

UNIVERSIDADE FEDERAL DE JUIZ DE FORA
FACULDADE DE ECONOMIA

WILLIAN ROCHA NEVES

**INTELIGÊNCIA ARTIFICIAL E MERCADO DE TRABALHO:
UMA MUDANÇA PERPÉTUA?**

JUIZ DE FORA – MG
2020

WILLIAN ROCHA NEVES

**INTELIGÊNCIA ARTIFICIAL E MERCADO DE TRABALHO:
UMA MUDANÇA PERPÉTUA?**

Trabalho monográfico apresentado ao curso de Ciências Econômicas da Universidade Federal de Juiz de Fora, como requisito parcial à obtenção do título de bacharel em Ciências Econômicas.

Orientador: Prof. Rogério Silva de Mattos

JUIZ DE FORA - MG
2020

Ficha catalográfica elaborada através do programa de geração automática da Biblioteca Universitária da UFJF, com os dados fornecidos pelo(a) autor(a)

Neves, Willian Rocha.

Inteligência artificial e mercado de trabalho: Uma mudança perpétua?

/ Willian Rocha Neves. -- 2021.

59 p. : il.

Orientador: Rogério Silva de Mattos

Trabalho de Conclusão de Curso (graduação) – Universidade Federal de Juiz de Fora, Faculdade de Economia, 2021.

1. Mercado de trabalho. 2. Inteligência artificial. 3. Previsão de salários.
4. Redes neurais artificiais. 5. Tecnologia. I. Mattos, Rogério Silva de, orient.
II. Título.

AGRADECIMENTOS

Primeiramente, agradeço a Deus por ter me presenteado com a dádiva da vida. Também agradeço à toda a minha família que me ajudou durante esta caminhada, em especial, minha mãe Marli, meu pai Reginaldo †, meu padrasto Maurício, Dona Gilda e Sr. Manoel †.

Aos meus grandes amigos que sempre estiveram comigo e me apoiaram também agradeço.

À UFJF e à Faculdade de Economia, especialmente aos economistas do Econs e os professores, que desde o primeiro dia ofereceram toda estrutura necessária para eu me tornar um bom economista.

E um agradecimento especial ao Professor Doutor Rogério Silva de Mattos por ter me apresentado esta inspiradora área de estudo, também agradeço pelo conhecimento compartilhado comigo no desenvolvimento desse trabalho e pela dedicação durante a orientação.

RESUMO

O presente estudo busca analisar a maneira que tecnologia se relaciona com o mercado de trabalho norte americano. São dois os objetivos principais: o primeiro é estudar como as inovações tecnológicas, mais especificamente a inteligência artificial, afetam os salários dos trabalhadores nos EUA. O segundo é realizar previsões com modelos SARIMAS e redes neurais para fazer uma projeção dos salários para os próximos três anos. Diante das mudanças nas relações sociais e produtivas provocadas pela tecnologia que foram observadas nas últimas décadas, juntamente com o papel fundamental da renda em um sistema econômico, esse estudo torna-se de grande importância. Para verificar tal relação, foi utilizado referencial teórico que categoriza os empregos em tipos de tarefas não rotineiras ou rotineiras e cognitivas ou manuais. Com base nesta classificação foi construída uma base de dados e foram adotados métodos econométricos e de machine learning no desenvolvimento do estudo, a fim de captar a trajetória das categorias dos salários americanos. Nota-se que o comportamento para o valor absoluto dos salários difere-se do comportamento observado para o valor dos salários como proporção do PIB. Essas diferenças são claras e representam importantes aspectos do fenômeno do desemprego tecnológico. De forma geral, destaca-se o bom desempenho preditivo das redes neurais MLP e a trajetória crescente de cada categoria de salários para o curto prazo, considerando os efeitos da tecnologia.

Palavras-chave: Inteligência Artificial, Redes Neurais, Salários, Tipos de Tarefa, Previsão, Tecnologia

ABSTRACT

This study seeks to analyze how technology relates to the United States labor market. The main objectives are twofold: the first is to study how technological innovations, more specifically artificial intelligence, affect workers' wages in the United States. The second is to perform with SARIMAS models and neural networks to make a projection of wages for the next three years. In view of the changes in social and productive relations brought about by technology that have been observed in recent decades, together with the fundamental role of income in an economic system, this study is of great importance. To verify this relationship, a theoretical framework was used that categorizes jobs into types of non-routine or routine and cognitive or manual tasks. Based on this classification, a database was built and econometric and machine learning methods were adopted in the development of the study, in order to capture the trajectory of the categories of American wages. It can be noted that the behavior for the absolute value of wages differs from the behavior observed for the value of wages as a proportion of GDP. These differences are clear and represent important aspects of the technological unemployment phenomenon. In general, the good predictive performance of the MLP neural networks stands out and the growing trajectory of each salary category for the short term, considering the effects of technology.

Key-words: Artificial Intelligence, Neural Networks, Salaries, Task categories, Forecasting, Technology

LISTA DE SIGLAS

IA – Inteligência artificial

RNA – *Rede Neural Artificial*

FRED - *Federal Reserve Economic Data*

BLS - *Bureau of Labor Statistics Latest Numbers*

SARIMA – *Seasonal Autoregressive Integrated Moving Average*

MLP – *Multilayer Perceptron*

LSTM – *Long-Short Term Memory*

CPS - *Current Population Survey*

RC – Rotineiras cognitivas

NRC – Não-rotineiras cognitivas

RM – Rotineiras Manuais

NRM - Não-rotineiras Manuais

SVM – *Support Vector Machine*

PLA – *Perceptron Learning Algorithm*

RNR – Rede neural recorrente

ARMA - *Autoregressive Moving Average*

MQO – Mínimos quadrados ordinários

LOO – *Leave-one-out*

TPU - *Tensor processing unit*

MSE - *Mean-square error*

Mape - *Mean Absolute Percentage*

RMSE - *Root-mean-square error*

LISTA DE FIGURAS

QUADRO 1: Capacitações humanas por tipos de tarefa.....	12
GRÁFICO 1: Evolução dos salários nos EUA por tipo de ocupação.....	15
FIGURA 1: A estrutura de um <i>Perceptron</i>	19
FIGURA 2: Aplicação do algoritmo PLA para encontrar hiperplanos.....	20
FIGURA 3: A estrutura de uma rede MLP.....	22
FIGURA 4: Representação gráfica da função linear.	23
FIGURA 5: Representação gráfica da função sigmoide.	24
FIGURA 6: Representação da recorrência	25
FIGURA 7: Módulo de uma LSTM.....	27
FIGURA 8: Mecânica de um módulo de LSTM.	28
FIGURA 9: Processamento final em um módulo.....	31
GRÁFICO 2: Salário Médio Semanal (em dólares) - EUA.....	39
GRÁFICO 3: Razão (Média Salarial Semanal)/PIB.....	40
QUADRO 2: SARIMAS testados para a série NRC.....	41
QUADRO 3: RNAs testadas para a série NRC.....	42
FIGURA 10 : Rede MLP de 150 Neurônios na Camada Interior.....	43
GRÁFICO 4: Previsão salários – NRC com rede MLP.....	44
QUADRO 4: SARIMAS testados para a série RC.....	44
QUADRO 5: RNAs testadas para a série RC.....	45
GRÁFICO 5: Erro da rede LSTM durante o treinamento.....	46
GRÁFICO 6: Previsão salários – RC com rede LSTM.....	47
QUADRO 6: SARIMAS testados para a série NRM.....	47
QUADRO 7: RNAs testadas para a série NRM.....	48
FIGURA 11: Rede MLP de 150 neurônios na camada interior.....	49
GRÁFICO 7: Previsão salários – NRM com rede MLP.....	50
QUADRO 8: SARIMAS testados para a série RM.....	50
QUADRO 9: RNAs testadas para a série RM.....	51
FIGURA 12: Rede MLP com 15 Neurônios de Processamento e 4 de Entrada.....	52
GRÁFICO 8: Previsão salários - RM com rede MLP.....	53
QUADRO 10: Previsões dos Melhores Modelos.....	54

SUMÁRIO

1	INTRODUÇÃO	5
2	REVISÃO DE LITERATURA E REFERENCIAL EMPÍRICO	10
3	METODOLOGIA	15
3.1	Base de dados.....	15
3.2	Redes Neurais Artificiais.....	16
3.3	Modelos SARIMA	31
3.4	Validação Cruzada	33
3.5	Medidas de desempenho.....	35
4	RESULTADOS.....	36
4.1	Análise descritiva.....	37
4.2	Análise Econométrica e de Machine Learning	40
4.2.1	Salário de tarefas Não-Rotineiras Cognitivas.....	40
4.2.2	Salário de tarefas Rotineiras Cognitivas.....	43
4.2.3	Salário de tarefas Não-Rotineiras Manuais	46
4.2.4	Salário de tarefas Rotineiras Manuais	49
4.2.5	Resultados da Modelagem.....	52
5	CONCLUSÃO	54
6	REFERÊNCIAS.....	56

1 INTRODUÇÃO

Desde o começo do desenvolvimento do sistema capitalista no período mercantilista (Séc. XVI a XVIII) até os dias atuais, o mundo passou por importantes revoluções na economia motivadas pela perseguição a menores custos e maior lucro.

Algumas dessas revoluções aconteceram pelo avanço tecnológico que algumas vezes apenas remodelou as relações entre os agentes na sociedade, porém, em outras ocasiões causou uma grande ruptura nas maneiras de produzir, alterando severamente a relação entre diversos grupos na sociedade como, por exemplo, trabalhadores e capitalistas. Com a introdução de novas tecnologias, o trabalho humano que antes era necessário para realizar diversas funções na produção deixa de ser usado em algumas ocupações, pois novas máquinas conseguiram substituí-lo na elaboração de algumas tarefas.

As transformações disruptivas vêm causando grandes impactos no mercado de trabalho. O avanço tecnológico cria novas relações servis que passaram a exigir dos trabalhadores novas habilidades e conhecimentos, direciona os trabalhadores para outros setores da economia, desalenta parte dos trabalhadores e gera consequências para toda a sociedade. Tais transformações se caracterizaram por um conjunto de novos métodos, tecnologias, invenções e novas práticas que redefinem a maneira que o trabalho humano é utilizado na fabricação de bens e na prestação de serviços. E isto é feito através da remodelagem das etapas de produção que visam tornar a produção mais eficiente e lucrativa.

Algumas das grandes transformações vieram através das revoluções industriais. A primeira revolução industrial, que começou na Inglaterra e ocorreu entre 1760 e 1860, teve como principal característica a incorporação do tear mecânico para substituição do trabalho artesanal de tecer. A primeira revolução também foi marcada pela adoção do sistema fabril, a introdução da máquina a vapor, e o uso generalizado de máquinas. Já a segunda revolução industrial (1850 - 1870) foi um processo que se caracterizou pela introdução da energia elétrica na produção industrial, pela massiva utilização do petróleo, do motor a combustão e do aço. Logo após a segunda guerra mundial, em 1940, houve uma terceira revolução industrial - também chamada de revolução informacional - neste momento a eletrônica surge como remodeladora da indústria e esta começa a se desenvolver impulsionada pela tecnologia da informação. As firmas com setores de pesquisa e desenvolvimento passam a aplicar na produção os avanços de áreas como robótica, genética, informática e eletrônica, e isto significa proporcionar uma maior participação de máquinas em tarefas repetitivas que exigiam pouca capacidade analítica.

Como continuação desse longo processo de automatização, alguns autores acreditam que estamos vivenciando a quarta revolução industrial, fase em que as transformações vêm através de nanotecnologia, engenharia genética, neurotecnologias e inteligência artificial. Temos vivenciado novas formas de tecnologia que cada vez mais tornam a capacidade das máquinas mais próximas da capacidade cognitiva humana. Temos, por exemplo, algoritmos de redes neurais *spiking* que são redes neurais artificiais que imitam com mais facilidade as redes neurais humanas quando comparada aos outros tipos de redes neurais artificiais.

Até a terceira revolução industrial as máquinas incorporadas no processo produtivo eram capazes de realizar tarefas rotineiras e manuais. No entanto, nesta possível quarta revolução industrial acontece algo que a humanidade ainda não experimentou: máquinas realizando também as tarefas não rotineiras, isto é, tarefas que exigem uma capacidade cognitiva até então possuída apenas pelo ser humano. A Inteligência artificial (IA) vem sendo usada amplamente na medicina, meteorologia, entretenimento, comunicações, robótica, engenharia e em diversos setores da sociedade.

Com o avanço da ciência, a IA vem se tornando cada vez mais versátil e mais capaz de realizar atividades que outrora eram tidas como exclusivamente humanas. As redes neurais artificiais (RNA) são um dos métodos de inteligência artificial e seu objetivo é imitar a rede neural humana no processamento de informação para que a tomada de decisões seja a melhor possível. A tentativa de aproximar a capacidade humana com uma RNA se materializa na utilização de um modelo matemático que procura replicar o processamento de informações do sistema nervoso humano. Uma importante característica deste tipo de inteligência artificial é a possibilidade de aprendizagem durante todo o processo, o que capacita a máquina para aprender e melhorar com seus erros sempre que é treinada para determinado objetivo.

Devido às poderosas habilidades que as máquinas passaram a possuir, nos dias atuais temos vivenciado a presença cada vez maior da IA em todos os setores da sociedade. Seja nos assistentes virtuais, como a Alexa da *Amazon*, no reconhecimento facial dos celulares ou nas máquinas da indústria automobilística, esta nova tecnologia está cada vez mais presente em nossa vida. Esta inserção da IA na sociedade pode causar mudanças severas nas relações de trabalho, já que estas novas máquinas não só conseguem fazer muitas coisas que o ser humano faz como também fazem melhor. As máquinas não se cansam, não são emocionalmente instáveis, e não necessitam de contrapartida salarial pelo trabalho, estas são características que podem dar às novas máquinas uma produtividade muito maior do que a humana e um espaço considerável dentro do sistema produtivo. Neste sentido a inteligência artificial pode ser vista

além de uma simples continuação do progresso tecnológico, ela se apresenta como o potencial ápice de toda essa evolução onde a disrupção gerada por ela afetará a dinâmica da sociedade de maneira sem precedentes, e até mesmo pode representar o que James Barrat (2013) chamou de "a invenção final".

Nesse cenário onde uma quarta revolução industrial pode causar mudanças ainda mais severas na sociedade, especialistas tentam entender quais serão as implicações disso nas tarefas, nos salários, nas relações servis e na sociedade como um todo. Na literatura econômica deste campo de estudo, a tentativa de entender qual será o resultado dessa disrupção geralmente se orienta com base nas duas forças advindas desse avanço tecnológico: a força destrutiva e a força de criacionista de empregos. Frey e Osborne (2013) utilizaram em sua pesquisa um universo de 702 ocupações e constataram que 47% delas corriam o risco de desaparecer até 2035 por causa da automação, ou seja, 47% das ocupações poderiam ter todas as tarefas relacionadas à ocupação realizadas exclusivamente por máquinas.

