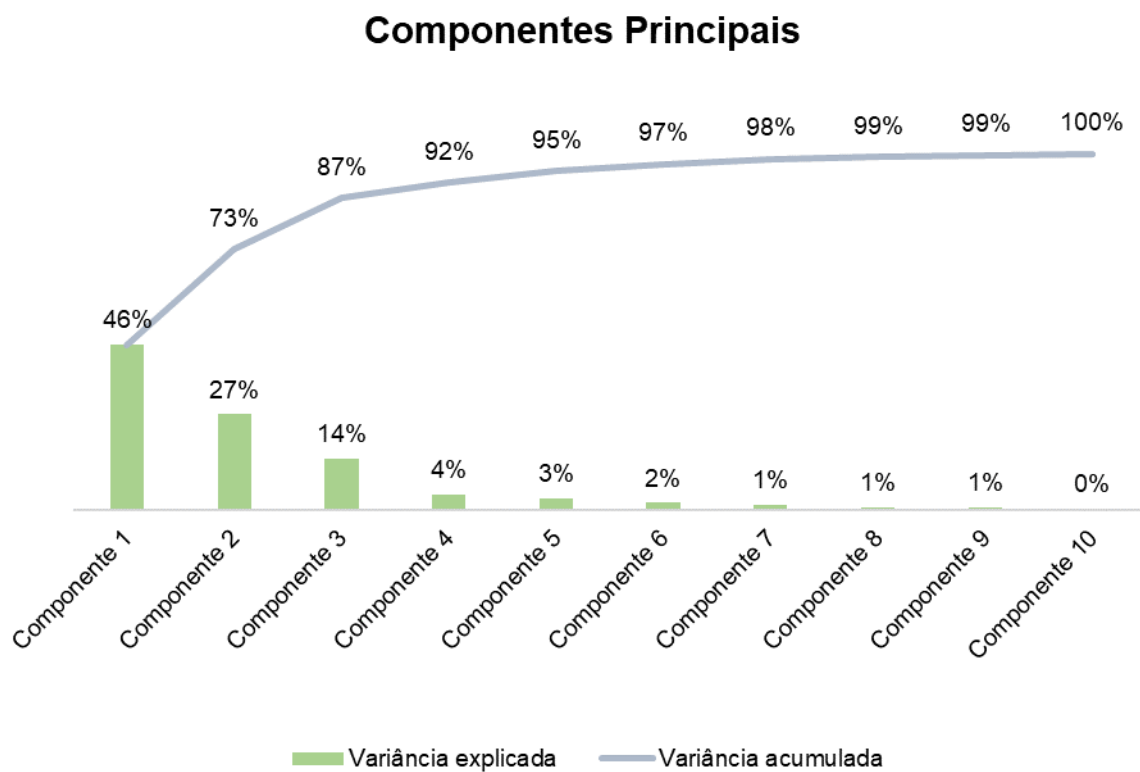
Componentes principais: Também conhecida como PCA (*Principal Componente Analysis*) é uma técnica multivariada que tem o objetivo de encontrar a menor quantidade de variáveis que tem perda mínima de informações no contexto de explicação da taxa de desemprego. Para conseguir alcançar esse objetivo é realizada a rotação e translação no espaço de todas as variáveis, de tal forma que os novos eixos são escolhidos pela variação que contêm. Para que essa análise seja eficiente, a informação contida nas medidas fica concentrada em um número reduzido de componentes que possuem variação significativa. As novas componentes são ordenadas de forma decrescente na quantidade de variação que apresentam, essas variáveis finais são chamadas de componentes principais.

Nesta etapa, antes de conseguir conduzir a análise é necessário popular os dados faltantes, pois só é possível conduzir o cálculo dessa forma. Como medida de contorno para os dados faltantes foi imputado a média da série, um contorno comum de tratamento estatístico, conforme mencionado por Klück, Fachel e Nunes (2009).

Depois foram construídas as componentes que são medidas por autovalores, como resultado podemos ver na figura abaixo que 3 componentes explicam cerca de 87% da variância dos dados.



Gráfico 1 – Percentual de variância explicada: Componentes Principais



Fonte: Elaboração própria a partir dos resultados obtidos no software R (2022).

Identificadas as componentes principais é necessário encontrar as variáveis que compõe cada uma delas, bem como seu grau de importância. A tabela a seguir traz o índice de correlação de cada variável dentro de cada componente, considerando que na etapa anterior optou-se por seguir com as três primeiras dimensões.

**Tabela 2 – Relevância das variáveis nas componentes**

| Variável                              | Componente 1 | Componente 2 | Componente 3 |
|---------------------------------------|--------------|--------------|--------------|
| PIB                                   | 34%          | 46%          | 1%           |
| Taxa Selic                            | 62%          | 12%          | 16%          |
| Dívida Pública                        | 6%           | 32%          | 55%          |
| Inflação                              | 15%          | 22%          | 28%          |
| Valor do Investimento                 | 1%           | 39%          | 44%          |
| Exportação                            | 76%          | 18%          | 0%           |
| Importação                            | 67%          | 22%          | 5%           |
| Salário Mínimo                        | 91%          | 1%           | 6%           |
| IDH                                   | 91%          | 1%           | 2%           |
| Taxa Não Alfabetizados                | 90%          | 0%           | 4%           |
| Despesa com Educação                  | 89%          | 4%           | 2%           |
| Dólar Comercial                       | 20%          | 42%          | 27%          |
| Índice de Gini                        | 74%          | 6%           | 3%           |
| Famílias atendidas pelo bolsa família | 54%          | 1%           | 2%           |
| Taxa de Pobreza                       | 82%          | 6%           | 2%           |
| ICC                                   | 1%           | 84%          | 2%           |
| IEC                                   | 3%           | 67%          | 1%           |
| ICEI Atual                            | 12%          | 49%          | 29%          |
| ICEI Expectativa                      | 29%          | 38%          | 28%          |
| ICEI Geral                            | 23%          | 43%          | 29%          |

Fonte: Elaboração própria a partir dos resultados obtidos no software R (2022).

Seguirão na análise as variáveis que tiveram mais de 50% de correlação com pelo menos uma dimensão, são elas: Taxa Selic, Dívida Pública, Exportações, Importações, Salário mínimo, IDH, Taxa de não alfabetizados, Despesa com educação, Índice de Gini, Número de famílias atendidas pelo bolsa família, Taxa de Pobreza, ICC e IEC.

Correlação de Spearman: Como já mencionado, trata-se de uma medida não paramétrica da dependência das variáveis. Feita a seleção das variáveis via Componentes Principais, as mesmas ainda passaram por uma etapa de análise de correlação. Isto é, tem-se como objetivo verificar se as variáveis não estão redundantes, usando como balizador para manter as variáveis que a correlação delas sejam menores do que 90%, em ambos sentidos.

As tabelas abaixo representam os indicadores de correlação das variáveis resultantes da análise de componentes principais:

Tabela 3 – Correlação de Spearman das variáveis

| VARIÁVEL             | TAXA SELIC | DÍVIDA PÚBLICA | EXPOR TACAO | IMPOR TACAO | SALÁRIO MÍNIMO | IDH  | TAXA Ñ ALFABETIZADOS | DESPESA EDUCACAO |
|----------------------|------------|----------------|-------------|-------------|----------------|------|----------------------|------------------|
| TAXA SELIC           | 100%       | 8%             | -78%        | -69%        | -81%           | -76% | 81%                  | -78%             |
| DÍVIDA PÚBLICA       |            | 100%           | -42%        | -59%        | -5%            | -8%  | 4%                   | -9%              |
| EXPORTACAO           |            |                | 100%        | 95%         | 79%            | 78%  | -79%                 | 80%              |
| IMPORTACAO           |            |                |             | 100%        | 73%            | 73%  | -72%                 | 73%              |
| SALÁRIO MÍNIMO       |            |                |             |             | 100%           | 99%  | -100%                | 98%              |
| IDH                  |            |                |             |             |                | 100% | -98%                 | 96%              |
| TAXA Ñ ALFABETIZADOS |            |                |             |             |                |      | 100%                 | -98%             |
| DESPESA EDUCACAO     |            |                |             |             |                |      |                      | 100%             |

Fonte: Elaboração própria a partir dos resultados obtidos no software SAS (2022).