Outros estudos procuraram evidências das implicações da IA na renda dos trabalhadores e os autores obtiveram resultados mistos. Chiacchio et al. (2018) capturaram uma relação não muito expressiva entre variações na exposição de empregos aos robôs e o salário por hora, o que impossibilita concluir que a automação teve um impacto significativo nos salários porque algumas estimativas sugeriram um impacto insignificante da automação nos salários, enquanto outros resultados obtidos apontam para a existência de impactos negativos, isto é, os robôs diminuiriam os salários dos trabalhadores. Ainda assim, os autores acreditam que as estimativas encontradas fornecem sugestiva evidência de um efeito negativo da automação nos salários dos trabalhadores.

Em contraste com estas estimativas, no Pwc Report (2018) pesquisadores utilizando um modelo de equilíbrio geral computável global da economia mundial buscaram compreender até que ponto o impacto de IA e outras tecnologias relacionadas causariam uma substituição de trabalhos formalmente feitos por humanos, e se este efeito seria compensado ou não pela criação de novos postos de trabalho. Foram relatadas evidências sugerindo que haveria um aumento do emprego na China de aproximadamente 12% nas próximas duas décadas¹, indicando que novos postos de trabalho surgiriam pelo fato de a produtividade e renda real aumentarem, além de novos e melhores bens serem ofertados no mercado. Sendo assim haveria um efeito em cadeia compensando a substituição do homem pela máquina em

¹ Os autores focaram nos impactos de longo prazo dos próximos 20 anos. Dado o ano base de 2017, as projeções são dos impactos até 2037.

algumas ocupações - o que os autores chamaram de efeito renda. Contudo, o efeito de criação de novos empregos não seria dominante em todos os setores. A agricultura, por exemplo, teria uma perda de empregos de 11%. No geral, o impacto estimado considerando todos os setores da economia seria positivo.

Logo, a substituição do trabalhador que temos vivenciado atualmente não necessariamente significaria perda de postos de trabalho, podendo o saldo até ser positivo quando há uma realocação no mercado de trabalho. Como destaca Bressen (2015), o uso de computadores está associado com crescimento em empregos que pagam bem e decréscimo em empregos que pagam mal, demonstrando uma realocação dos postos de trabalho e uma exigência do aprendizado de novas habilidades por parte dos trabalhadores para se realocarem no mercado de trabalho.

Juntamente com uma potencial força criacionista de empregos, pode haver uma relação de complementaridade. A automação pode acontecer não sobre a ocupação do indivíduo, mas sobre algumas tarefas de tal ocupação, como por exemplo, tarefas rotineiras e repetitivas. Portanto ainda existiria a necessidade da presença do trabalhador para que todas as tarefas de uma ocupação sejam realizadas, mesmo que o computador automatize algumas.

O presente estudo procura entender as mudanças causadas pela evolução tecnológica nos salários de trabalhadores dos Estados Unidos. Para isso, baseando na literatura sobre o tema, foi analisado o comportamento dos salários americanos ao longo das últimas décadas focando em uma agregação do trabalho com base nas quatro categorias de tarefas mencionadas: Rotineiras, não rotineiras, cognitivas e manuais. Devido a grande complexidade acerca do tema, o enfoque deste estudo é somente na análise e previsão das séries temporais geradas pela agregação citada e, posteriormente, a verificação de possíveis características nas séries que indicam mudança de comportamento dos salários frente às disrupções provocadas pelo avanço tecnológico.

Foram analisados os dados dos Estados Unidos pois sua economia se apresenta em um estágio de desenvolvimento mais avançado do que a maioria das economias do mundo. Além disso, é evidente que o fato da tecnologia estar mais inserida na dinâmica do mercado de trabalho de países mais desenvolvidos faz com que os efeitos do avanço tecnológico sejam mais visíveis e de mais fácil percepção nessas economias. Outro fator que favorece a análise dos Estados Unidos é a disponibilidade de dados que possibilitam a agregação para termos as séries de salários, tais como o sistema FRED do Banco Central de Saint Louis e o *US Bureau of Labor Statistics* (BLS). Por fim, foram feitas análises das séries de salários, modelagem e

previsão com três métodos úteis para a previsão de séries temporais: SARIMA, Rede neural *Perceptron* de multicamadas (MLP) e rede neural *Long Short Term Memory* (LSTM).

Além dessa introdução, o trabalho está estruturado da seguinte forma: A segunda seção aborda a Revisão de Literatura e apresenta o referencial teórico, nesta seção discorro sobre a literatura pesquisada referenciando as principais publicações. Na terceira seção temos uma descrição da metodologia, dos métodos e dados utilizados na análise. Esta parte aborda a teoria das quatro grandes categorias relacionadas às ocupações americanas, também são expostos os modelos e medidas de desempenho utilizadas. A quarta seção apresenta um os resultados que dizem respeito ao efeito do desemprego tecnológico nos salários e também os resultados referentes a previsão de cada modelo em cada série. Por fim, temos na quinta seção uma conclusão com uma discussão sobre os resultados encontrados neste trabalho.

2 REVISÃO DE LITERATURA E REFERENCIAL EMPÍRICO

Autor, Levy, e Murname (2003) buscaram compreender o impacto da rápida adoção de computadores nas atividades exercidas por trabalhadores americanos e na demanda por trabalho. Baseando na literatura sobre TI e mercado de trabalho, eles investigaram as tarefas exercidas pelos computadores e se essas habilidades laborais das máquinas complementam ou substituem as atividades realizadas pelos trabalhadores.

A teoria de Autor, Levy, e Murname (2003) introduziu uma metodologia muito influente para mensurar e analisar mudanças na demanda por qualificação do mercado de trabalho. ALM(2003) propões realizar uma distinção entre competência e tarefas. Segundo esta teoria, uma tarefa é uma unidade de atividade de trabalho que gera um produto. Uma competência é um estoque de capacidade do trabalhador que serve para desempenhar várias tarefas. Considerando esta importante distinção, ALM conseguiu uma visão ampliada de como avanços tecnológicos, mudanças no mercado de trabalho e mudanças nas oportunidades de comércio afetam a divisão de trabalho entre humanos e máquinas, as tarefas exercidas pelos trabalhadores e as competências exigidas no mercado de trabalho.

Os autores utilizaram uma classificação que se baseia em dois critérios:

1. Se a tarefa é rotineira ou não-rotineira.
2. Se a tarefa é manual ou cognitiva.

O primeiro critério diz respeito às regras para realizar a tarefa. **Tarefa rotineira** é aquela em que se conhece explicitamente as regras necessárias para sua execução e **tarefa não-rotineira** é aquela em que não se conhece tais regras ou então tais regras não são compreendidas de forma explícita, apenas de forma tácita (não transmissível diretamente). O segundo critério é sobre as características humanas para realizar as tarefas. **Tarefa manual** depende fundamentalmente das capacidades físicas e motoras do trabalhador para sua execução, enquanto que **tarefa cognitiva** depende de processamento mental, interação humana e, portanto de habilidades e competências cognitivas (conhecimento/nível educacional) do trabalhador.

Esta classificação gera uma matriz de quatro classes de tarefas (QUADRO 1). As conclusões obtidas pelos autores vão de encontro aos estudos empíricos relacionados ao fenômeno da polarização do trabalho.

Eles concluíram que as máquinas poderiam substituir os trabalhadores em trabalhos intensivos em tarefas rotineiras, também destacaram uma relação de complementaridade entre o trabalho humano e o trabalho dos computadores em atividades não rotineiras

manuais. Por fim, salientaram que a tecnologia da informação iria continuar substituindo trabalhadores qualificados em atividades manuais e atividades cognitivas que são rotineiras, pois a existência de regras bem definidas permite aos computadores e algoritmos entender a sequência lógica de instruções para realizar determinada atividade.

QUADRO 1: Capacitações humanas por tipos de tarefa

Tipos de Tarefa	Rotineira	Não rotineira
Manual	Substituição substancial	Substituição limitada ou complementaridade
Cognitiva	Substituição substancial	Fortes complementaridades

Fonte: Adaptado da Tabela 1 - (Autor, Levy e Murnane, 2003).

David Autor publicou vários estudos que relatam as mudanças no mercado de trabalho dos EUA nos últimos 30 anos, em especial mudanças nos empregos e salários. Em Autor (2010), ele mostrou que a economia norte americana aumentou a demanda por de alta e baixa qualificação, enquanto trabalhadores de qualificação média perderam espaço no mercado de trabalho. Este é o fenômeno da polarização do trabalho. Segundo Autor (2010) as possíveis razões para este fenômeno são variadas, mas ele apresenta quatro hipóteses principais:

1. Comércio internacional e terceirização diminuindo a quantidade de empregos nos Estados Unidos.
2. Substituição de tarefas desenvolvidas por humanos pelo uso da tecnologia.
3. Declínio real do salário mínimo norte americano.
4. Redução do poder e penetração dos sindicatos.

Tendo como referência esta literatura, muitos autores investigaram os efeitos do avanço tecnológico no mercado de trabalho. Grande parte dos estudos se concentrou em entender como os salários e postos de trabalho se comportam diante deste fenômeno.

Boehm (2014) investigou os efeitos da polarização do trabalho nos salários de trabalhadores com 27 anos de idade residentes nos Estados Unidos. O autor utilizou os dados da Pesquisa Nacional Longitudinal da Juventude e um modelo de escolha ocupacional para

comparar trabalhadores e alocá-los em diferentes categorias². Seguindo os estudos empíricos sobre polarização do trabalho, Boehm utilizou características dos trabalhadores para classificá-los em ocupações de pouca, média e alta qualificação.

Os resultados obtidos sugerem que trabalhadores com talentos que os tornariam adequados para ocupações de média qualificação experimentaram um declínio substancial em seus salários relativos e possivelmente um declínio em seus salários absolutos. Além disso, segundo o autor, o efeito estimado da polarização do trabalho pode explicar o aumento da desigualdade na metade superior da distribuição salarial nos Estados Unidos, mas não foi possível inferir nada sobre alterações na metade inferior da distribuição salarial norte-americana. Sendo que este último resultado é devido aos dados e premissas assumidas no estudo.

Kantega (2018) utilizou dados transversais dos EUA observados de 1979 a 2010 para investigar os efeitos na estrutura salarial provocados pela demanda de habilidades que foram afetadas pela polarização do trabalho. Os estudos realizados pelo autor sugeriram que a demandas por habilidade dos trabalhadores foram polarizadas nas últimas três décadas, significando uma diminuição na demanda de ocupações de habilidades médias e um aumento na demanda em ocupações cujas habilidades necessárias para realizá-las são altas ou baixas. O autor encontrou que as tendências industriais, o progresso tecnológico e o comércio foram responsáveis por 57% da mudança na demanda por habilidades dos trabalhadores. Além disso, as mudanças nas complementaridades de produção foram cruciais, porém insuficientes para replicar as mudanças observadas na estrutura salarial e ocupacional dos trabalhadores.

²O autor utilizou dados da Pesquisa Longitudinal Nacional da Juventude (NLSY) de 1979 e 1997. Os estudos NLSY fornecem informações detalhadas sobre histórico dos indivíduos, educação e resultados do mercado de trabalho nos Estados Unidos.

Os salários dos diversos tipos de atividades foram agregados com base nos estudos de Autor, Levy, e Murnane (2003) e Acemoglu e Autor (2011). No primeiro estudo foi feita uma divisão entre tarefas rotineiras e não rotineiras conforme a natureza da ocupação e as atividades exercidas nela. Uma ocupação pode ser classificada como rotineira se as atividades relacionadas a ela puderem ser resumidas em um conjunto de atividades específicas a serem realizadas, seguindo instruções e procedimentos bem definidos. Já uma ocupação não rotineira seria uma em que as atividades exercidas exigem criatividade, capacidade de resolução de problemas, flexibilidade e interação humana constante.

Acemoglu e Autor (2011) utilizam desta categorização e classificam as ocupações em: “Cognitiva” ou “Manual” e “Rotineira” ou “Não rotineira”. Na distinção entre trabalhos cognitivos e manuais, se considera o dispêndio físico empregado na atividade. Disto resultam quatro grandes grupos referentes às ocupações compostos pelas mais diversas profissões presentes em uma economia. Uma classificação com base nesta literatura pode ser a seguinte³:

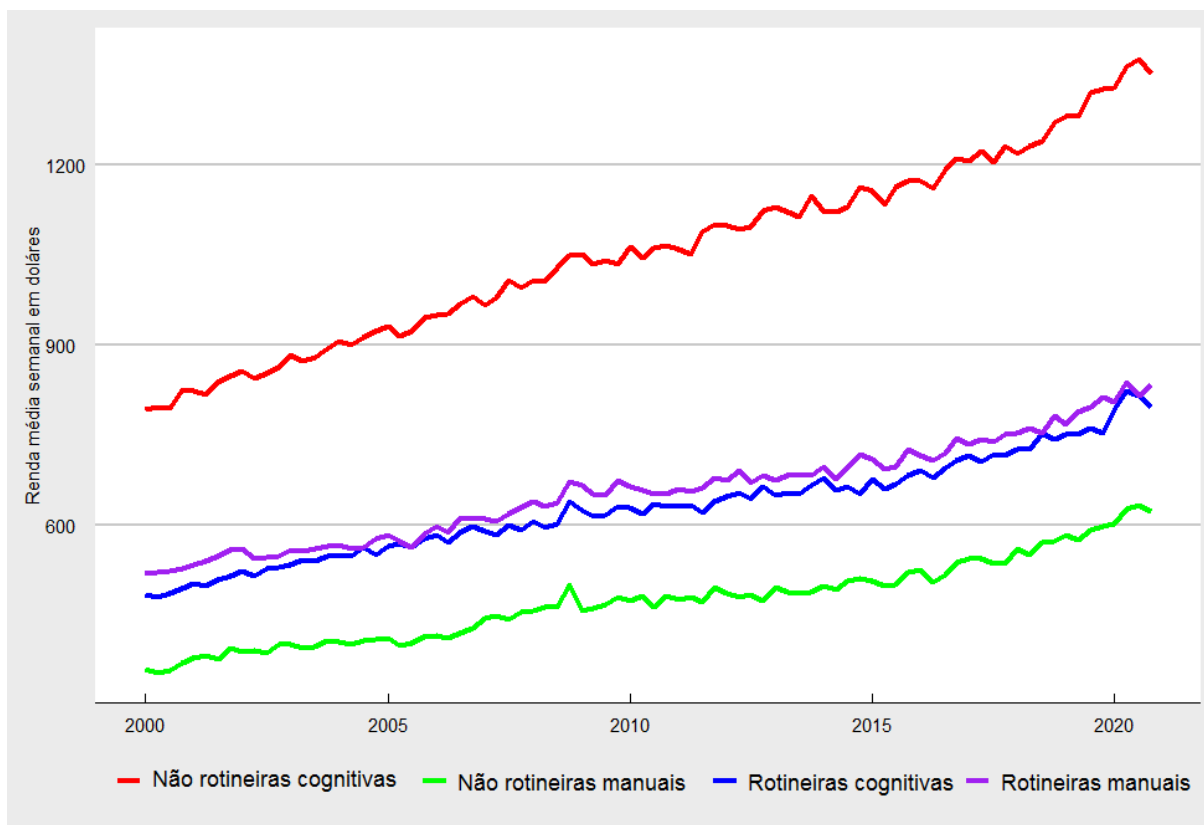
- **Rotineiras manuais:** São trabalhos tais como lenhadores, açougueiros, operador de máquinas, costureiros, mecânicos automotivos e eletricitas. Estes são os chamados trabalhos “*blue collar*”.⁴
- **Não rotineiras manuais:** São os trabalhos que fornecem serviços à sociedade, como policiais, bombeiros, zeladores, jardineiros, auxiliares de atendimento domiciliar e profissionais de cuidados pessoais.
- **Rotineiras cognitivas:** São as atividades relacionadas a vendas, auxílio administrativo e trabalho de escritório; exemplos incluem secretários, atendentes, caixas de banco, vendedores de varejo, corretores de imóveis e agentes de viagens.
- **Não rotineiras cognitivas:** Alguns exemplos são dentistas, gerentes financeiros e administrativos, médicos, engenheiros, analistas financeiros, programadores, geógrafos, agrônomos e economistas.