Tabela 4 – Correlação de Spearman das variáveis

| VARIÁVEL               | DESPESA EDUCACAO | ÍNDICE GINI | FAMILIAS BOLSA FAMILIA | TAXA POBREZA | ICC  | IEC  |
|------------------------|------------------|-------------|------------------------|--------------|------|------|
| TAXA SELIC             | -78%             | 62%         | -67%                   | 62%          | -42% | -42% |
| DÍVIDA PÚBLICA         | -9%              | 49%         | -20%                   | 51%          | -59% | -36% |
| EXPORTACAO             | 80%              | -72%        | 72%                    | -78%         | 53%  | 47%  |
| IMPORTACAO             | 73%              | -77%        | 71%                    | -84%         | 55%  | 43%  |
| SALÁRIO MÍNIMO         | 98%              | -76%        | 82%                    | -81%         | 12%  | 14%  |
| IDH                    | 96%              | -79%        | 83%                    | -82%         | 7%   | 8%   |
| TAXA Ñ ALFABETIZADOS   | -98%             | 76%         | -81%                   | 80%          | -11% | -12% |
| DESPESA EDUCACAO       | 100%             | -71%        | 80%                    | -83%         | 11%  | 12%  |
| ÍNDICE GINI            |                  | 100%        | -75%                   | 84%          | -24% | -16% |
| FAMILIAS BOLSA FAMILIA |                  |             | 100%                   | -86%         | -2%  | -3%  |
| TAXA POBREZA           |                  |             |                        | 100%         | -2%  | -7%  |
| ICC                    |                  |             |                        |              | 100% | 91%  |
| IEC                    |                  |             |                        |              |      | 100% |

Fonte: Elaboração própria a partir dos resultados obtidos no software SAS (2022).

Para decidir entre os pares correlacionados foi observado a correlação da variável individualmente com a taxa de desemprego, fazendo escolha sempre pela variável de maior índice de associação. A tabela abaixo traz os valores de correlação individual de cada variável com a taxa de desemprego.

**Tabela 5 – Correlação de Spearman individual das variáveis vs Taxa de Desemprego**

| VARIÁVEL               | CORRELAÇÃO |
|------------------------|------------|
| ICC                    | -65,2%     |
| IEC                    | -48,4%     |
| TAXA Ñ ALFABETIZADOS   | -35,2%     |
| IMPORTACAO             | -23,9%     |
| EXPORTACAO             | -7,7%      |
| TAXA SELIC             | -1,9%      |
| TAXA POBREZA           | 4,1%       |
| ÍNDICE GINI            | 8,7%       |
| FAMILIAS BOLSA FAMILIA | 14,1%      |
| DESPEZA EDUCACAO       | 33,1%      |
| IDH                    | 34,9%      |
| SALÁRIO MÍNIMO         | 35,2%      |
| DÍVIDA PÚBLICA         | 77,4%      |

Fonte: Elaboração própria a partir dos resultados obtidos no software SAS (2022).

Nesta etapa, as variáveis que se mantiveram foram: Taxa Selic, Dívida Pública, Importações, Salário Mínimo, Índice de Gini, Número de famílias atendidas pelo bolsa família, Taxa de Pobreza e ICC.

## 4.2 PREDIÇÃO

As próximas subseções abordarão a respeito das predições construídas através dos algoritmos descritos na metodologia deste trabalho.

### 4.2.1 Modelos clássicos de série temporal

A primeira aplicação na modelagem dos dados foi univariada, isto é, utilizando apenas a própria taxa de desemprego aplicando as classes de modelos de séries temporais mencionadas anteriormente, os modelos da classe ARIMA. Todo banco de dados foi utilizado para desenvolvimento do modelo preditivo.

A figura a seguir traz os dados anuais de taxa de desemprego, com o objetivo de observar se os dados têm média e variância constante, o que é um pressuposto dos modelos da classe ARIMA.

Gráfico 2 – Série histórica da Taxa de Desemprego anual média

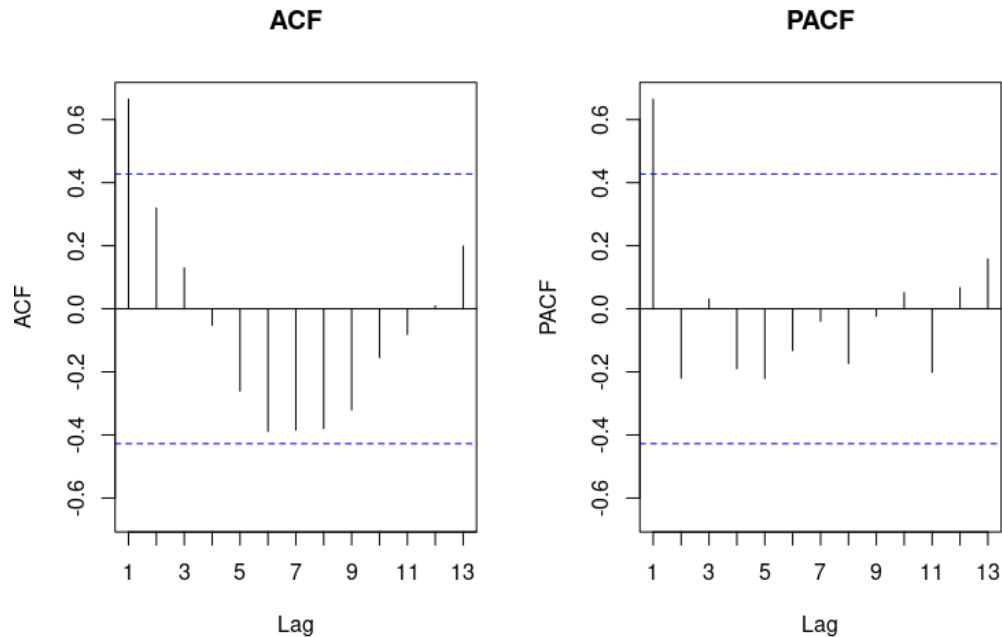


Fonte: Dados do IBGE (2021). Elaboração própria (2022).

O gráfico não é suficiente para essa tomada de decisão, então foi realizado o teste KPSS (Kwiatkowski, Phillips, Schmidt e Shin), conhecido também como teste da raiz unitária, que tem como hipótese nula a estacionariedade dos dados. O p-valor do teste foi de 0,259 ( $>0,05$ ), o que indica que não se rejeita a hipótese de estacionariedade dos dados.

Além disso, também foram observados os gráficos de ACF (autocorrelação) e PACF (autocorrelação parcial), para observar a autocorrelação da série e já visualizar o ajuste do modelo. A correlação expõe o quão relacionado é um valor do presente com outro valor do passado, a diferença entre a autocorrelação e a autocorrelação parcial é que em na ACF há a correlação direta e indireta e em na PACF apenas a correlação direta.