3 Uma descrição completa dos postos de trabalho contidos na base de dados está disponível no documento de classificação ocupacional do censo de 2010, que pode ser acessado através do endereço eletrônico: <https://www.bls.gov/cps/cenocc2010.pdf>

4 Uma tradução para o português é “colarinho azul”. É uma expressão utilizada nos EUA para identificar trabalhadores que realizam atividades predominantemente manuais.

Assim, agregando os salários dos postos de trabalho conforme a classificação acima obtemos quatro séries temporais que representam o comportamento dos salários nos Estados Unidos entre 2000 e 2020.

GRÁFICO 1: Evolução dos salários reais nos EUA por tipo de ocupação



Os salários nominais foram corrigidos com base no Índice de Preços ao Consumidor dos Estados Unidos. Fonte: Elaboração própria a partir de dados obtidos na pesquisa *Current Population Survey (CPS)* fornecidos pelo *United States Census Bureau*.

Conforme está apresentado na seção sobre a base de dados, os dados são disponibilizados pelo *United States Census Bureau*, mas também podem ser obtidos nos endereços eletrônicos de outras instituições do governo norte americano, como o FRED de Saint Louis⁵.

⁵ *Federal Reserve Economic Data (FRED)* – É um banco de dados de mais de 500.000 séries de dados econômicos de mais de 80 fontes cobrindo questões e informações relacionadas a bancos, negócios, índices de preços ao consumidor e ao produtor, emprego, população, taxas de câmbio, produto interno bruto, taxas de juros, comércio e transações internacionais, e dados financeiros dos EUA. Pode ser acessado em <https://fred.stlouisfed.org/>

3 METODOLOGIA

3.1 Base de dados

O governo americano disponibiliza diversas bases de dados que possibilitam uma análise muito detalhada dos setores de sua economia. O *United States Census Bureau*, órgão principal que coordena a produção de estatísticas, conduz pesquisas que abastecem diversas bases de dados. Uma dessas bases é a *Current Population Survey (CPS)*, que disponibiliza dados sobre 60.000 domicílios americanos⁶. Com a posse desses dados, alguns centros de pesquisa analisam os setores e produzem balanços mensais para o acompanhamento do desempenho da economia. O *U.S. Bureau of Labor Statistics* (órgão estatístico vinculado ao ministério do trabalho estadunidense) é um desses centros, ele mensura a atividade do mercado de trabalho, condições de trabalho, mudanças de preços, e produtividade na economia norte-americana. Os relatórios e bases de dados disponibilizadas por este órgão trazem informações sobre o mercado de trabalho de todos os estados, regiões metropolitanas, condados e cidades. Dentre as informações sobre os salários nacionais estão dados de 800 ocupações, dados sobre 400 indústrias de diversos setores, dados categorizados por características do posto de trabalho e o nível de dificuldade e complexidade necessário para efetuar as tarefas.

De forma agregada, foram utilizados neste estudo os dados do *U.S. Bureau of Labor Statistics* que trazem informações sobre:

- *Production, transportation, and material moving occupations* (Ocupações em produção, transporte e logística).
- *Service occupations* (Ocupações relacionadas a serviços);
- *Sales and office occupations* (Ocupações comerciais e de escritório);
- *Management, professional, and related occupations* (Gestão, profissionais e ocupações relacionadas);
- *Natural resources, construction, and maintenance occupations* (Ocupações relacionadas a recursos naturais, construção e manutenção);

⁶ Uma descrição completa da metodologia utilizada na pesquisa está disponível em <https://www.census.gov/programs-surveys/cps/technical-documentation/methodology.html>

Seguindo a literatura sobre o tema, foi feita a categorização dos salários destes grupos de acordo com a natureza das tarefas:

- **Rotineiras Cognitivas (RC)**
- **Não rotineiras Cognitivas (NRC)**
- **Rotineiras Manuais (RM)**
- **Não rotineiras Manuais (NRM).**

3.2 Redes Neurais Artificiais

Tendo como inspiração o cérebro humano, cientistas desenvolveram máquinas que tentam imitar parte do funcionamento deste mecanismo. Pela alta complexidade do processamento cerebral, por barreiras computacionais e pela falta de desenvolvimento em áreas paralelas, as tentativas de imitar e criar artificialmente as funções do cérebro não foram completamente bem sucedidas, no sentido de que os sistemas de inteligência artificial (AI) que existem atualmente são incompletos.

Uma expressiva evolução no desenvolvimento dessas máquinas veio através das chamadas Redes Neurais Artificiais (RNA), que tentam simular o processo humano do pensar e tomar decisão tendo como base a maneira que o cérebro funciona. Com a ajuda desta importante ferramenta, softwares atualmente podem se tornar mestres em jogos complexos como GO, podem tornar veículos autônomos ou mesmo serem aplicados em áreas mais gerais como medicina, negócios e finanças.

Portugal e Fernandes (1996) analisaram a modelagem de RNA para fazer previsões. Apoiados na característica das RNA's de tentar incorporar algumas qualidades do cérebro humano, eles utilizaram esta modelagem para identificar e aprender as características de uma série temporal, como tendência, ciclo e sazonalidade, possibilitando assim a realização de previsões. Os autores concluíram que as RNA's representam um instrumento indispensável para economistas e profissionais que trabalham criando cenários e fazendo previsões econômicas.

As redes neurais são um dos algoritmos de *machine learning* mais utilizados para resolver problemas computacionalmente complexos, são modelos estatísticos não lineares que podem ser utilizados em problemas de classificação e regressão. Antes de aplicar as Redes Neurais e outros modelos de *machine learning*, temos que treinar os modelos com as informações que temos para que possamos aplica-los em dados novos.

Considerando uma variável y a ser entendida e outras variáveis x_1, x_2, \dots, x_i que podem ser úteis para explicar y , um problema de machine learning é a tentativa de modelar como se dá a

relação entre essas variáveis. Uma representação para esta situação é apresentada pela fórmula $y = f(X)$, onde y = variável resposta, X = vetor de variáveis explicativas e ' f ' é uma função que representa alguma relação desconhecida entre y e X . Os problemas de machine learning normalmente envolvem utilizar n dados previamente coletados para y e X , na forma $(y_i, x_i), i = 1, \dots, n$, visando-se analisar esses dados para estimar a função ' f ', como acontece em problemas de regressão e classificação.

Baseado na estrutura do problema, mais especificamente no tipo da variável dependente, existem duas principais maneiras de treinar um modelo:

- a. **Aprendizado supervisionado** – Neste tipo de treinamento de modelo, uma parte dos dados de treinamento é utilizada para ensinar o algoritmo. Para cada valor assumido pelas variáveis independentes x_i , temos uma resposta y_i associada. O aprendizado do modelo é feito associando cada conjunto de valores das variáveis independentes a um valor da variável de resposta, de modo que o erro que obtemos ao comparar o dado real y_i com a previsão \hat{y}_i é o menor possível. Esse processo de ajuste ou treinamento é realizado utilizando uma fração dos dados que temos sobre as variáveis, no sentido de que uma parte dos dados mostra o caminho que o modelo deve seguir para melhor entender a relação entre as variáveis. Hastie et al. (2009)

Diversos métodos utilizam esta forma de treinamento para ajustar um modelo, dentre eles a regressão linear, regressão logística, algumas redes neurais, os vetores de suporte as máquinas (SVM) e a grande maioria de modelos utilizados em previsão de séries temporais.

- b. **Aprendizado não supervisionado** – Representa um tipo de aprendizado um pouco mais desafiador. Neste caso, não podemos supervisionar o resultado do modelo, pois não temos uma resposta y_i associada aos diferentes valores assumidos pelas variáveis independentes x_i .

Este é um tipo de aprendizado com diversas aplicações, entre elas os modelos de clusterização.

Como destaca Bishop (2006, p.23), “o objetivo em tais problemas de aprendizagem não supervisionados pode ser descobrir grupos de exemplos semelhantes presente nos dados, o que é chamado de agrupamento, ou determinar a distribuição dos dados de entrada, conhecido como estimativa de densidade, ou ainda projetar os dados de um espaço de alta dimensionalidade em até duas ou três dimensões para fins de visualização”. (Tradução nossa).

O Perceptron

A forma mais simples de uma rede neural é o Algoritmo de Aprendizado *Perceptron* (*Perceptron Learning Algorithm* - PLA). Desenvolvido por Rosenblatt (1958), o PLA é um simples modelo matemático que representa a base dos modelos de redes neurais desenvolvidos nas décadas de 1980 e 1990. De fato, ele é uma rede neural - a mais simples possível. É um algoritmo pra problemas de aprendizado supervisionado de classificação binária. O mais simples *Perceptron* é composto por uma única unidade de processamento, um neurônio, que é alimentado por dados de entrada e produz uma resposta com base em um critério de decisão adotado (FIGURA 1).

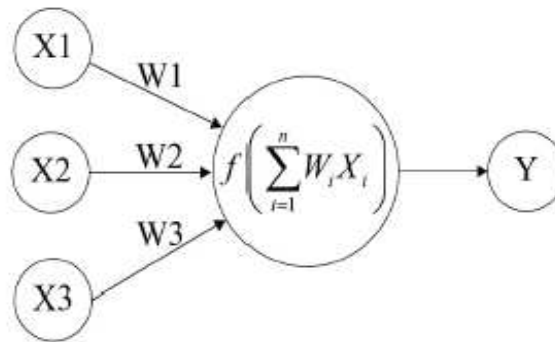


FIGURA 1 – A estrutura de um *Perceptron*. Fonte: Neural Networks with R. Ciaburro e Venkateswaran (2017)

A figura 1 ilustra a estrutura deste simples modelo matemático. Os três círculos da esquerda representam três neurônios (nomeados X1, X2 e X3) que compõem os dados de entrada de do modelo. O círculo maior localizado no centro da figura representa o núcleo do neurônio de processamento, nele os valores de entrada são processados e é produzida uma saída. O círculo à direita representa o resultado do modelo. No caso do *Perceptron*, é uma simples resposta binária que pode ser “1” ou “0” ou ainda “sim” ou “não”.

Os elementos acima das flechas nomeados por W1, W2 e W3 representam os pesos do modelo. Estes pesos são um dos componentes principais do *Perceptron*, eles passam por diversas atualizações durante processo de aprendizagem para que o modelo produza bons resultados. Outro elemento importante do *Perceptron* está no interior do núcleo representado por $f()$, este elemento é uma função chamada função de ativação. Seu papel é realizar uma simples operação: converter a soma ponderada dos pesos e dados de entrada em um resultado binário.

O mecanismo que permite a aprendizagem funciona com base na maneira que o *Perceptron* percebe erro que ele cometeu na sua classificação.

O erro é simplesmente a diferença entre a resposta correta e a resposta produzida pelo modelo. Assumindo que podemos fornecer dados de entrada n vezes ao *Perceptron*, podemos definir o n -ésimo erro como sendo:

$$e(n) = y(n) - \hat{y}(n) \quad (1)$$

Onde $y(n)$ é a resposta correta e $\hat{y}(n)$ é a resposta obtida como saída do modelo.

Quando existe erro, o modelo começa o aprendizado para tentar não errar na próxima vez que for testado. Isto acontece através de uma atualização dos pesos W que foram utilizados na n -ésima vez que o *Perceptron* foi utilizado. Deste modo, temos que o valor do peso W_i , $i = 1, \dots, n$, na tentativa seguinte passa a ser:

$$W_i(n+1) = W_i(n) + \Delta i \quad (2)$$

Onde $W(n)$ é o valor do peso W_i na etapa n e Δi representa a quantidade de mudança em W_i após a aprendizagem. A variável Δ está em função do erro $e(n)$: Quando não há erro, $\Delta i = 0$ e o valor do peso W_i não se altera, quando há erro, $\Delta i \neq 0$, o *Perceptron* atualiza o valor do peso W_i para que haja mais respostas corretas na próxima utilização do algoritmo.

A atualização dos pesos é a essência dos algoritmos de aprendizado, tanto do *Perceptron* simples quanto a rede neural multicamadas ou de qualquer outra rede neural.

Em tarefas de classificação, o *Perceptron* pode ser entendido como um classificador que objetiva encontrar uma fronteira linear (um hiperplano) que divide os dados linearmente separáveis em dois grupos distintos. É um classificador da forma (1) que computa uma combinação linear das variáveis explicativas e retorna uma resposta binária.

$$x : \hat{w}_0 + \hat{w}_1 x_1 + \hat{w}_2 x_2 = 0 \quad (3)$$

O hiperplano separador (Figura 2) é encontrado pela mecânica do PLA, que minimiza a distância de observações classificadas incorretamente em relação à fronteira de decisão.

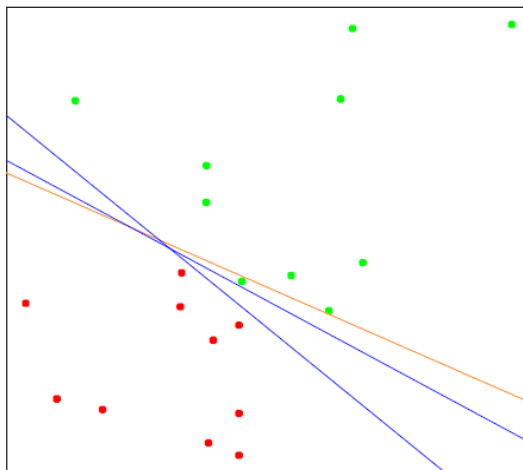


FIGURA 2: Aplicação do algoritmo PLA para encontrar hiperplanos. Fonte: Elements of Statistical Learning. Hastie et al. (2009)

No entanto, como destacam Hastie et al. (2009) existem alguns problemas com este algoritmo, tais como:

- 1 – Múltiplas soluções possíveis (hiperplanos separadores).
- 2 – É útil apenas para um conjunto de dados linearmente separáveis e problemas de solução binária.
- 3 – A capacidade computacional de um único *Perceptron* é limitada.