Gráfico 3 – ACF E PACF da série



Fonte: Elaboração própria a partir dos resultados obtidos no software R (2022).

Dado que neste trabalho há apenas 22 observações e que o teste KPSS não rejeitou a hipótese de estacionariedade optou-se por seguir a análise considerando a série no formato inicial, mesmo o teste ADF não tendo rejeitado a hipótese nula, que se refere a ter raiz unitária. Quanto as visualizações de ACF e PACF há indícios de que o ajuste de modelo será autorregressivo de ordem um.

Para fazer o ajuste do modelo foi utilizada a função `auto.arima` do *software* R, que tem como objetivo encontrar o melhor modelo da classe ARIMA para o ajuste dos dados. O modelo estimado por esse algoritmo foi um modelo autorregressivo de ordem um, isto é AR(1), como indicavam os gráficos de autocorrelação e autocorrelação parcial.

Os resultados do modelo, em termos de coeficientes e medidas de acurácia estão postos a seguir:

Tabela 6 – Resultados do modelo AR

| Amostra         | Coefficiente | Erro médio absoluto (MAE) | Raíz do Erro Quadrático Médio (RMSE) | R <sup>2</sup> |
|-----------------|--------------|---------------------------|--------------------------------------|----------------|
| Desenvolvimento | 0,7805       | 1,2278                    | 1,6003                               | 58,44%         |

Fonte: Elaboração própria a partir dos resultados obtidos no software R (2022).

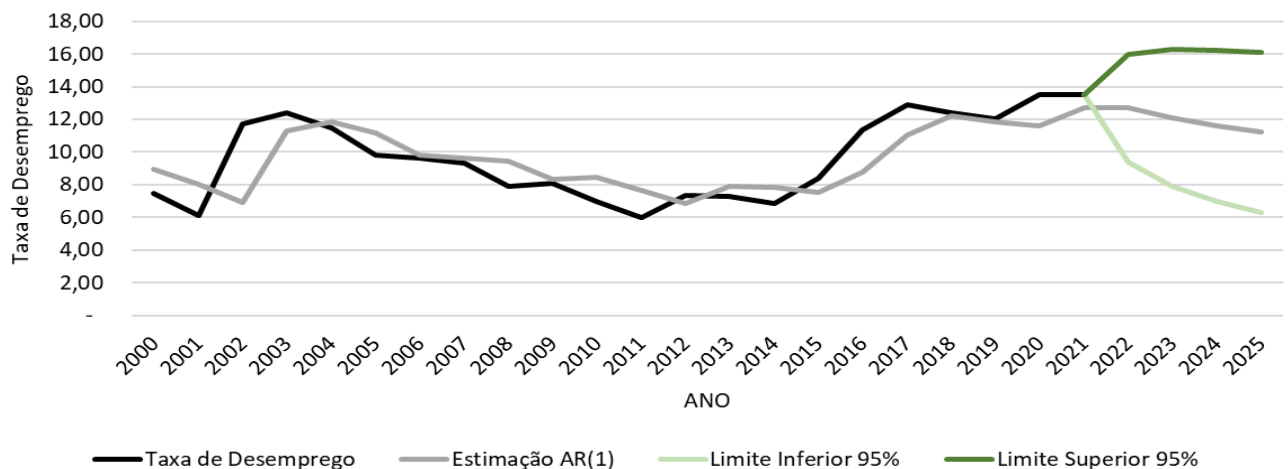
Também foi aplicado o teste Box-Ljung (1978) que visa avaliar a falha de ajuste de um modelo de série temporal. O teste é aplicado aos resíduos de uma série temporal

após o ajustamento de um modelo aos dados e verifica a autocorrelações dos resíduos. Se as autocorrelações forem muito pequenas, conclui-se que o modelo não exhibe falha significativa de ajuste. O resultado do teste foi 0,3628, não rejeitando a hipótese nula quanto a qualidade do ajuste, indicando como um bom modelo e sem indicativos de autocorrelação dos resíduos.

Considerando estes testes e comprovando a qualidade do modelo foram estimadas previsões para o período de 2022 até 2025, bem como seu intervalo de confiança (95%), dado o objetivo deste trabalho ser em relação a previsão do patamar da taxa e não propriamente o ponto em si, o ajuste do modelo e as previsões estão dispostas no gráfico a seguir. É válido destacar que a média anual da taxa de desemprego de 2022 até o momento (setembro/2022) é de 9,96%.

**Gráfico 4 – Ajuste do modelo AR**

**Ajuste do Modelo e Previsão**



Fonte: Dados do IBGE (2021). Elaboração própria a partir dos resultados obtidos no software R (2022).

A próxima seção trata do ajuste de modelos via algoritmos de *machine learning*, onde são consideradas as variáveis selecionadas anteriormente para compor a previsão. Posteriormente, serão comparados os resultados com os ajustes e previsões expostos até aqui.

#### 4.2.2 Modelos de *machine learning*

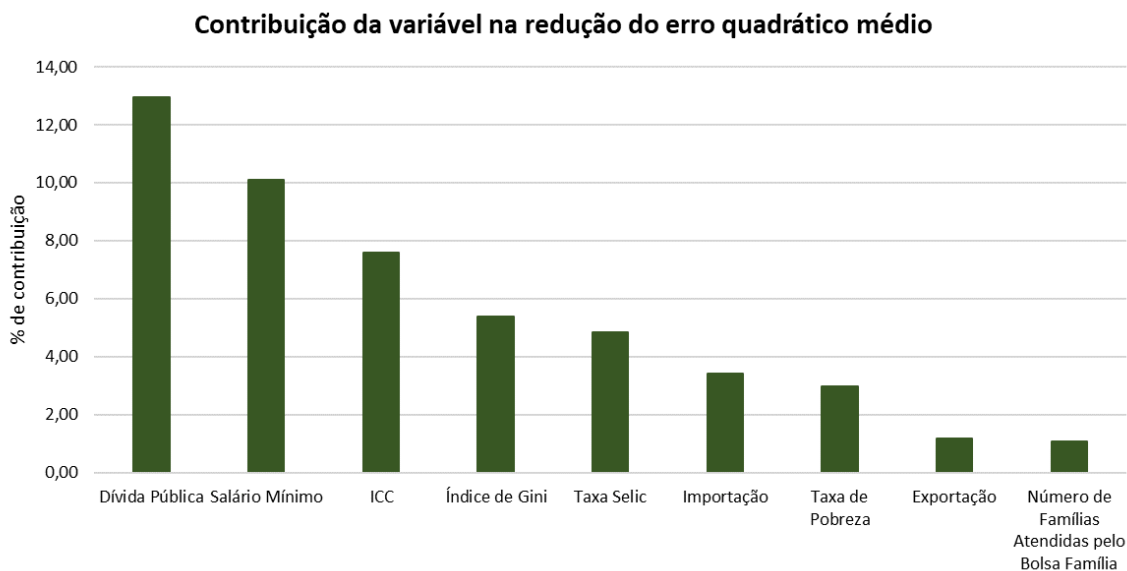
Conforme descrito na metodologia do trabalho foram testados diversos algoritmos de *machine learning* e estarão aqui dispostos os de menor erro.

*Random Forest*: A modelagem dos dados partiu do conjunto de dados estipulado no capítulo de seleção de variáveis, considerando todas as variáveis no ano corrente e a taxa de desemprego resultante no ano seguinte, de forma que seja possível selecionar as variáveis mais importantes em um ano zero, para previsão do patamar da taxa de desemprego do ano seguinte. Da mesma forma em que foi construído o ajuste de modelos através de séries temporais, aqui também foi utilizada toda base para desenvolvimento.

Para fazer o ajuste do modelo foi utilizada a função *randomForest* no *software R*, que tem como objetivo ajustar modelos dessa classe.

O ajuste do modelo tem como resultado a explicação de 51,97% de explicação da variância dos dados, selecionando as 5 variáveis mais importantes para tal, são elas seguindo o grau de importância: Dívida Pública, Salário Mínimo, ICC, Índice de Gini e Taxa Selic. A figura a seguir mostra a contribuição de cada variável para redução do erro quadrático médio.

**Gráfico 5 – % de contribuição da variável na redução do erro.**



Fonte: Elaboração própria a partir dos resultados obtidos no *software R* (2022).



Quanto ao relacionamento das variáveis os modelos dessa classe não nos permitem interpretação do coeficiente.

Os resultados do modelo, em termos de medidas de acurácia estão postos a seguir:

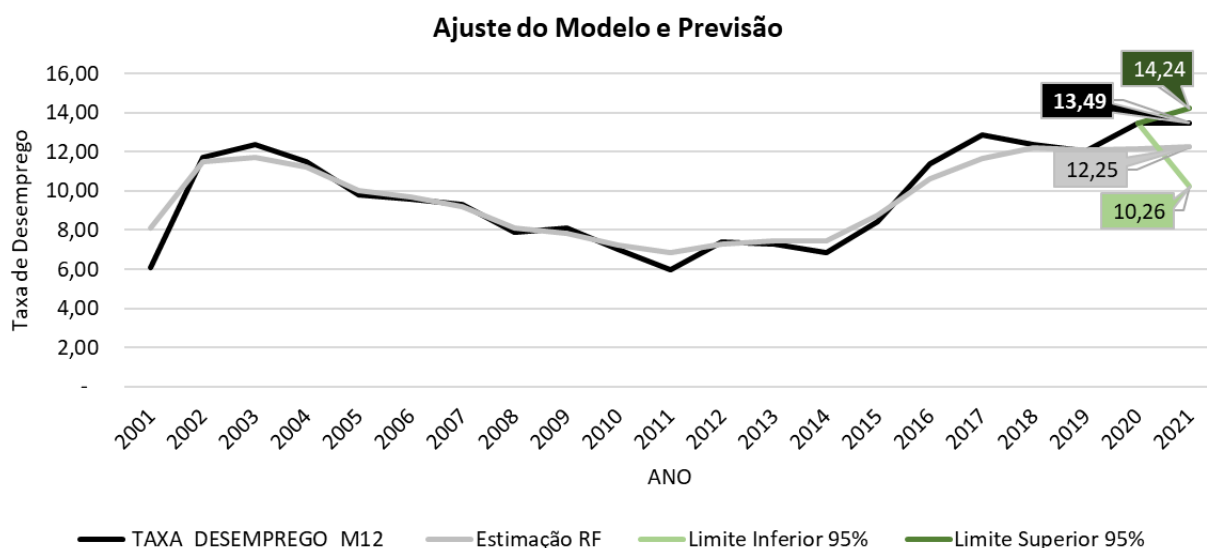
**Tabela 7 – Resultados do modelo Random Forest**

| Amostra         | Erro médio absoluto (MAE) | Raíz do Erro Quadrático Médio (RMSE) | R <sup>2</sup> |
|-----------------|---------------------------|--------------------------------------|----------------|
| Desenvolvimento | 0,4915                    | 0,7042                               | 94,66%         |

Fonte: Elaboração própria a partir dos resultados obtidos no software R (2022).

Como este algoritmo é baseado nas variáveis mencionadas não é possível construir previsões para os próximos anos sem tê-las mapeadas, por isso o gráfico a seguir contém apenas os dados reais.

**Gráfico 6 – Ajuste do modelo Random Forest**



Posteriormente, serão comparados os resultados com os ajustes e previsões expostos até aqui.

*Support Vector Machine:* A modelagem dos dados também partiu do conjunto de dados estipulado no capítulo de seleção de variáveis, considerando todas as variáveis no ano corrente e a taxa de desemprego resultante no ano seguinte, de forma que seja possível realizar a previsão do patamar da taxa de desemprego do ano seguinte. Diante desta proposta, todos os dados foram utilizados para o desenvolvimento.

Para fazer o ajuste do modelo foi utilizada a função svm no *software* R, que tem como objetivo ajustar modelos dessa classe. É importante destacar que esse tipo de ajuste considera todas as variáveis preditoras, sem considerar uma nova seleção, por isso a importância da etapa anterior de seleção de variáveis.

Os resultados do modelo, em termos de medidas de acurácia estão postos a seguir:

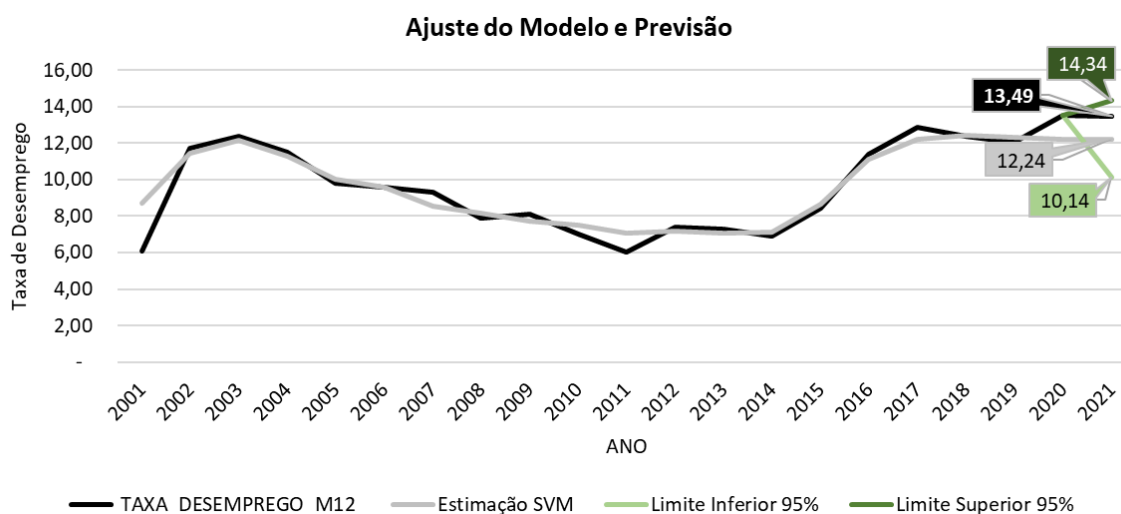
**Tabela 8 – Resultados do modelo SVM**

| Amostra         | Erro médio absoluto (MAE) | Raiz do Erro Quadrático Médio (RMSE) | R <sup>2</sup> |
|-----------------|---------------------------|--------------------------------------|----------------|
| Desenvolvimento | 0,5796                    | 0,7313                               | 91,21%         |

Fonte: Elaboração própria (2022).

Como este algoritmo é baseado nas variáveis mencionadas não é possível construir previsões para os próximos anos sem tê-las mapeadas, por isso o gráfico a seguir contém apenas os dados reais.

**Gráfico 7 – Ajuste do modelo SVM**



Fonte: Dados do IBGE (2021). Elaboração própria a partir dos resultados obtidos no software R (2022).

Posteriormente, serão comparados os resultados com os ajustes e previsões expostos até aqui.