Apesar de sua baixa aplicabilidade e desempenho que deixa a desejar, o PLA foi fundamental para o desenvolvimento de importantes métodos de *machine learning* que lidam com problemas não lineares, tais como o SVM e principalmente as RNA's.

A rede neural *Perceptron* de multicamadas

Uma rede neural *Perceptron* de multicamadas (Multi-Layer *Perceptron* – MLP), assim como o PLA, também é um modelo matemático com um fluxo de dados unidirecional da camada de entrada para a de saída, cujo resultado final é obtido através de transformações dos dados de entrada. Mas, diferentemente do PLA, esta rede MLP é caracterizada por diversas camadas com neurônios interligados que formam diferentes caminhos paralelos de processamento dos dados, do início ao final da rede neural (FIGURA 3).

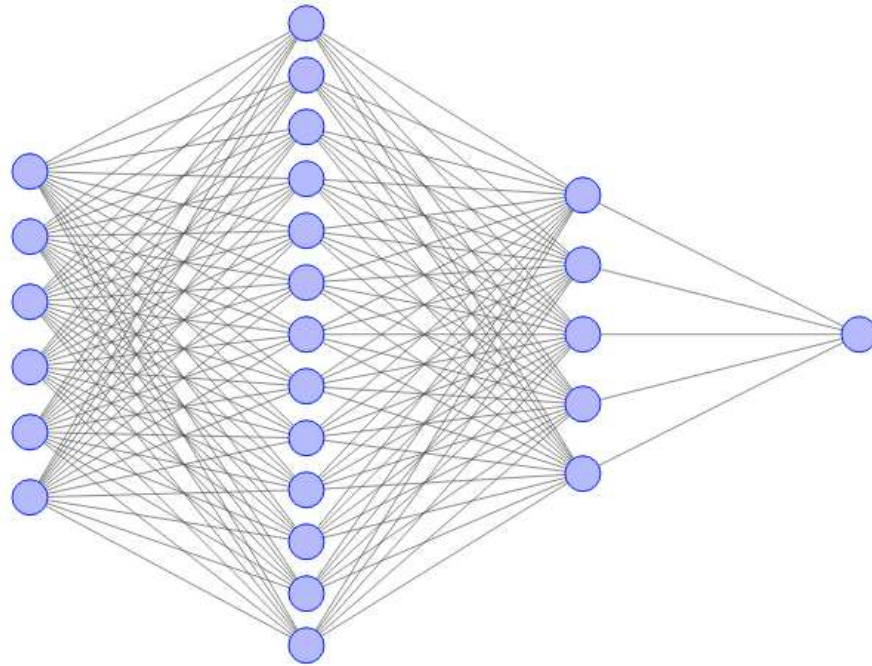


FIGURA 3 – A estrutura de uma rede MLP com uma camada de entrada de 6 neurônios, duas camadas de processamento com 13 e 5 neurônios cada e uma camada de saída com um neurônio. Fonte: Elaboração própria utilizando software R.

Assim como nas redes neurais naturais, nas redes artificiais MLP a unidade central de processamento são os neurônios, que operacionalizam as entradas para gerar uma saída. Similarmente ao que acontece no PLA, a saída de um neurônio da rede MLP é uma função da soma ponderada dos dados de entrada. Cada neurônio desempenha uma operação simples que envolve a ativação se o sinal total de informação recebida atingir um certo limite. Basicamente, o objetivo de rede neural como um todo, é simplesmente computar a saída de todos os neurônios, que é um cálculo inteiramente determinístico (Ciaburro e Venkateswaran, 2017).

As redes neurais são treinadas para ajustar os pesos (w e constantes) que são calibrados sequencialmente com base nos valores de saída da rede. Suas arquiteturas se baseiam em multi-camadas de neurônios: a camada de entrada, as camadas ocultas (processamento) e a de saída. Na i -ésima saída do j -ésimo neurônio $x_{j,i}$ é aplicado um peso (w), este é um parâmetro numérico que determinam o quanto um neurônio em uma camada vai afetar outro na camada precedente. Por fim, é aplicada uma função de ativação $f()$, que é uma transformação no somatório dos dados ponderados pelo pesos.

A função matemática não linear usada no ajuste dos dados é a função de ativação $f()$, ela que converte as observações que entram em uma única saída do neurônio. É nas funções de ativação que a mágica da rede neural acontece, pois sem as funções de ativação as redes neurais seriam apenas transformações lineares. Como muitos problemas da sociedade são não

lineares e complexos, então se faz necessário ir além da linearidade para alcançar uma boa modelagem e é por este objetivo que as funções de ativação ajudam a tornar as RNA's ferramentas importantes para solucionar problemas reais⁷.

As figuras abaixo apresentam dois exemplos de funções de ativação amplamente utilizadas em redes neurais. A figura 4 mostra graficamente uma função de ativação linear. Para um neurônio com esta função de ativação a saída é igual a entrada, a função é representada pela fórmula:

$$f(x) = x \quad (4)$$

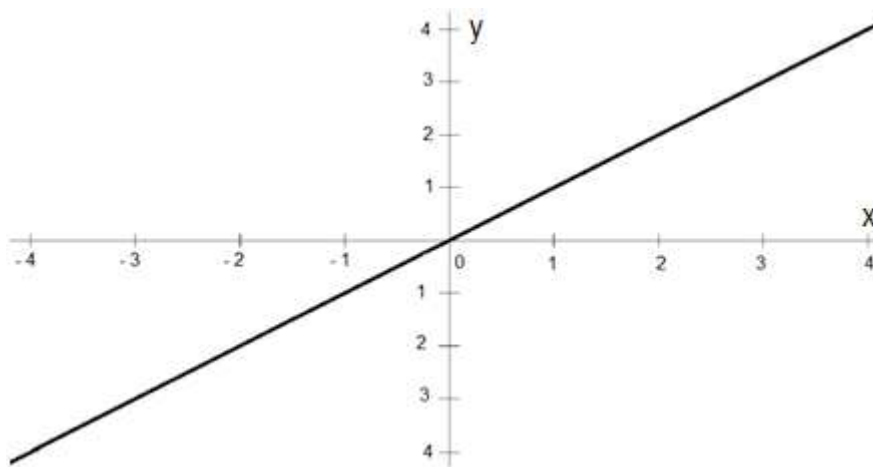


Figura 4 – Representação gráfica da função linear. Fonte: Adaptado de Ciaburro e Venkateswaran (2017)

A figura 5 mostra graficamente uma função de ativação sigmoide (representada pelo símbolo σ). Esta é a mais antiga e mais usada função de ativação (Ciaburro e Venkateswaran, 2017). Nesta função de ativação, todos os valores de entrada são transformados em um valor entre 0 e 1. A função σ é representada pela fórmula:

$$f(x) = 1/(1 + e^{-x}) \quad (5)$$

⁷ São as funções de ativação que permitem uma rede neural lidar com problemas não lineares. Existem diversas funções de ativação disponíveis, algumas delas são a linear, *Tanh*, *sigmoide* e *ReLU*. Para mais detalhes ver Neural Ciaburro e Venkateswaran (2017)

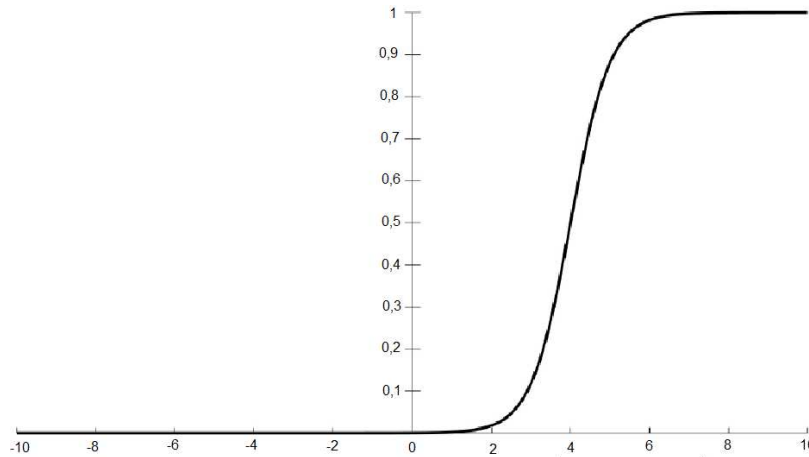


Figura 5 – Representação gráfica função sigmoide. Fonte: Adaptado de Ciaburro e Venkateswaran (2017)

A ideia central das redes MLP se baseia na extração de combinações lineares das saídas dos neurônios e na modificação do resultado por funções de ativação como essas. O resultado é um poderoso modelo de aprendizagem amplamente utilizado nas mais diversas e para os mais diferentes objetivos.

Uma importante desvantagem das redes neurais MLP é que elas sofrem com o problema de memória de curto prazo. Se uma sequência de dados for longa as redes MLP podem deixar de considerar informações passadas importantes, pois existe uma dificuldade em transportar informações das etapas anteriores para as posteriores.

Redes Neurais recorrentes

Ao contrário de uma rede MPL que tem um fluxo de dados unidirecional da camada de entrada para a de saída, uma rede neural recorrente (RNR) tem um fluxo bidirecional que propaga os dados para frente e também para trás, isto é, dos estágios de processamento posteriores para os anteriores. Neste tipo de rede os dados fluem bidirecionalmente durante o processo de treinamento, e é isto que caracteriza a recorrência.

As RNR são bem adequadas para problemas de aprendizado supervisionado em que o conjunto de dados tem uma natureza sequencial. São redes neurais com memória, elas tem a capacidade de se lembrar do passado, e isto dá a elas a capacidade de produzir um comportamento dinâmico Ciaburro e Venkateswaran (2017). Tal característica aliada com a capacidade da rede de aprender sequencialmente a torna uma boa opção para problemas de modelagem de series temporais, tradução de textos, reconhecimento de fala, entre outros; pois tais problemas requerem uma memória interna para reforçar o processo de aprendizagem.

A FIGURA 6 mostra esquematicamente a recorrência de uma rede neural, esta característica permite pensarmos uma RNR como várias “cópias” de uma mesma rede neural MLP (essas cópias estão apresentadas como um quadrado verde com a palavra “módulo”). As “cópias” são chamadas de módulos, que é simplesmente a rede neural em uma das iterações. Em cada repetição do módulo, os dados de entrada são três elementos: os valores das variáveis explicativas no período t , X_t (na figura, os círculos azuis na parte inferior), a saída do módulo anterior, \hat{y}_{t-1} (representada pelos círculos roxos na parte superior) e também o estado da célula (que são informações sobre os módulos passados persistentes na rede LSTM durante todas as iterações dos módulos). O fato de a saída de um módulo ser dado de entrada para o próximo módulo é o que caracteriza a recorrência. Em problemas de natureza sequencial a saída do módulo representa a resposta da rede LSTM no instante t . Na previsão de séries de tempo, por exemplo, a previsão do período t é a saída do módulo no instante t e também é dado de entrada para criar a previsão no instante $t+1$.

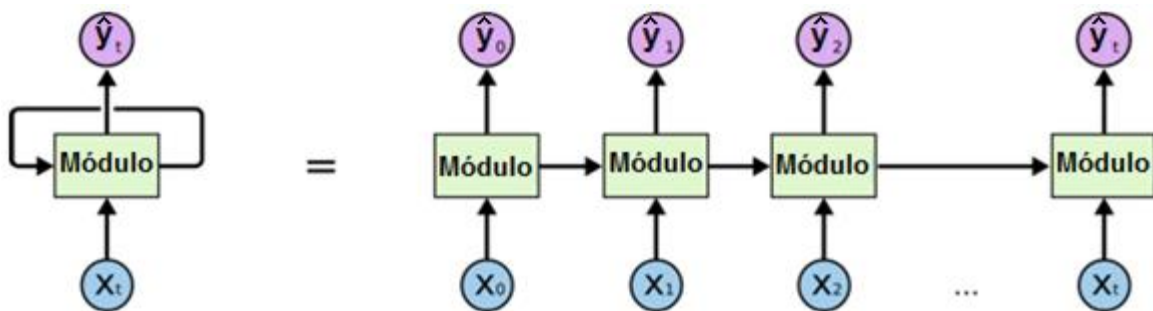


FIGURA 6 – Representação da recorrência. Fonte: Olah (2015)

Este comportamento de “*looping*” nos módulos é que agrega o caráter dinâmico à rede neural e é o que permite a persistência de informações de um período passado distante dentro da rede neural.

A mecânica de processamento é a mesma do *Perceptron* e da rede MLP: é um sistema de somatórios ponderados e funções diferenciáveis que capturam os padrões complexos presente nos dados através de funções não lineares.

Long Short Term Memory

Um tipo especial de rede neural foi desenvolvida para corrigir o problema de memória de curto prazo da rede MLP: a rede de memória de curto-longo prazo (Long-Short Term Memory – LSTM; Hochreiter e Schmidhuber, 1997). Esta rede neural tem na sua estrutura,

mecanismos chamados de *Gates*. Eles cumprem o papel de uma comporta, regulando o fluxo de informação do passado, fazendo a rede se lembrar do que é necessário e assim corrigindo o problema da memória curta. São os *gates* que tornam a LSTM útil na aplicação de problemas que exigem aprendizado sequencial com cadeias mais longas de dados, como por exemplo a previsão de séries de tempo.

A rede neural LSTM pode ser vista como uma classe especial de RNRs que representou grandes avanços. Lembrar dos padrões de períodos passados distantes é a sua função. Ela tem esta capacidade devido aos *gates* presentes nos módulos, que armazenam informações passadas e aplicam operações sobre elas. Os *gates* filtram o que é relevante para a previsão e descartam o que não é.

Por ser recorrente, a rede LSTM também tem a característica de cadeia de repetição, no entanto o módulo tem uma estrutura bem específica e um pouco mais complexa: existem cinco camadas de neurônios, interagindo de uma maneira bem peculiar.

A figura 7 apresenta a estrutura de um módulo de LSTM. Na parte inferior do módulo existem os três *gates*, estes são simples camadas de neurônios representadas por quadrados amarelos e rotulados com o símbolo σ . Estes rótulos de camadas de neurônios são comum na literatura sobre redes neurais, eles dizem respeito à função de ativação dos vários neurônios presentes na camada. Conforme exposto na seção sobre redes MLP, a função pode ser linear, *tanh*, *sigmoide*, etc. Existem outras duas camadas de neurônios *tanh*, uma localizada na parte superior direita do módulo e a outra na parte inferior central – ambas representadas por elipses roxas.