**Lasso Regression:** A modelagem dos dados partiu do conjunto de dados estipulado no capítulo de seleção de variáveis, considerando todas as variáveis no ano corrente e a taxa de desemprego resultante no ano seguinte, de forma que seja possível selecionar as variáveis mais importantes em um ano zero, para previsão do patamar da

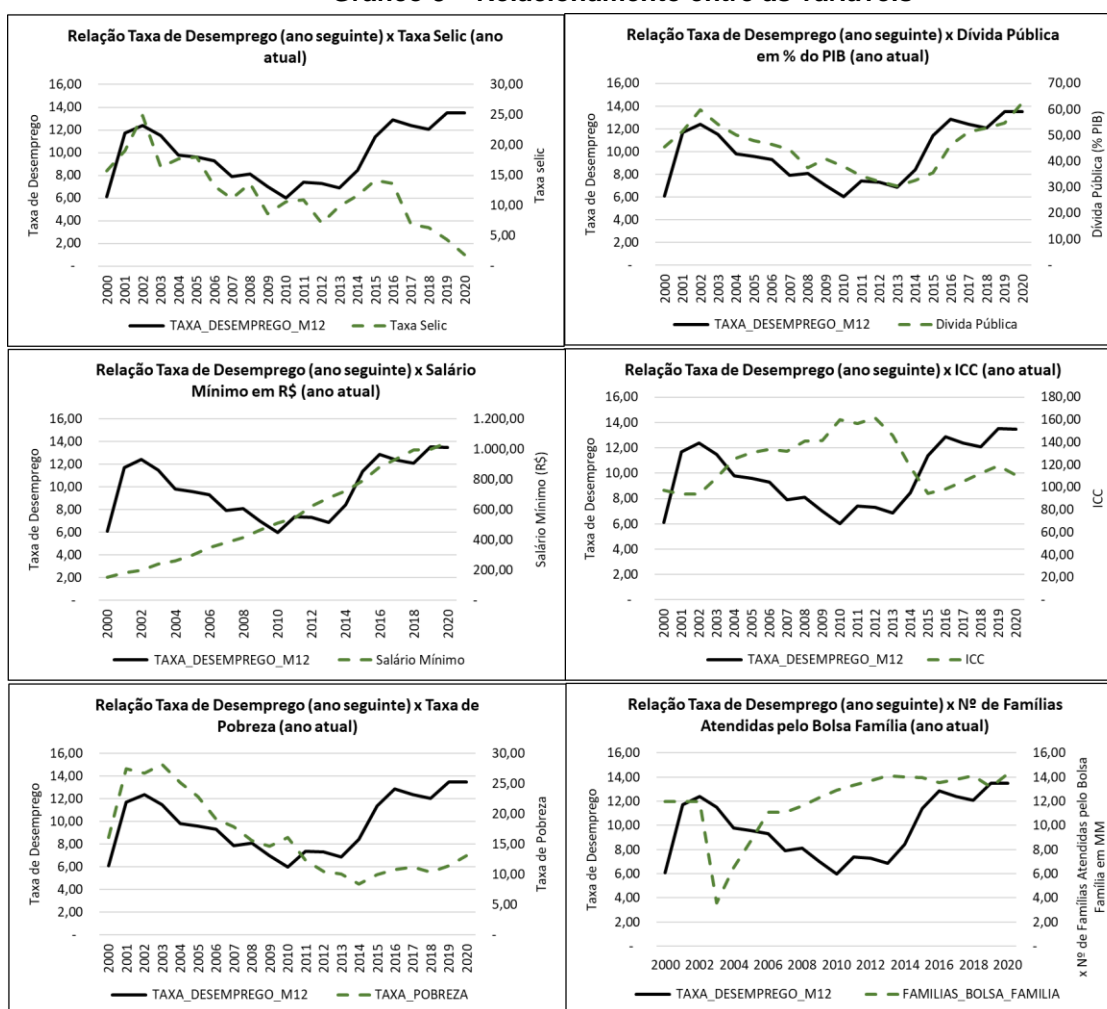
taxa de desemprego do ano seguinte. Diante desta proposta, toda base de dados foi utilizada para fins de desenvolvimento.

Para fazer o ajuste do modelo foi utilizada a função *glmnet* no software R, que tem como objetivo ajustar modelos dessa classe.

O ajuste do modelo selecionou as 6 variáveis mais importantes para a previsão do patamar de taxa de desemprego, são elas seguindo o grau de importância: Taxa Selic, Dívida Pública, ICC, Salário Mínimo, Taxa de Pobreza e Número de Famílias Atendidas pelo Bolsa Família.

Nesta classe de modelos é possível interpretar a relação de cada uma dessas variáveis com a Taxa de Desemprego, observando como elas contribuem para o ajuste do modelo, o que está disposto nos gráficos a seguir.

**Gráfico 8 – Relacionamento entre as variáveis**



Fonte: Dados do IBGE (2021). Elaboração própria a partir dos resultados obtidos no software R (2022).

Segundo Chatterjee e Lahiri (2011) quando as variáveis de erro têm uma média finita, o estimador de Lasso é fortemente consistente, desde que o parâmetro de penalidade seja de ordem menor que o tamanho da amostra, o que é o caso desse resultado. Portanto, abaixo estão dispostas as relações, entretanto é importante trazer a ressalva de que não se trata de uma simples relação de causalidade.

A taxa Selic apresentava semelhante inclinação à taxa de desemprego até 2016, depois desse período passou a apresentar comportamento inverso. Dentro da estimação do modelo, foi considerada como se estivesse sempre andando no mesmo sentido, isto é, um aumento na taxa Selic no ano  $x$  causa um aumento na taxa de desemprego do ano seguinte, permanecendo as demais variáveis constantes.

Os gráficos mostram alguns relacionamentos interessantes, a inclinação da curva de dívida pública em % do PIB anda sempre muito próxima da inclinação da taxa de desemprego. Dessa forma, se em um dado ano  $x$ , há um aumento de dívida pública em relação ao PIB, pode-se esperar um aumento da taxa de desemprego no ano seguinte, permanecendo as demais variáveis constantes. Vale também destacar que a variável de Dívida Pública está sendo utilizada aqui sob o PIB, então não necessariamente se tratam de aumentos de dívida, em alguns anos o que ocorreu foi a queda do PIB, fazendo com que a relação apresente crescimento.

A partir de 2012 os aumentos de salário mínimo acompanham uma elevação da taxa de desemprego, mas este não era o comportamento no período anterior, o que faz com que se reflita se pode haver mais informações implícitas nessa variável. É interessante retomar também as divergências que foram trazidas no capítulo de perspectivas teóricas, em que alguns autores indicavam que o salário mínimo não era um fator determinante e neste trabalho ele apareceu como determinante na taxa de desemprego.

Quanto ao ICC (indicador de custo de crédito), este apresenta um movimento inverso, isto é, se em um dado ano  $x$  o ICC apresenta um aumento, espera-se que no ano seguinte a taxa de desemprego apresente uma queda, permanecendo as demais variáveis constantes.

A variável de Taxa de Pobreza entrou no modelo com peso positivo, indicando que segue a mesma inclinação que a Taxa de Desemprego, o que é possível observar no

gráfico, mesmo que a partir de 2014 tem tido bem menor volatilidade do que a Taxa de Desemprego.

A variável referente ao Número de Famílias Atendidas pelo Bolsa Família tem peso pequeno no modelo e de cunho negativo, o que indica relação inversa com a Taxa de Desemprego.

Os resultados do modelo, em termos de medidas de acurácia estão postos a seguir:

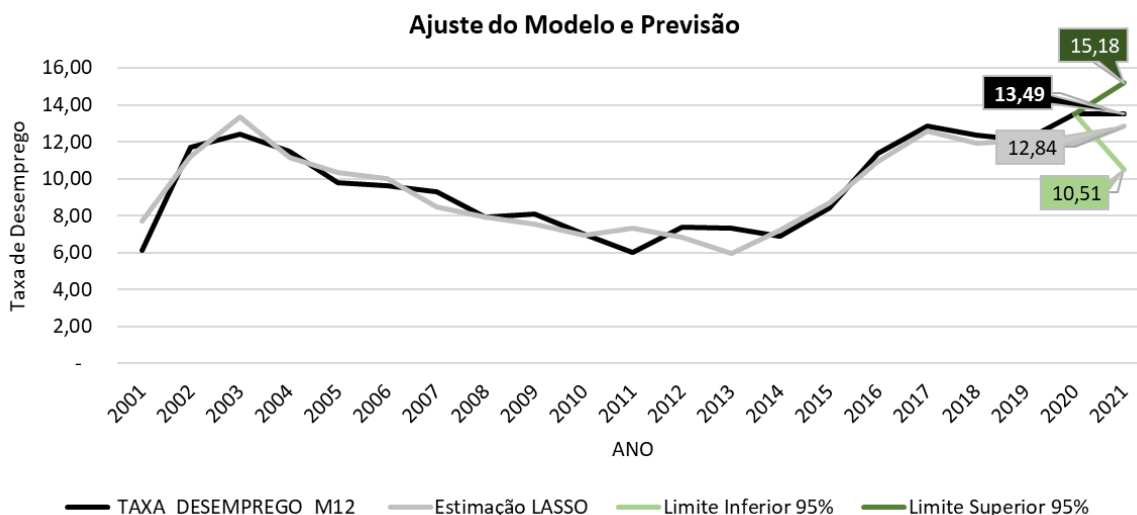
**Tabela 9 – Resultados do modelo LASSO**

| Amostra         | Erro médio absoluto (MAE) | Raíz do Erro Quadrático Médio (RMSE) | R <sup>2</sup> |
|-----------------|---------------------------|--------------------------------------|----------------|
| Desenvolvimento | 0,6248                    | 0,7941                               | 89,19%         |

Fonte: Elaboração própria a partir dos resultados obtidos no software R (2022).

Como este algoritmo é baseado nas variáveis mencionadas não é possível construir previsões para os próximos anos sem tê-las mapeadas, por isso o gráfico a seguir contém apenas os dados reais.

**Gráfico 9 – Ajuste do modelo LASSO**



Fonte: Dados do IBGE (2021). Elaboração própria a partir dos resultados obtidos no software R (2022).

### 4.3 COMPARATIVO

Retomando os resultados de todos os ajustes construídos, através dos modelos clássicos de séries temporais e de alguns algoritmos de *machine learning* é possível

verificar que em termos de ajuste de dados na amostra de desenvolvimento todos os modelos de *machine learning* tem performance superior em termos de erro e aderência dos dados e o ajuste através de *Random Forest* tem resultado superior, considerando essas métricas.

Analisando puramente os indicadores de erro e aderência fica claro que há ganho na utilização de outros dados macroeconômicos para previsão da taxa de desemprego.

**Tabela 10 – Comparativos dos ajustes dos modelos**

| Técnica | Erro médio absoluto (MAE) | Raíz do Erro Quadrático Médio (RMSE) | R <sup>2</sup> |
|---------|---------------------------|--------------------------------------|----------------|
| AR      | 1,2278                    | 1,6003                               | 58,44%         |
| RF      | 0,4915                    | 0,7042                               | 94,66%         |
| SVM     | 0,5796                    | 0,7313                               | 91,21%         |
| LASSO   | 0,6248                    | 0,7941                               | 89,19%         |

Fonte: Elaboração própria a partir dos resultados obtidos no software R (2022).

Sobre as variáveis macroeconômicas que foram relevantes para previsão e considerando o ajuste feito através do Lasso, que permite interpretação, é importante retomar a expectativa que se tinha delas e como elas se comportaram na modelagem.

A respeito da variável de dívida pública em relação ao PIB, esperava-se o relacionamento encontrado, de correlação positiva com a taxa de desemprego. Entretanto é importante reforçar que essa é uma variável ampla, em que o aumento específico da dívida não necessariamente é um provocador do aumento da taxa de desemprego se vier acompanhado de um aumento de PIB, ou a depender também da forma que se compõe. Cabe ainda destacar que a elevação da dívida em relação ao PIB pode indicar um efeito indireto do PIB sobre o emprego, pois além de estar no denominador, a queda da atividade econômica reduz a arrecadação e conseqüentemente afeta o endividamento. Ou seja, dado que há uma certa defasagem entre a queda da atividade econômica e o emprego, a relação dívida/PIB pode captar melhor esse efeito.

Quanto ao salário mínimo, é válido salientar a relação positiva, conforme esperado, que poderia ser melhor aprofundada investigando o crescimento real do salário mínimo anualmente, neste trabalho, foi considerada a informação nominal.

A variável ICC (indicador de custo de crédito), também havia expectativa de correlação positiva, entretanto a relação é negativa, sinalizando que se há um aumento no ICC no ano zero, mantendo o restante constante, há uma redução na taxa de

desemprego. Esse é o resultado de mais difícil elucidação, pois o encarecimento do crédito deveria desestimular o consumo e investimento. Como possível explicação, pode-se entender que o encarecimento do crédito seja reflexo de um aumento na demanda por crédito e que seu aumento seria, portanto, um reflexo da maior dinâmica econômica.

A taxa Selic foi outra variável em que havia a expectativa de que a relação fosse positiva, e foi assim que a variável entrou nos modelos. Entretanto, é importante observar que a partir de 2016 essa relação teve comportamento inverso, período em que houve início as menores taxas já aplicadas através dessa política monetária.

A taxa de pobreza também era considerada com expectativa de relacionamento positivo, e foi assim que entrou no modelo. Mesmo que a taxa de pobreza seja também uma consequência da situação da taxa de desemprego, no período considerado, o modelo estima que um aumento na taxa de pobreza no ano zero impacta em um aumento na taxa de desemprego no ano seguinte.

E, por último, considerando a variável Número de Famílias Atendidas pelo Bolsa Família, havia expectativa de relação inversa que se concretizou.

No geral, as variáveis macroeconômicas se comportaram dentro do esperado teórico dentro das modelagens realizadas.

## 5 CONCLUSÃO

O objetivo deste trabalho era mapear as variáveis macroeconômicas mais relevantes para previsão do patamar da taxa de desemprego no Brasil. Como mencionado, a situação neste ano de 2022, com a inflação alta e juros alto preocupa todos os âmbitos do cenário econômico, tornando-o cada vez mais desafiador. Dos diversos componentes que são foco de estudo e atenção em momentos como esse há o componente de taxa de desemprego, que é uma das principais variáveis econômicas e sociais, uma vez que reflete diretamente em todos os pontos citados acima, em âmbitos de políticas públicas e redistribuição de renda.

Para atender este objetivo, foram mapeadas variáveis macroeconômicas no período de 2000 a 2021, observadas as correlações e explicações correlatas que elas forneciam e depois construídos modelos de séries temporais (sem o uso das variáveis macroeconômicas) e modelos de *machine learning* (fazendo uso dessas variáveis).

Por meio dos critérios estabelecidos para as análises, considera-se que o objetivo foi atendido de forma satisfatória, uma vez que todos os modelos ajustados apresentam intervalos (patamares) que contêm a taxa de desemprego realizada no ano de 2021.