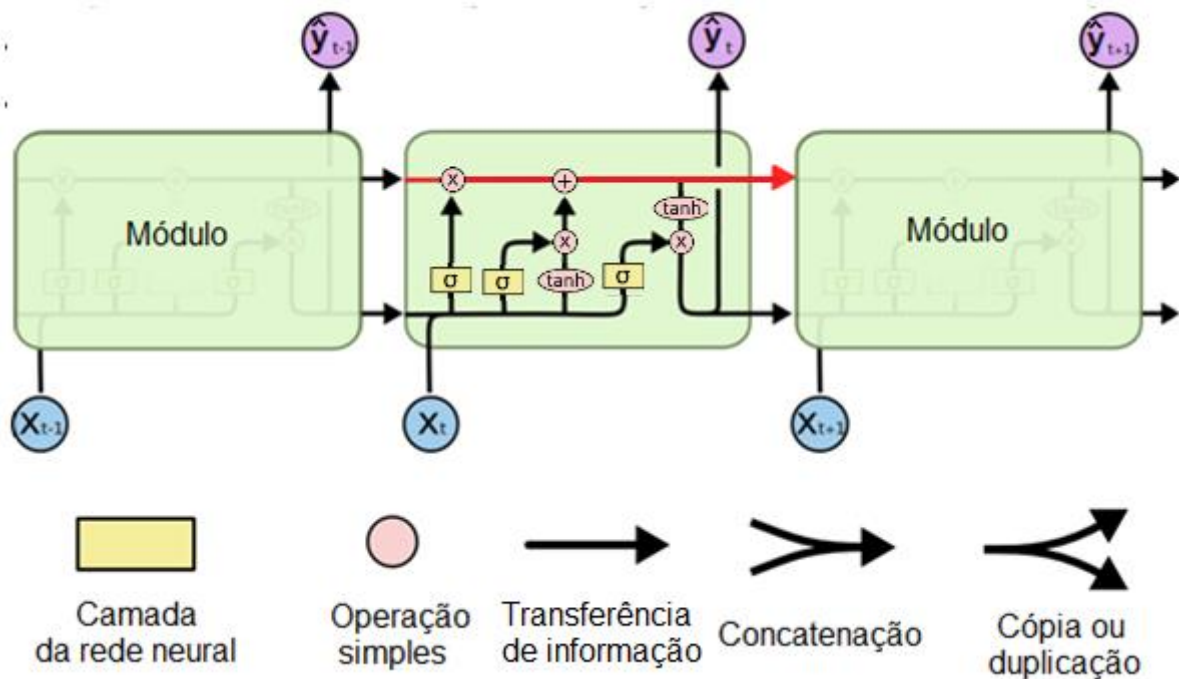


FIGURA 7 – Módulo de uma LSTM. Fonte: Adaptado de Olah, C. (2015)

A figura 7 também apresenta o significado das flechas e dos outros elementos do módulo, que representam simples operações aritméticas tais como adição e multiplicação. Uma parte fundamental de um módulo é o chamado estado da célula que é uma informação do passado que o módulo recebe e repassa alterando-a. O caminho pelo qual esta informação (o estado da célula) passa no módulo está representado na figura 7 como uma seta vermelha na parte superior do módulo.

O entendimento de como as informações são processadas dentro de uma rede LSTM à primeira vista é um pouco confusa, mas a chave para compreender o funcionamento de uma LSTM é entender o caminho que os dados fazem dentro do módulo da rede neural. Entender a função dos *gates* e do estado da célula⁸ é fundamental.

Iniciando pelo estado da célula. Este é o elemento do módulo que traz informações dos módulos anteriores, ele sofre alterações no módulo atual e é repassado para módulos posteriores. O estado da célula é um dos principais componentes que são utilizados para produzir a saída da rede LSTM no período t .

As informações são alteradas, adicionadas ou removidas do estado da célula pelos *gates*. Os *gates* são camadas de neurônios que decidem qual informação é relevante devendo ser

⁸ O estado da célula leva informação durante a sequência de processamento, ele cria um elo temporal entre os módulos. As informações de módulos anteriores são trazidas e levadas reduzindo assim os efeitos de uma memória curta - que é característica das redes MLP. Para mais detalhes sobre como o estado da célula gere a memória de uma rede LSTM, ver Chung et al. (2014)

mantida/lembrada e qual é irrelevante devendo ser descartada/esquecida. Eles são gestores da memória de longo prazo.

Através das alterações feitas no estado da célula, os *gates* regulam o fluxo de informações na rede neural. A função de ativação dos *gates* é *sigmoide* (por isso a representação ilustrativa por um quadrado com σ). Quando um neurônio de um *gate* recebe informações e produz como resultado o valor 0, isto significa que essas informações e seus derivados devem ser esquecido; quando produz 1, significa que toda a informação que recebeu e seus derivados são relevantes e devem ser considerados pela rede LSTM (este é um dos momentos em que a rede LSTM pega informações do passado e constrói a memória).

Uma maneira visual de apresentar o interior de um módulo de uma rede LSTM está apresentado na Figura 8.

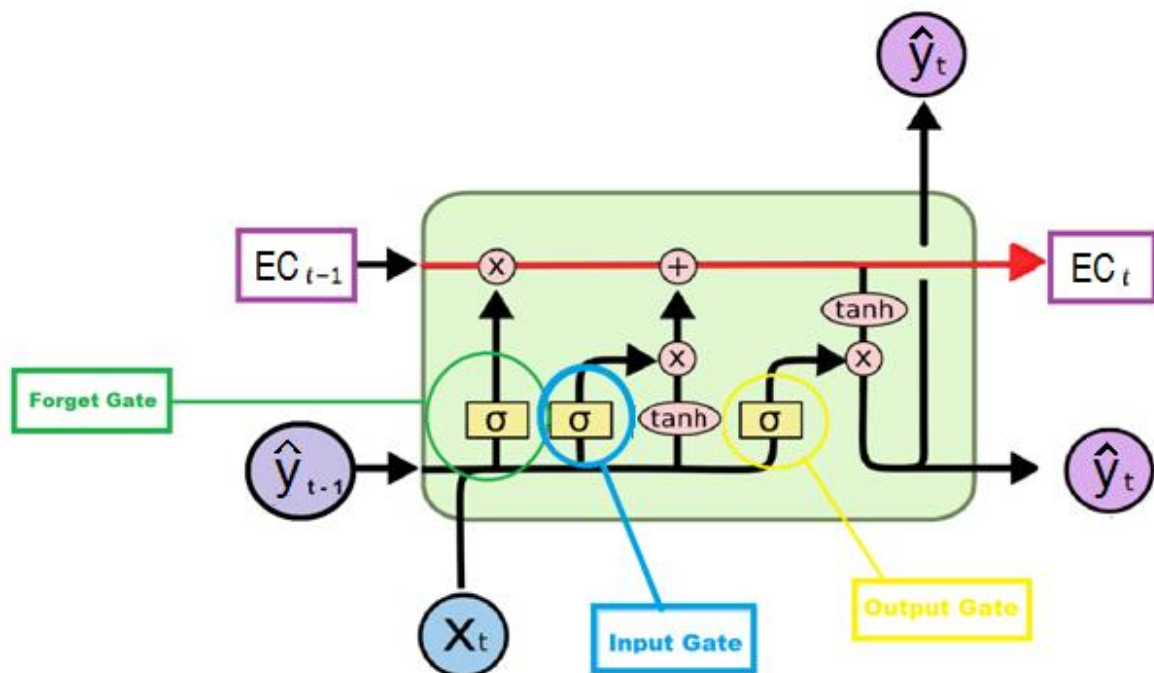


FIGURA 8 – Mecânica de um módulo de LSTM. O caminho percorrido pelo estado da célula está representado como linha horizontal vermelha no topo do módulo. Fonte: Adaptado de Olah (2015)

Inicialmente um módulo recebe três dados de entrada:

- \hat{y}_{t-1} - A saída do módulo anterior
- X_t - Os valores das variáveis explicativas no período t .
- EC_{t-1} - O estado da célula do módulo anterior que é transmitida para o módulo atual

Na realização do processamento de todas essas informações, primeiro temos o *forget gate*. Ele é o primeiro a modificar o estado da célula do módulo atual, decidindo qual informação deve ser jogada fora. Nesta etapa é realizada uma transformação da saída do módulo anterior (\hat{y}_{t-1}) e dos dados de entrada (X_t). Isto é feito através de uma camada de neurônios *sigmoide* (do *forget gate*) que analisa as informações X_t e \hat{y}_{t-1} retornam um número entre zero e um que será utilizado para atualizar o estado da célula do módulo passado que chegou no módulo atual, representado na figura 8 por EC_{t-1} . Mais próximo de 0 significa que o estado da célula anterior, \hat{y}_{t-1} e X_t não são importantes e podem ser esquecidos, mais próximo de 1 significa são importantes, isto é, devem ser lembrados durante as etapas de processamento posteriores do módulo atual.

O resultado do *forget gate* é definido como:

$$f_t = \sigma(W_f \cdot [\hat{y}_{t-1}, x_t]), \text{ onde:} \quad (6)$$

- σ é a função de ativação sigmoide
- W_f são os pesos das conexões entre os dados de entrada e os neurônios do *input gate*.
- $[\hat{y}_{t-1}, x_t]$ são os dados de entrada: a saída do módulo passado concatenada com os dados de entrada x_t .

Como já foi dito, a função de ativação da camada é uma sigmoide, por isso o resultado deste *gate* - representado por f_t - é um número entre 0 e 1⁹. É importante frisar que o *forget gate* é simplesmente uma camada de neurônios, e a fórmula de seu resultado é apenas uma transformação de somas ponderadas por uma função e ativação, assim como foi apresentado no caso do *Perceptron*. Veja a ilustração na figura 1.

Com o resultado desta operação do *forget gate*, obtemos a primeira alteração do estado da célula:

$$EC'_t = EC_{t-1} * f_t, \text{ onde:} \quad (7)$$

- EC'_t é o estado da célula atualizado no módulo atual
- EC_{t-1} é o estado da célula do módulo anterior
- f_t é o resultado do *forget gate*.

⁹ Rigorosamente falando, em uma rede LSTM não são introduzidos nem produzidos resultados de um único elemento escalar. A operacionalização de uma rede LSTM envolve vetores, matrizes e tensores. Para mais detalhes ver Hochreiter, 1997

Todas estas alterações no estado da célula acontecem no caminho destacado por uma flecha vermelha no topo do módulo (FIGURA 8). Neste momento, a rede LSTM já começa esquecer o que deveria ser esquecido e lembrar o que deve ser lembrado. Prosseguimos para atualizar o estado da célula novamente, desta vez utilizando o *input gate* e uma camada de neurônios com função de ativação *tanh*.

No *input gate*, passamos novamente as mesmas informações - saída do módulo anterior (\hat{y}_{t-1}) e dos dados de entrada (X_t) - em uma camada *sigmoide*, que decide o quanto estas informações são importantes para este outro *gate*. Mais próximo de 0 significa que tudo deve ser esquecido, mais próximo de 1 significa que tudo é importante para a memória da rede. A diferença no caso do *input gate* é que a atualização do estado da célula **não é imediata**. Paralelamente ao processo do *input gate*, os dados de entrada e a saída do módulo anterior são inseridos em uma camada de neurônios *tanh* (Veja a ilustração na figura 6 novamente). Isto é feito para reescalonar os valores entre -1 e 1, o objetivo é ajuda a regular o funcionamento da rede LSTM e calibrar melhor os pesos. Em seguida, multiplicamos a saída da *tanh* com a do *input gate* e utilizamos este produto para finalmente atualizar o estado da célula.

Seja o resultado do *input gate*, i_t definido como:

$$i_t = \sigma(W_i \cdot [\hat{y}_{t-1}, x_t]) \text{ , onde:} \quad (8)$$

- σ é a função de ativação sigmoide.
- W_i são os pesos das conexões entre os dados de entrada e os neurônios do *input gate*.
- $[\hat{y}_{t-1}, x_t]$ são os dados de entrada: a saída do módulo passado concatenada com os dados de entrada.

Seja o resultado da camada *tanh*, representado por $tanh_t$, ele é definido como:

$$tanh_t = \tanh(W_{tanh} \cdot [\hat{y}_{t-1}, x_t]) \text{ , onde:} \quad (9)$$

- *tanh* é a função de ativação *tanh*.
- W_i são os peso das conexões entre os dados de entrada e os neurônios da camada *tanh*.
- $[\hat{y}_{t-1}, x_t]$ são os dados de entrada: a saída do módulo passado concatenada com os dados de entrada.

Então obtemos a última atualização do estado da célula no módulo atual:

$$EC_t = EC'_t + i_t * tanh_t \text{ , onde:} \quad (10)$$

- EC_t é o estado da célula atualizado no módulo atual
- EC'_t é o estado da célula que foi atualizado pelo *forget gate* na etapa passada
- $i_t * \tanh_t$ é produto entre os resultados do *input gate* e da camada *tanh*. ($0 \leq i_t \leq 1$) e ($-1 \leq \tanh_t \leq 1$)

O estado da célula atualizado EC_t é o estado da célula que será dado de entrada do próximo módulo, porém não é a saída deste módulo da rede LSTM (não é a previsão no período t). O EC_t será usado para calcular a saída (\hat{y}_t) do módulo atual que é igual a previsão pontual da LSTM para o período t .

Simultaneamente às transformações nos *forget* e *input gate*, temos a operacionalização das informações por parte do *output gate*.

A saída deste *gate* define a saída da rede LSTM (\hat{y}_t). Primeiro passamos novamente saída do módulo anterior (\hat{y}_{t-1}) e dos dados de entrada (X_t) na camada *sigmoide* do *output gate*. Ao mesmo tempo, o estado da célula EC_t é passado na camada *tanh* localizada no lado superior direito do módulo da figura 9.

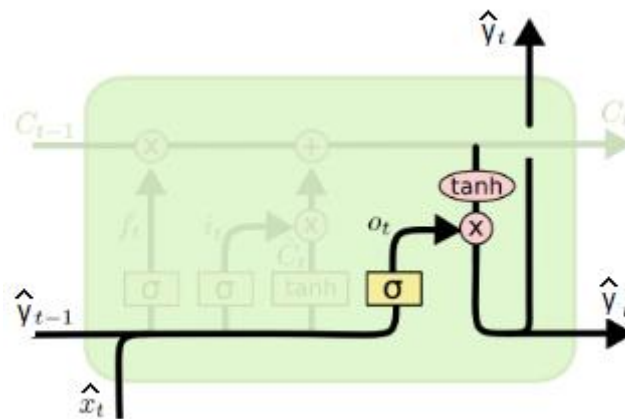


Figura 9 – Processamento final em um módulo. Fonte: Adaptado de Olah(2015)

Então multiplicamos a saída do *output gate* pelo estado da célula EC_t que foi transformado pela camada *tanh*, e como resultado temos a saída (\hat{y}_t) que será insumo para próximo módulo e também a previsão no período t da rede LSTM.

$$Y_{lstm,t} = \hat{y}_t = o_t * \tanh(EC_t) \quad (11)$$

Onde:

- $Y_{lstm,t}$ é a saída e previsão da rede neural LSTM no instante t .
- \hat{y}_t é a saída deste módulo que será utilizada como dado de entrada do próximo módulo.

$$- o_t \text{ é a saída do } output \text{ gate. Onde } o_t = \sigma(W_o \cdot [\hat{y}_{t-1}, x_t]) \text{ e } 0 \leq o_{t,i} \leq 1 \quad (12)$$

Então o estado da célula EC_t e a saída (\hat{y}_t) são levados para o próximo módulo onde serão os insumos para a recalibração dos pesos juntamente com novos dados de entrada (X_t). Assim como no caso do *Perceptron* e da rede MLP, as alterações no valor dos pesos são feitas com base no erro de previsão da rede, que é a diferença entre a saída $Y_{Istm,t}$ e o verdadeiro dado no período t , Y_t .

Por fim, \hat{y}_t e X_t serão utilizados para calcular qual deve ser a saída em um período $t + 1$ utilizando a mesma mecânica de funcionamento deste módulo.

3.3 Modelos SARIMA

É um importante método para fazer modelagem e previsão de séries temporais foi introduzido por Box-jenkins (1970). Esta metodologia consiste de modelos lineares da classe ARMA (*autoregressive integrated moving average*), que são particularmente uteis para representar processos estocásticos estacionários e desempenham um papel fundamental na modelagem de séries temporais. Os modelos ARMA(p,q) são modelos adequados para séries univariadas, nesta metodologia se utiliza as informações passadas da série temporal para fazer previsões. Sua representação é através de dois polinômios: um de médias móveis dos erros de ordem q e outro autorregressivo de ordem p .

Um polinômio de média móvel dos erros $\theta(B)$:

$$(1 - \theta_1 B^1 - \theta_2 B^2 - \dots - \theta_q B^q) \varepsilon_t \quad (13)$$

Um polinômio Autorregressivo da série, $\phi(B)$:

$$(\phi_1 B^1 + \phi_2 B^2 + \dots + \phi_p B^p) Y_t \quad (14)$$

Definindo Y_t como uma série temporal estacionária, temos um Modelo ARMA(p,q) dado por:

$$(\phi_1 B^1 + \phi_2 B^2 + \dots + \phi_p B^p) Y_t = \delta + (1 - \theta_1 B^1 - \theta_2 B^2 - \dots - \theta_q B^q) \varepsilon_t \quad (15)$$

Neste modelo, pode ser feita a aplicação direta do método de mínimos quadrados ordinários (MQO) para estimar os valores dos parâmetros $(\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q)$ e assim ajustar o modelo para a tarefa de previsão.

Para a metodologia ser aplicável o termo de erro ε_t precisa ser um ruído branco, tendo média nula e variância constante. Também é condição necessária para a aplicação dos modelos $ARMA(p,q)$ que o polinômio autorregressivo atenda condição de estacionariedade e que o polinômio de média móvel atenda a condição de invertibilidade¹⁰.

Esta modelagem é aplicável apenas em séries estacionárias. Quando a série não é estacionária devemos utilizar métodos como a diferenciação (representada por Δ) para fazer a série estacionária e tornar possível a aplicação do método.

Seja Y_t um processo não-estacionário homogêneo de ordem d . Diferenciando Y_t d vezes transformamos Y_t em estacionário, ou seja, $W_t = \Delta^d Y_t$ é um processo estacionário. Além de estacionarizar e estabilizar a série temporal através da diferenciação simples, pode também ser necessário estabilizar a variância da série. Isto pode ser feito através da transformação de Box-Cox (1964). Neste caso, quando utilizamos a diferenciação, o modelo que se assume para a série é o $ARIMA(p,d,q)$ que é derivado do $ARMA(p,q)$, porém com a diferença de que é um processo integrado de ordem d .

Como muitas das séries econômicas são caracterizadas por padrões sazonais, também precisamos incorporar esse fator importante em nossa modelagem. A maneira de aplicar a metodologia Box-Jenkins em séries sazonais é torná-las diferenciadas sazonalmente.

MODELOS ARIMAS SAZONAIS

Também conhecidos como modelos SARIMA, são úteis para modelar o componente sazonal presente em séries de tempo. A construção se assemelha à do $ARIMA(p,d,q)$ simples.

Podemos ter os seguintes tipos de sazonalidade (representados pela letra “s”) :

- Mensal: $s = 12$
- Trimestral: $s = 4$
- Quadrimestral: $s = 3$
- Semestral: $s = 6$

Similarmente ao caso $ARIMA(p,d,q)$ simples, a modelagem se dá através de dois polinômios, um de média móvel sazonal de ordem Q : $\Theta(B^s)$; e outro autorregressivo sazonal de ordem P : $\Phi(B^s)$.

¹⁰ A condição de estacionariedade tem que ser satisfeita pelo polinômio AR. Já o polinômio MA deve satisfazer a condição de invertibilidade, isto é, o polinômio MA possui raízes fora do círculo unitário. Esta é um hipótese sustentada por Box e Jenkins durante o desenvolvimento da metodologia, se trata de uma conveniência matemática para que a razão dos polinômios resulte em um polinômio convergente. Para mais detalhes, ver Mattos (2018).

Unindo esses dois modelos, temos uma modelagem multiplicativa que abrange de forma mais completa as características da série temporal. Este novo modelo é uma combinação do $ARIMA(p,d,q)$ simples com o $ARIMA(P,D,Q)$ sazonal, sua representação na versão compacta é a seguinte:

$$\phi(B)\Phi(B^s)\Delta^d\Delta_s^D Y_t = \theta(B)\Theta(B^s)\varepsilon_t, \text{ onde:} \quad (16)$$

- $\phi(B)$ é o polinômio autoregressivo simples de ordem p
- $\Phi(B^s)$ é o polinômio autoregressivo sazonal de ordem P
- Δ^d é o operador de diferença simples de ordem d
- Δ_s^D é o operador de diferença sazonal de ordem D
- Y_t é a série temporal
- $\theta(B)$ é o polinômio média móvel simples de ordem q
- $\Theta(B^s)$ é o polinômio média móvel sazonal de ordem Q
- ε_t é o termo de erro

A construção de um modelo $ARIMA(p, d, q)$ passa pela identificação dos lags máximos p do polinômio AR e q do polinômio MA. Faz-se o uso das funções de autocorrelação simples e parcial, os padrões presentes nessas funções são úteis para mensurar a dependência linear entre as observações da série permitindo assim identificar os valores dos parâmetros p e q . Depois de definida as ordens dos polinômios, estimamos os parâmetros do modelo por MQO e aplicamos o teste Q Box e Pierce (1970), a fim de testar se os erros seguem um processo ruído branco. Após estas etapas, podemos utilizar o modelo $ARIMA$ para modelar e fazer previsão da série temporal, o caso do $ARIMA(p, d, q)(P, D, Q)$ é análogo.

3.4 Validação Cruzada

No processo de treinamento de um algoritmo de machine learning, precisamos utilizar somente os dados que temos para verificar a acurácia e fazer o modelo aprender como é a relação entre as variáveis. A técnica de validação cruzada é útil neste caso onde temos que conhecer a capacidade de generalização do modelo com base no conjunto de dados que temos.

Esta técnica realiza um particionamento do conjunto de dados entre um conjunto de treino e um conjunto de teste. Com o primeiro conjunto o modelo aprende com os dados, com o segundo conjunto verifica-se a capacidade de generalização do modelo para dados novos.

Existem diversas maneiras de efetuar esse particionamento, duas das principais são a abordagem do conjunto de validação e o *leave-one-out* (LOO).

A abordagem do conjunto de validação consiste em dividir aleatoriamente o conjunto de observações entre conjunto de treino e conjunto de teste. Esta abordagem é conceitualmente simples e fácil de implementar, porém como pontuou Hastie et al. (2009), existem duas desvantagens. A primeira delas é que estimativa do erro utilizando o conjunto de teste pode ter uma variabilidade grande, dependendo do particionamento que foi feito aleatoriamente. A segunda desvantagem é que apenas um subconjunto das observações é utilizado para o ajuste do modelo. Como os métodos de aprendizado estatístico tendem a desempenhar mal quando se usa poucos dados, é possível que o erro do conjunto de teste superestime o verdadeiro erro do modelo (considerando um treinamento com todos os dados disponíveis).

Uma segunda maneira de dividir os dados é o método LOO. É um método parecido com a abordagem do conjunto de validação, mas o LOO tenta mitigar as desvantagens. LOO também envolve dividir o conjunto de dados em um subconjunto de teste e outro de treino. Todavia, uma única observação é utilizada para validação e todas as $n - 1$ observações restantes são utilizadas para o treinamento do modelo. É obtida uma previsão \hat{Y}_i para a observação excluída do treino e assim temos um erro de previsão denotado por:

$$\hat{e} = (Y_i - \hat{Y}_i)^2 \quad (17)$$

Esta é uma estimativa aproximadamente não-viesada para o erro de teste mas com muita variabilidade já que é feita com base em uma única previsão; (Hastie et al.; 2009). Repetindo este método n vezes produzimos n erros, $\hat{e}_1, \hat{e}_{21}, \dots, \hat{e}_{n1}$.

A estimativa do método LOO para o erro de teste de validação cruzada é a média destes erros:

$$VC_n = \frac{1}{n} \sum_{i=1}^n \hat{e} \quad (18)$$

O método LOO é um método de uso geral e pode ser utilizado em várias modelagens cujo objetivo é a previsão. Além disso, o LOO tem algumas vantagens comparado a abordagem do conjunto de validação, tais como:

- Menor viés

- Nesta abordagem, executar o LOO múltiplas vezes sempre vai produzir o mesmo resultado: não há variabilidade no tamanho do particionamento.

3.5 Medidas de desempenho

Para avaliar o poder de previsão dos modelos, faremos uso de três medidas de desempenho que são amplamente utilizadas na literatura de séries temporais. São elas:

Erro quadrático médio (*Mean Squared Error* - MSE), é calculada como uma média quadrática do erro de previsão e é definida pela fórmula:

$$\text{MSE} = \frac{1}{n} \sum_1^n (y_i - \hat{y}_i)^2 \quad (19)$$

Raiz do erro quadrático médio (*Root Mean Squared Error* - RMSE), é calculada como a raiz da média quadrática do erro de previsão e é definida pela fórmula:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_1^n (y_i - \hat{y}_i)^2} \quad (20)$$

Raiz do erro quadrático médio (*Root Mean Squared Error* - RMSE), expressa o erro do modelo em percentagem e é definida pela fórmula:

$$\text{MAPE} = \frac{100}{n} \sum_1^n \frac{y_i - \hat{y}_i}{y_i} \quad (21)$$

Onde y_i são os dados da série, \hat{y}_i são as previsões do modelo e n o número de observações. Note que se $\hat{y}_i = y_i$ todas as métricas de desempenho são iguais a zero, ou seja, não existe erro nas previsões.

4 RESULTADOS

Nesta seção foram feitas análises descritivas das séries que refletem o comportamento dos salários e também foram realizadas aplicações dos métodos expostos na seção de metodologia para gerar previsões dos níveis salariais para os próximos doze trimestres. As previsões foram feitas utilizando a linguagem R, o software Rstudio e a plataforma *Google Colab*¹¹. A modelagem ARIMA foi feita no Rstudio, já as redes neurais foram treinadas no Google Colab.

As técnicas de aprendizado profundo como as redes neurais são intensivas em computação, por isso as previsões das RNAs foram realizadas no Google Colab. Este serviço oferece oito unidades de processamento de tensor (TPUs) o que possibilita um melhor treinamento das redes e em menor tempo. Quando mais recursos computacionais estão disponíveis, mais experimentos podem ser executados com mais dados e mais *epochs*. Isso pode fornecer um treinamento melhor ao modelo, o que significa hiperparâmetros melhor estimados e previsões mais corretas.

O presente trabalho limita-se a ajustar arquiteturas básicas de LSTM, aquém do verdadeiro poder preditivo do modelo. Caso haja maior poder computacional, outras redes neurais mais poderosas podem ser aplicadas de forma semelhante e com a mesma mecânica de funcionamento.

As previsões foram realizadas para as quatro categorias de salários de forma absoluta, ou seja, tendo em vista o valor total dos salários em cada uma das categorias. O horizonte de previsão destas análises foi de doze trimestres (ou três anos). A interpretação dos resultados do modelo ARIMA se baseia na identificação dos parâmetros e suas significâncias, na verificação estrutural de correlação dos erros e na análise das medidas de desempenho dos modelos. As redes neurais têm pouca interpretabilidade, mas temos informações sobre a quantidade de neurônios em cada camada, a checagem de *overfitting* e a verificação das medidas de desempenho. Portanto, nesse estudo foi necessária a observação do erro quadrado médio (MSE), das funções de perda e do teste de Ljung-Box; cujos objetivos são respectivamente: identificar quanto o modelo erra ao projetar o próximo período, identificar a

¹¹ O **Google Collaboratory** é um serviço de armazenamento e execução em nuvem de *rotinas computacionais* de criação e execução de códigos em Python e R. Seu funcionamento é diretamente em um navegador, sem a necessidade de nenhum tipo de instalação de software em um computador pessoal.

etapa de treinamento ótima e avaliar a estrutura de correlação dos erros. Esse conjunto de informações permite expor a solidez de cada um dos modelos.

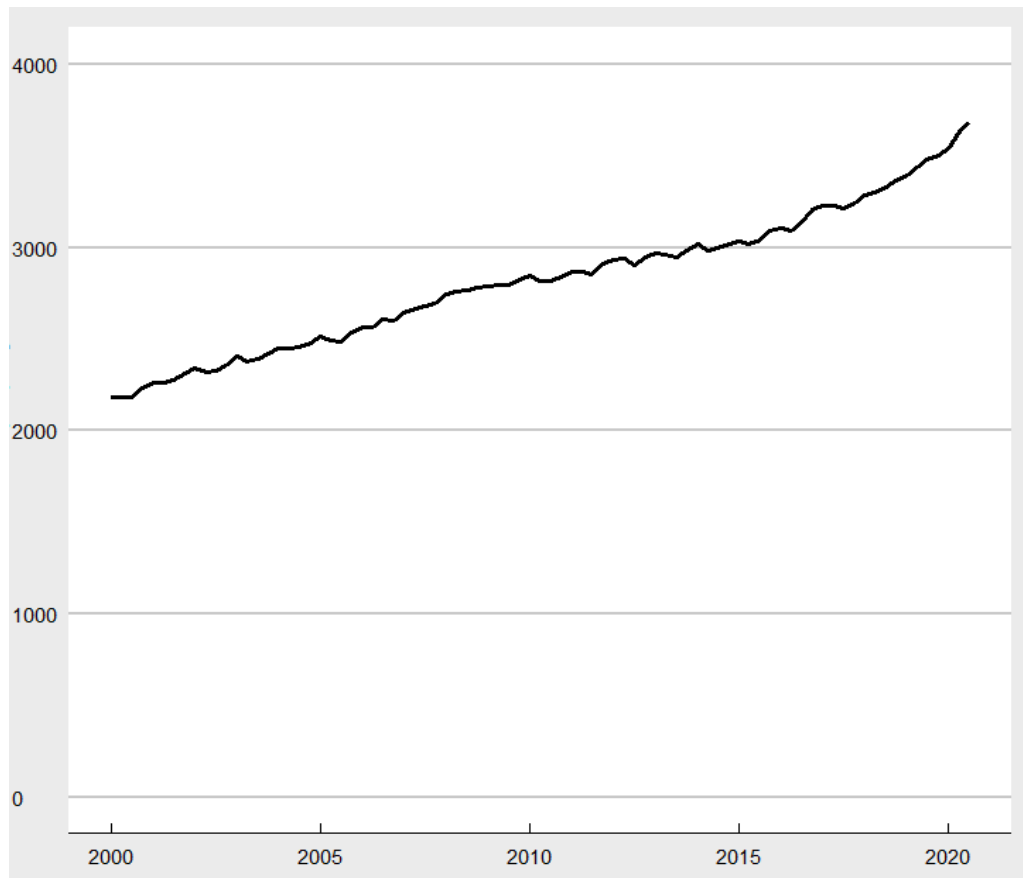
Para a calibração dos pesos e parâmetros as séries foram divididas em uma parte contendo 90% dos dados que serviu de treinamento e outra parte contendo 10% dos dados que serviu para avaliar o desempenho fora da amostra dos modelos. Foram configuradas diferentes tipos de redes neurais, com arquiteturas contendo diferentes números de neurônios e diferentes etapas de treinamento.