No que diz respeito aos resultados, pode-se dizer que as variáveis mais relevantes para previsão do patamar da taxa de desemprego, em ordem de relevância, considerados os modelos de *machine learning*, são: Dívida Pública, Salário Mínimo, ICC, Taxa Selic, Índice de Gini, Taxa de Pobreza e Número de Famílias Atendidas pelo Bolsa Família.

Em relação a melhor técnica, considerando modelos de séries temporais e modelos de *machine learning*, há prós e contras para se levar em consideração. Como os resultados mostraram, os modelos de *machine learning* apresentam menor erro e também menor intervalo de confiança, entretanto é necessário mapear uma gama de variáveis para construir a previsão. Os modelos de séries temporais, embora apresentem maior erro, são mais fáceis de construir, tem previsões em níveis satisfatórios e permitem previsões de mais anos à frente. Dessa forma, a escolha da técnica ideal vai depender do objetivo e do tempo que o pesquisador tem disponível.

Quanto as variáveis que compuseram o modelo construído através do algoritmo Lasso, que permite interpretação, e seus respectivos relacionamentos com a taxa de



desemprego, o comportamento foi dentro do esperado para a maioria delas. Segundo os resultados construídos, as variáveis de Dívida Pública, Salário Mínimo, Taxa Selic, e Taxa de Pobreza, apresentam a mesma direção da taxa de desemprego, isto é, enquanto uma cresce em um ano zero a taxa de desemprego cresce no ano seguinte, mantendo o restante constante. Quanto ao ICC e Número de Famílias Atendidas pelo Bolsa Família, a relação é inversa, quando essas variáveis crescem em um ano zero, a taxa de desemprego decresce no ano seguinte, mantendo o restante constante. Desses resultados, cabe uma maior reflexão futura para compreensão da relação inversa entre ICC e a taxa de desemprego, bem como novas simulações e testes econométricos para verificar se essa relação se mantém.

Outros aspectos relevantes para trabalhos futuros é a revisão das variáveis macroeconômicas, verificando se não há mais variáveis que possam ser consideradas, ou explorando as relações intrínsecas das variáveis já consideradas com outros fatores macroeconômicos. Uma melhoria adicional que pode ser considerada é analisar apenas variáveis que tenham disponibilidade de dados trimestrais, permitindo aumento de dados. Analisando as técnicas de séries temporais, também é interessante testar outros algoritmos como os utilizados por Weber (2016) em termos de modelo sazonal autorregressivo e de médias móveis (SARIMA), modelo vetorial autorregressivo (VAR) e modelo de correção de erros (VEC).

## REFERÊNCIAS

- ARAÚJO, João; ANTIGO, Mariangela. Desemprego e qualificação da mão de obra no Brasil. **Revista de economia contemporânea**. Rio de Janeiro, v. 20, n. 2, p. 308-335, maio-ago 2016.
- AUSTIN RATING. **Agência classificadora de risco**. São Paulo, 2021. Disponível em: <https://www.austin.com.br/Utimos-Ratings.html>. Acesso em: 06 ago. 2022.
- BANCO CENTRAL DO BRASIL. **Taxas de juros básicas – Histórico**. 2022. Disponível em: <https://www.bcb.gov.br/controleinflacao/historicotaxasjuros>. Acesso em: 6 ago. 2022.
- BARROS, R. P.; CAMARGO, J. M.; MENDONÇA, R. S. P. **A estrutura do Desemprego no Brasil**. Rio de Janeiro: IPEA, 1997. (Texto para discussão, 478).
- BREIMAN, L. (2001). **Random forests, Machine Learning**. [S.l.], v. 45, n. 1, p. 5–32. Disponível em: <https://doi.org/10.1023/A:1010933404324> Acesso em: 06 ago. 2022.
- CADIÑANOS, Leonardo. **Desemprego entre jovens de 1991 a 2014: uma investigação com base nos dados do Brasil e OCDE**. Trabalho de conclusão de curso (Graduação em Ciências Econômicas) – Faculdade de Ciências Econômicas, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2019.
- CHATTERJEE, A; LAHIRI, S. Strong Consistency of Lasso Estimators. **The Indian Journal of Statistics**, Sankhyā, v. 73-A, n. 1, p. 55-78, 2011.
- CHOI. S. OH, J.; CHOI, C.; KIM, C. Input Variable Selection for Feature Extraction in Classification Problems. **Signal Processing**. [S.l.], v. 92, n. 3, p. 636-648, mar 2004.
- CHRISTO, Dirce. **Taxa de Sacrifício da Economia Brasileira: Uma Discussão sobre a Premissa de Linearidade da Curva de Phillips**. Trabalho de conclusão de curso (Graduação em Ciências Econômicas) – Faculdade de Ciências Econômicas, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2009.
- CORREA, Humberto. **Evolução do desemprego no Brasil de 1995 a 2010: análise dos governos FHC e Lula**. Trabalho de conclusão de curso (Graduação em Ciências Econômicas) – Faculdade de Ciências Econômicas, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2013.
- COUNTRY ECONOMY. **Índice de desenvolvimento Humano**. 2022. Disponível em: <https://pt.countryeconomy.com/demografia/idh>. Acesso em: 6 ago. 2022.
- CUNHA, Denis; ARAÚJO, Araci; LIMA, João. Determinantes do desemprego e inatividade de jovens no Brasil metropolitano. **Revista de economia e agronegócio**, Viçosa, v. 9, n. 3, p. 369-392, 2011.

DOWNIE, N. M.; HEATH, R. W. **Basic statistical methods**. New York, Harper & Brothers, p. 289, 1959.

ESTADÃO. **Como investimentos em educação potencializam o mercado de trabalho**. 2021. Disponível em: <https://einvestidor.estadao.com.br/colunas/marcelo-biasoli/investimentos-educacao-potencializam-mercado-trabalho>. Acesso em: 6 ago. 2022.

FOLHA DE SÃO PAULO. **Taxa de desemprego entre os mais pobres é de 36%, aponta estudo**. 2021. Disponível em: <https://www1.folha.uol.com.br/mercado/2021/09/taxa-de-desemprego-entre-os-mais-pobres-e-de-36-aponta-estudo.shtml#:~:text=Taxa%20de%20desemprego%20entre%20os,09%2F2021%20%2D%20Mercado%20%2D%20Folha>. Acesso em: 6 ago. 2022.

FRAGA, Gilberto; DIAS, Joilson. Taxa de desemprego e a escolaridade dos desempregados nos Estados brasileiros: estimativas dinâmicas de dados em painéis. **Revista de economia aplicada**, São Paulo, v. 11, n. 3, p. 407-424, jul-set 2007.

FURTADO, Adolfo C.A.R. **Estudo sobre o desemprego**. Brasília, DF: Biblioteca digital da câmara dos deputados, 1997. Disponível em: <https://bd.camara.leg.br/bd/handle/bdcamara/29579>. Acesso em: 06 ago. 2022.

GÉRON, Aurélien. **Hands-on Machine Learning with Scikit-Learn, Keras & TensorFlow**. Sebastopol: O'Reilly Media 2. ed., 2019.

GOV.BR – MINISTÉRIO DA ECONOMIA. **Balança Comercial - Dados consolidados**. Brasília, 2022. Disponível em: <https://www.gov.br/produtividade-e-comercio-exterior/pt-br/assuntos/comercio-exterior/estatisticas/balanca-comercial-brasileira-acumulado-do-ano>. Acesso em: 6 ago. 2022.

GRANGER, C.W.J.; NEWBOLD, P. (1977) **Forecasting Economic Time Series**. Academic Press, New York: Elsevier, 1997.