Para acessar o poder preditivo de cada modelo foram utilizadas as três medidas de desempenho já expostas neste texto: MSE, MAPE E RMSE. Os resultados das previsões foram consolidados em quadros para facilitar o entendimento ao leitor, e logo após foram apresentadas as análises individuais dos modelos que tiveram melhor desempenho.

4.1 Análise descritiva

O GRÁFICO 2 apresenta gráfico do valor do salário semanal médio(em dólares) dos trabalhadores dos EUA de todos os tipos de ocupações. Fica claro um padrão ascendente da série nos últimos 20 anos, o que revela uma tendência de aumento dos ganhos salariais durante a modernização da economia e mecanização das atividades que aconteceu neste século. Porém este efeito pode estar mascarado pelo crescimento econômico, que traz consigo maiores correções do valor do salário e um aumento do mesmo. Uma melhor forma de analisar o comportamento dos salários é visualizá-los como proporção do pib, o que é feito a seguir.

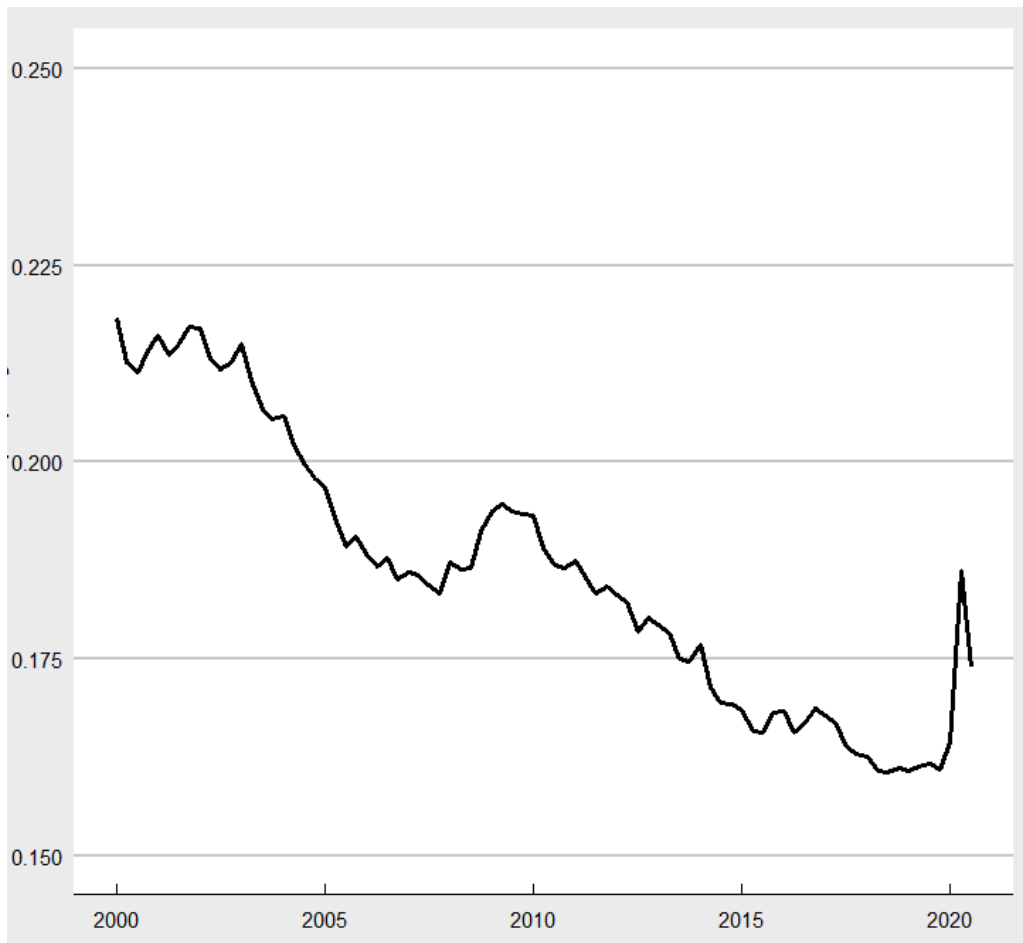
GRÁFICO 2 – Salário Real Médio Semanal (em dólares) - EUA



Fonte: Elaboração própria utilizando o software R.

Na GRÁFICO 3, temos o salário real médio mensal dividido pelo produto interno bruto em bilhões de dólares, ambas as quantidades medidas a valores correntes. A série indica que, nas últimas duas décadas, a quantidade de salário necessária para produzir a mesma quantidade de produto tem diminuído. Isto é um indício de que o fenômeno da polarização dos empregos transborda para além dos postos de trabalho, neste caso específico, para os salários. Este resultado corrobora com os resultados de alguns estudos empíricos que destacam a diminuição dos salários com a automação de cadeias produtivas. Algumas estimativas de Chiacchio et al. (2018), por exemplo, sugeriram um impacto negativo da automação nos salários, isto é, os robôs e a tecnologia da informação diminuindo os salários dos trabalhadores ao longo do tempo.

GRÁFICO 3 – Razão (Média Salarial Semanal)/PIB - EUA



Fonte: Elaboração própria utilizando o software R.

Como pode ser observado no gráfico, nos anos 2000 os Estados Unidos precisavam gastar aproximadamente 22% ¹² do valor do PIB (**mensurado em bilhões**) para pagar o salário semanal médio de um trabalhador, esta proporção diminuiu 6 pontos percentuais até antes da pandemia da COVID-19, atingindo o nível mínimo de 16%. Com a crise mundial, o PIB dos EUA teve uma queda expressiva de 32,9% no segundo trimestre de 2020, o que contribuiu para que a razão salário/PIB ficasse maior. O aumento foi de 2,6 p.p. com o pico de 18,6% sendo alcançado no terceiro trimestre de 2020.

¹² Esta porcentagem é obtida considerando a razão (média salarial semanal)/(PIB_bilhões). O denominador está mensurado em dólares, enquanto que o numerador está em bilhões de dólares. Se formos utilizar ambos os termos na mesma unidade, a razão fica 10^8 vezes menor. Temos, por exemplo, que nos anos 2000 os EUA gastava 0,0000000022% do PIB para pagar o salário médio semanal de um trabalhador. Nesta última proporção, considera-se a razão (média salarial semanal)/(PIB), ambos os termos mensurados em dólares a valores correntes.

4.2 Análise Econométrica e de Machine Learning

4.2.1 Salário de tarefas Não-Rotineiras Cognitivas

Como pode ser observado no QUADRO 2 e no QUADRO 3, foram testados diversos modelos com diversos parâmetros, configurações e arquiteturas diferentes. Para esta série de salários, o modelo ARIMA que melhor desempenhou fora da amostra foi o ARIMA(0,1,1)(1,0,1)[4], este modelo teve seus parâmetros significantes e o teste Ljung-Box indicou ausência de correlação dos termos de erro. A rede neural MLP que melhor ajustou-se aos dados foi uma rede com 150 neurônios na camada interior de processamento, ela foi treinada 20 vezes e sem taxa de *dropout*¹³. Dentre diversas arquiteturas de redes LSTM testadas, a que obteve melhores resultados foi uma LSTM com 500 *epochs* no treinamento, 50 unidades de neurônios por camada e sua taxa de *dropout* foi de 30%.

QUADRO 2 – Sarimas Testados Para a Série NRC

Modelo	Equação	RMSE	MAPE	MSE
ARIMA(0,1,1)(1,0,1)[4]	$(1 + 0,02B^4)(1 - B)Z_t = (1 + 0,75B)(1 - 58B^4)a_t$	54,53	3,53%	2973,83
ARIMA(1,1,1)(1,0,1)[4]	$(1 + 0,08B^4)(1 - 0,3B)(1 - B)Z_t = (1 + B)(1 - 0,57B^4)a_t$	60,07	3,97%	3608,0

¹³ O *dropout* é uma técnica eficaz de regularização que previne a rede neural de coadaptar durante o treinamento. Em uma rede neural, coadaptação refere-se a uma situação em que diferentes neurônios de processamento têm comportamento altamente correlacionado, e isto é uma das principais causas de *overfitting*. Quando uma camada possui *dropout* ela pode descartar aleatoriamente algumas das conexões entre as camadas e seus neurônios (os correlacionados), o que auxilia no combate ao *overfitting*. Para mais detalhes, ver Srivastava et. al (2014).

QUADRO 3 – RNAs Testados Para a Série NRC

Modelo	Neurônios	Etapas de treinamento	Taxa <i>dropout</i>	RMSE	MAPE	MSE
LSTM	50/camada	3000	50%	58,91	3,56%	3.469,98
	50/camada	1000	50%	53,66	3,54%	2.879,02
	50/camada	500	50%	72,68	4,62%	5.282,60
	50/camada	50	50%	71,27	4,64%	5.078,64
	15/camada	1000	50%	65,92	4,17%	4.345,24
	15/camada	500	50%	62,99	3,93%	3.967,69
	15/camada	50	50%	61,71	3,94%	3.808,44
	100/camada	1000	50%	71,13	4,58%	5.058,79
	100/camada	500	50%	62,99	3,87%	3.967,64
	100/camada	50	50%	65,22	4,09%	4.253,21
	50/camada	1000	30%	68,4	4,25%	4670,5
	50/camada	500	30%	60,0	3,45%	3450,2
	50/camada	50	30%	67,0	4,17%	4491,2
	50/camada	1000	5%	60,51	3,87%	3661,9
	50/camada	500	5%	84,04	5,6%	7063,26
	50/camada	50	5%	79,5	5,125	6235,24
MLP	5/camada	20	0%	49,45	2,8%	2445,7
	15/camada	20	0%	47,42	2,67%	2249,6
	45/camada	20	0%	47,15	2,60%	2224,0
	150/camada	20	0%	28,41	1,6%	807,04

Considerando todos estes modelos e as três medidas de desempenho (MSE, MAPE e RMSE), o modelo que melhor desempenhou fora da amostra foi a rede MLP de 150 neurônios em uma única camada interior. A rede considerou como dado de entrada os lags 1, 2 e 4 da série temporal. Inicialmente o valor dos parâmetros foram dados aleatoriamente e a rede precisou de 20 etapas de treinamento para ajustá-los até ter um bom ajuste. A FIGURA 10 apresenta a estrutura desta rede neural.

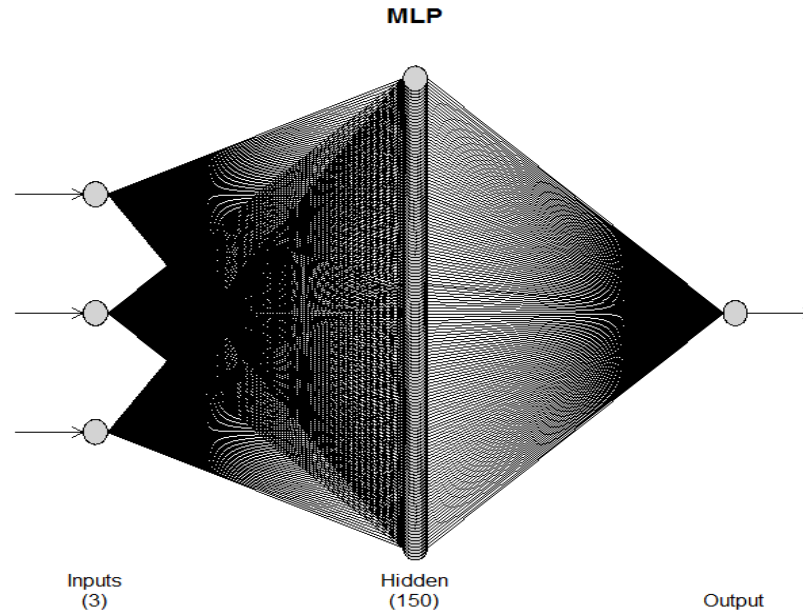
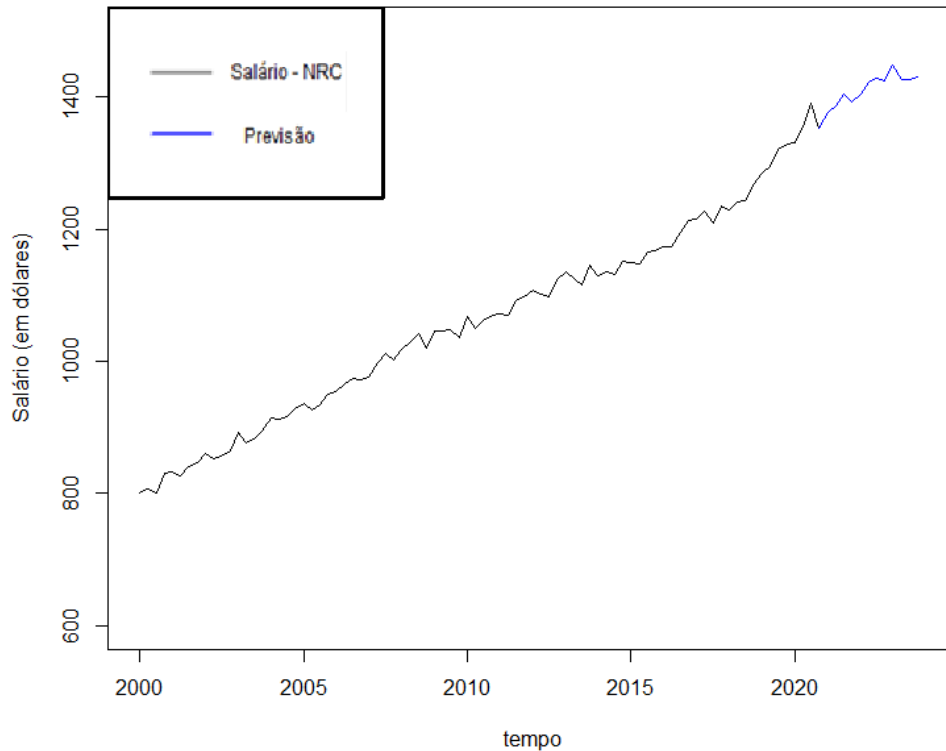


FIGURA 10 – Rede MLP de 150 neurônios na camada interior. Fonte: Elaboração própria utilizando o software R.