HONGYU, Kuang, SANDANIELO; Vera, JUNIOR; Gilmar. Análise de Componentes Principais: resumo teórico, aplicação e interpretação. **Engineering and Science**, v1, n.5, p.1-8, 2015.

HOTELLING, H. Analysis of a complex of statistical variables into principal components. **The Journal Educational Psychology**, Cambridge, v.24, p.498-520, 1933.

HOTELLING, H. Simplified calculation of principal components. **Psychometrika, Williamsburg**, v.1, p.27-35, 1936.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE. **PNAD Contínua - Pesquisa Nacional por Amostra de Domicílios Contínua**. Rio de Janeiro, 2022. Disponível em: <https://www.ibge.gov.br/estatisticas/sociais/trabalho/9173-pesquisa-nacional-por-amostra-de-domicilios-continua-trimestral.html?=&t=series->

historicas&utm\_source=landing&utm\_medium=explica&utm\_campaign=desemprego. Acesso em: 6 ago. 2022.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE. **Rendimento mensal e índice de Gini**. Rio de Janeiro, 2015. Disponível em: [https://seriesestatisticas.ibge.gov.br/lista\\_tema.aspx?de=16&no=6&op=0](https://seriesestatisticas.ibge.gov.br/lista_tema.aspx?de=16&no=6&op=0). Acesso em: 6 ago. 2022.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE. **SCNT - Sistema de Contas Nacionais Trimestrais**. Rio de Janeiro, 2022. Disponível em: [https://www.ibge.gov.br/estatisticas/economicas/contas-nacionais/9300-contas-nacionais-trimestrais.html?=&t=series-historicas&utm\\_source=landing&utm\\_medium=explica&utm\\_campaign=pib#evolucao-taxa](https://www.ibge.gov.br/estatisticas/economicas/contas-nacionais/9300-contas-nacionais-trimestrais.html?=&t=series-historicas&utm_source=landing&utm_medium=explica&utm_campaign=pib#evolucao-taxa). Acesso em: 6 ago. 2022.

INSTITUTO DE PESQUISA ECONOMICA APLICADA – IPEA. **Percepção e expectativa**. Brasília, 2012. Disponível em: <http://ipeadata.gov.br/beta3/#!/lista-de-serie?ascOrder=desc&base=&busca=&columnOrdering=SERATUALIZACAO&fonte=&skip=0&tema=Percep%C3%A7%C3%A3o%20e%20expectativa&territoriality=>. Acesso em: 6 ago. 2022.

INSTITUTO DE PESQUISA ECONOMICA APLICADA – IPEA. **Pobreza - taxa de pobreza (P0)**. Brasília, 2014. Disponível em: <http://www.ipeadata.gov.br/ExibeSerie.aspx?serid=38590&module=M>. Acesso em: 6 ago. 2022.

INSTITUTO DE PESQUISA ECONOMICA APLICADA – IPEA. **Salário Mínimo Real**. Brasília, 2022. Disponível em: <http://ipeadata.gov.br/ExibeSerie.aspx?serid=37667&module=M>. Acesso em: 6 ago. 2022.

INSTITUTO DE PESQUISA ECONOMICA APLICADA – IPEA. **Séries de Estatísticas Conjunturais**. Brasília, 2022. Disponível em: <https://www.ipea.gov.br/cartadeconjuntura/index.php/series-estatisticas-conjunturais-2/>. Acesso em: 6 ago. 2022.

INSTITUTO DE PESQUISA ECONOMICA APLICADA – IPEA. **Taxa de câmbio comercial para compra: real (R\$) / dólar americano (US\$) - média**. Brasília, 2022. Disponível em: <http://www.ipeadata.gov.br/ExibeSerie.aspx?serid=38590&module=M>. Acesso em: 6 ago. 2022.

KLÜCK, Mariza; FACHEL, Jandira; NUNES, Luciana. Uso da imputação múltipla de dados faltantes: uma simulação utilizando dados epidemiológicos. **Cad. Saúde Pública**, Rio de Janeiro, v25, n.2, p.268-278, 2009.

LANTZ, Brett. **Machine Learning with R**. Birmingham: Pack Publishing, 2. ed., 2015.

LUCAS, Kézia. **O impacto do salário mínimo sobre o desemprego no Brasil.**

Trabalho de dissertação de mestrado (Programa de mestrado em Ciências Econômicas) – Universidade Estadual de Maringá, Maringá, 2006.

MARSTON, S.T. Two views of the geographic distribution of unemployment. **Quarterly Journal of Economics**, v.10, n.1, p. 57-79, 1985.

MATTOS, R. A. de. Os Sentidos da integralidade: algumas reflexões acerca dos valores que merecem ser defendidos. In: PINHEIRO, R; MATTOS, R. A. de (Org.). **Os sentidos da integralidade**. Rio de Janeiro: IMS/UERJ/ABRASCO, p. 39-64, 2001.

MORETTIN, P; TOLOI, C. **Análise de séries temporais**. 2 ed. São Paulo: Edgar Blucher, 474p, 2006.

NASCIMENTO, R. et al.. O algoritmo Support Vector Machines (SVM): avaliação da separação ótima de classes em imagens CCD-CBERS-2. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 14., 2009, Natal. **Anais [...]**. São Paulo: INPE, p. 2079-2086. 2009.

ORAIAR, Rodrigo; SIQUEIRA, Fernando. Investimento público no Brasil e suas relações com ciclo econômico e regime fiscal. **Economia e Sociedade**, Campinas, v. 27, n. 3, p. 939-969, 2019. Disponível em: <https://periodicos.sbu.unicamp.br/ojs/index.php/ecos/article/view/8657079>. Acesso em: 7 ago. 2022.

PORTAL DA TRANSPARENCIA. **Detalhamento de Benefícios Bolsa Família**. Brasília, 2021. Disponível <https://www.portaltransparencia.gov.br/beneficios/bolsa-familia?ordenarPor=beneficiario&direcao=asc>. Acesso em: 6 ago. 2022.

PRONI, Marcelo W. **Teorias do desemprego: um guia de estudo**. Texto para Discussão. Campinas, SP: IE/Unicamp, 2015. (Texto para discussão, 256).

TEIXEIRA, Evandro; COSTA, Jaqueline. Desemprego no Brasil no período de 1995-2002: uma análise dos seus determinantes, enfatizando a discriminação segundo o gênero. **Cadernos de Economia - Curso de Ciências Econômicas – Unochapecó**. Chapecó, v. 12, n. 23, p. 102-123, jul-dez 2008.

TESOURO NACIONAL TRANSPARENTE. **Despesas da União - Séries Históricas**. Brasília, 2022. Disponível <https://www.tesourotransparente.gov.br/publicacoes/despesas-da-uniao-series-historicas/2019/11>. Acesso em: 6 ago. 2022.

VIRGOLINO, Rodrigo. **Análise multivariada de características clínicas de PET/MAH e níveis de expressão gênica e derivação de modelos de predição diagnóstica em pacientes infectados com HTLV-1**. Dissertação de Mestrado (Mestrado em Doenças Tropicais) – Programa de pós graduação em doenças tropicais, Universidade Federal do Pará, Belém, 2017.

WEBER, Camila. **Previsão de desemprego em regiões Metropolitanas do Brasil**: um exercício Empírico usando modelos sarima, var e vec. Trabalho de conclusão de curso (Graduação em Estatística) – Instituto de matemática e estatística, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2016.

WORLDWIDE INFLATION DATA. **Inflação histórica Brasil – IPC**. 2022. Disponível <https://www.inflation.eu/pt/taxas-de-inflacao/brasil/inflacao-historica/ipc-inflacao-brasil.aspx>. Acesso em: 6 ago. 2022.