Esta rede apresentou um MAPE de 1,6%, um MSE de 807,04 e um RMSE de 28,41; sendo todas as medidas computadas com previsões fora da amostra. Sendo assim este modelo foi escolhido para fazer a previsão do salário de tarefas Não-Rotineiras Cognitivas (NRC).

O GRÁFICO 4 apresenta as previsões da rede neural MLP para os próximos 3 anos. A projeção interrompe a tendência de alta, trazendo uma estabilização do valor do salário No nível de 1400 dólares.

GRÁFICO 4 – Previsão salários – NRC com rede MLP



Fonte: Elaboração própria utilizando o software R.

4.2.2 Salário de tarefas Rotineiras Cognitivas

Conforme exposto no QUADRO 4 e no QUADRO 5, para esta série de salários foram testadas 16 redes neurais LSTM com diferentes arquiteturas, 2 modelos ARIMA e 4 redes MLP. O modelo ARIMA que melhor desempenhou fora da amostra foi o ARIMA(0,1,1)(1,0,1)[4], este modelo teve seus parâmetros significantes e teste Ljung-Box indicou ausência de correlação dos termos de erro. A rede neural MLP que melhor ajustou-se aos dados foi uma rede com 5 neurônios na única camada interior de processamento, ela foi treinada 20 vezes e sem taxa de *dropout*. Dentre diversas arquiteturas de redes LSTM testadas, a que obteve melhores resultados foi uma LSTM com 50 neurônios por camada, ela foi treinada 500 vezes e sua taxa de *dropout* é de 50%.

QUADRO 4 – Sarima Testados Para a Série RC

Modelo	Equação	RMSE	MAPE	MSE
ARIMA(0,1,2)(1,0,1)[4]	$(1 + 0,3B^4)(1 - B)Z_t = (1 + 0,65B)(1 + 0,2B)(1 - 0,5B^4)a_t$	41,87	4,43%	1753,5
ARIMA(0,1,1)(1,0,1)[4]	$(1 + 0,3B^4)(1 - B)Z_t = (1 + 0,65)(1 - 0,5B^4)a_t$	41,7	4,40%	1736,2

QUADRO 5 – RNAs Testados Para a Série RC

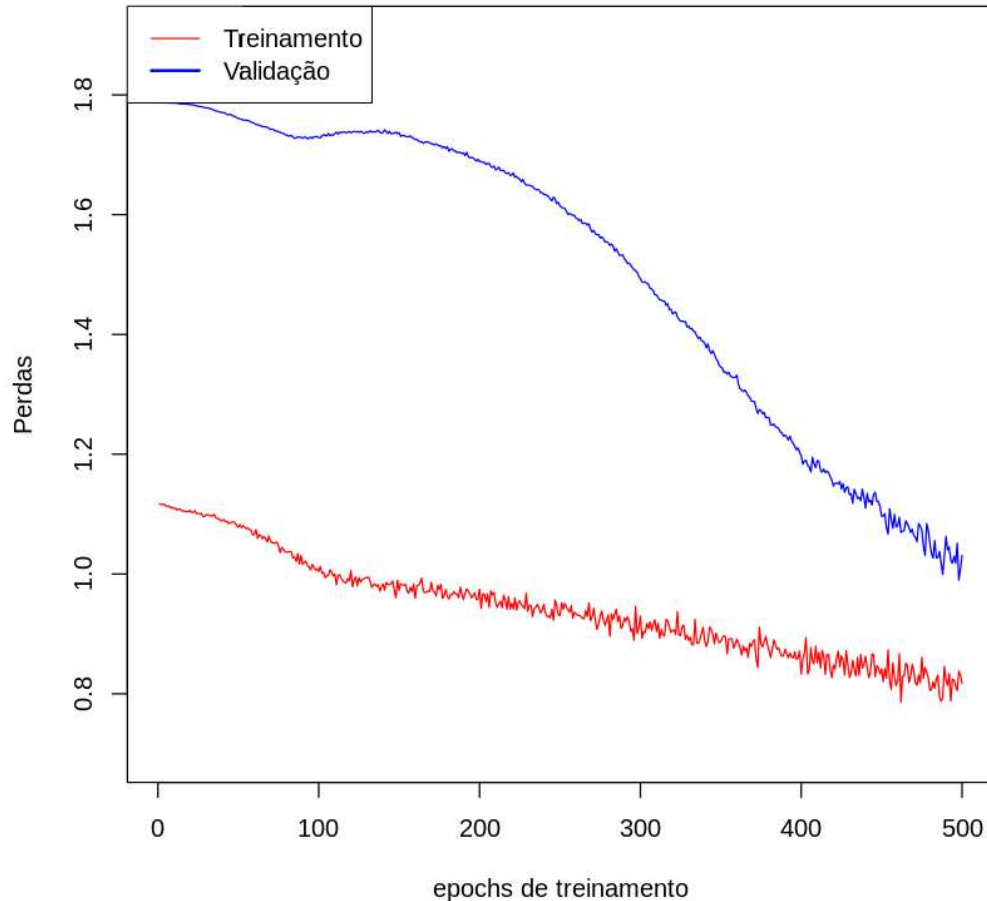
Modelo	Neurônios	Etapas de treinamento	Taxa <i>dropout</i>	RMSE	MAPE	MSE
LSTM	50/camada	1000	50%	22,61	2,22%	511,3
	50/camada	500	50%	16,58	1,56%	274,9
	50/camada	50	50%	39,16	4,34%	1533,7
	15/camada	1000	50%	36,7	3,85%	1349,5
	15/camada	500	50%	22,9	2,14%	523,6
	15/camada	50	50%	43,6	4,84%	1902,7
	100/camada	1000	50%	25,4	2,23%	645,8
	100/camada	500	50%	18,8	1,66%	354,2
	100/camada	50	50%	23,8	2,28%	568,5
	15/camada	1000	5%	37,09	3,61%	1375,9
	15/camada	500	5%	22,5	2,13%	505,3
	15/camada	50	5%	26,4	2,53%	698,9
	50/camada	1000	5%	21,1	1,89%	445,2
	50/camada	500	5%	22	2,13%	482,6
	50/camada	50	5%	18,12	1,63%	328,36
MLP	15/camada	1000	0%	33,2	3,46%	1105,1
	100/camada	20	0%	66,71	7,28%	4450,2
	5/camada	20	0%	30,35	2,84%	921,2
	15/camada	20	0%	36,55	3,61%	1336,2

Considerando todos estes modelos e as três medidas de desempenho (MSE, MAPE e RMSE), o modelo que melhor desempenhou fora da amostra foi a rede LSTM de 50 neurônios por camada. Esta RNA possui 30.651 parâmetros, distribuídos entre as camadas dos três gates e das camadas regularizadoras. Existem 50 neurônios em cada camada, e a rede foi treinada 500 vezes.

O GRÁFICO 5 apresenta o aprendizado do modelo durante as etapas de treinamento. O treinamento foi interrompido na etapa de número 500, mas poderia ter continuado até que houvesse convergência para um ponto ótimo no problema de otimização dos parâmetros (o que seria representado pelo ponto mais baixo do formato de “U” dessas curvas). A interrupção do treinamento da rede com 500 etapas aconteceu pelo fato de que nesta configuração a rede

neural desempenhou melhor fora da amostra – em comparação com as outras arquiteturas testadas.

GRÁFICO 5 – Erro da rede LSTM durante o treinamento



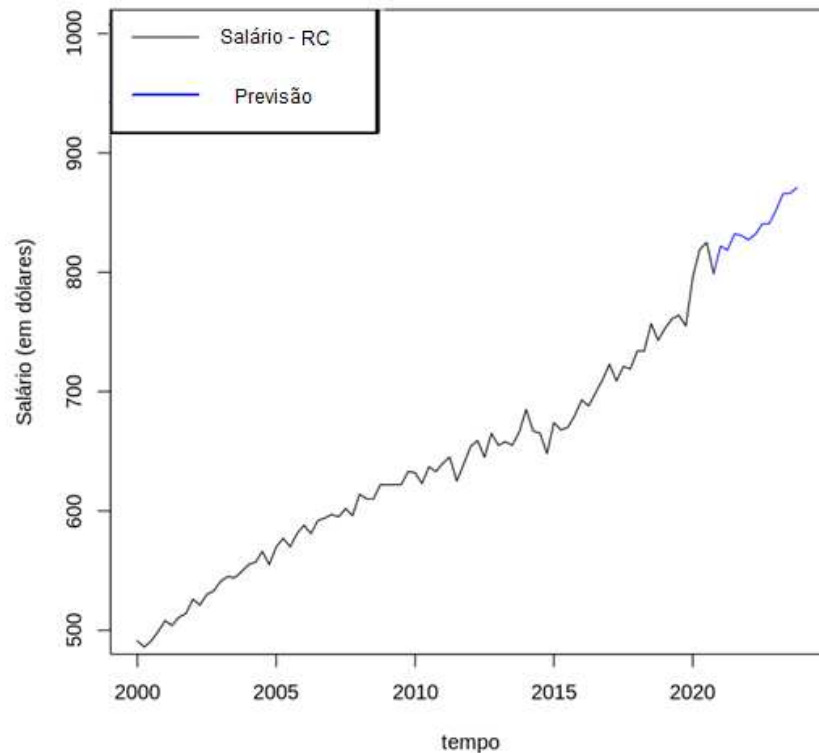
Fonte: Elaboração própria pelo software R.

A pouca quantidade de observações da série foi um fator que induziu o modelo ao *overfitting* durante o treinamento. Algumas técnicas de regularização foram usadas para tentar evitar o *overfitting*, uma delas foi estabelecer a taxa de *dropout* em 50% para algumas camadas de neurônios.

Esta rede apresentou um MAPE de 1,56%, um MSE de 274,9 e um RMSE de 16,58; sendo todas as medidas computadas com previsões fora da amostra. Deste modo, este modelo foi escolhido para fazer a previsão do salário de tarefas Rotineiras Cognitivas.

O GRÁFICO 6 apresenta as previsões da rede neural LSTM para os próximos 3 anos. A projeção produzida pela rede neural segue a tendência de alta das últimas duas décadas e estima o valor do salário em 870 dólares para o último trimestre de 2023.

GRÁFICO 6 – Previsão salários – RC com rede neural LSTM



Fonte: Elaboração própria pelo software R.

4.2.3 Salário de tarefas Não-Rotineiras Manuais

Como pode ser observado no QUADRO 6 e no QUADRO 7, para esta série também foram testados diversos modelos com diferentes configurações e arquiteturas. Para esta série de salários, o modelo ARIMA que melhor desempenhou fora da amostra foi o ARIMA(0,1,1)(0,0,1)[4], este modelo teve seus parâmetros significantes e teste Ljung-Box novamente indicou ausência de correlação dos termos de erro. A rede neural MLP que melhor ajustou-se aos dados foi uma rede com 15 neurônios na camada interior de processamento, ela foi treinada 20 vezes e sem taxa de *dropout*. Dentre diversas arquiteturas de redes LSTM testadas, a que obteve melhores resultados foi uma LSTM com 100 unidades de neurônios por camada, treinada 50 vezes e sua taxa de *dropout* é de 30%.

QUADRO 6 – Sarimas Testados Para a Série NRM

Modelo	Equação	RMSE	MAPE	MSE
ARIMA(0,1,1)(0,0,1)[4]	$(1 - B)Z_t = (1 + 0,62B)(1 - 0,49B^4)a_t$	33,56	4,90%	1126,3
ARIMA(0,1,2)(0,0,1)[4]	$(1 - B)Z_t = (1 + 0,62B)(1 - 0,02B)(1 - 0,49B^4)a_t$	34	4,98%	1144,31

QUADRO 7 - RNAs Testados Para a Série NRM

Modelo	Neurônios	Etapas de treinamento	Taxa <i>dropout</i>	RMSE	MAPE	MSE
LSTM	50/camada	1000	50%	49	7,20%	2396,3
	50/camada	500	50%	39,22	5,59%	1538,6
	50/camada	50	50%	42,4	5,57%	1798
	15/camada	1000	50%	36	5,15%	1299,22
	15/camada	500	50%	33,9	4,69%	1147,4
	15/camada	50	50%	35,1	4,70%	1225,9
	100/camada	1000	50%	35,4	5,35%	1253,4
	100/camada	500	50%	41,2	6,21%	1700,76
	100/camada	50	50%	34,35	4,42%	1180,14
	50/camada	1000	5%	40,15	5,75%	1612
	50/camada	500	5%	41,4	6,04%	1711,8
	50/camada	50	3%	38,1	4,94%	1452
	15/camada	1000	5%	38,11	5,25%	1453
	15/camada	500	5%	38,3	5,20%	1467,72
	15/camada	50	5%	36,1	4,88%	1303,02
15/camada	1000	0%	35,71	5,18%	1275,5	
MLP	100/camada	20	0%	31,1	4,40%	968,9
	5/camada	20	0%	29	4,02%	844,1
	15/camada	20	0%	28,54	3,94%	815

Considerando todos estes modelos e as três medidas de desempenho (MSE, MAPE e RMSE), o modelo que melhor desempenhou fora da amostra foi a rede MLP de 15 neurônios na camada interior. A rede considerou como dado de entrada os lags 1, 2 e 3 da série temporal. Inicialmente o valor dos parâmetros foi dado aleatoriamente e a rede precisou de 20 etapas de treinamento para ajustá-los até ter um bom ajuste. A FIGURA 11 apresenta a estrutura da rede neural.

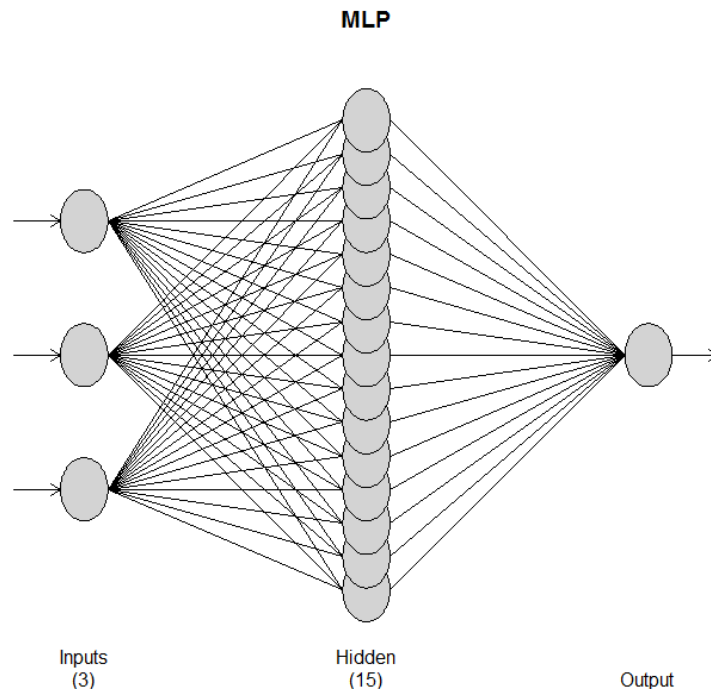
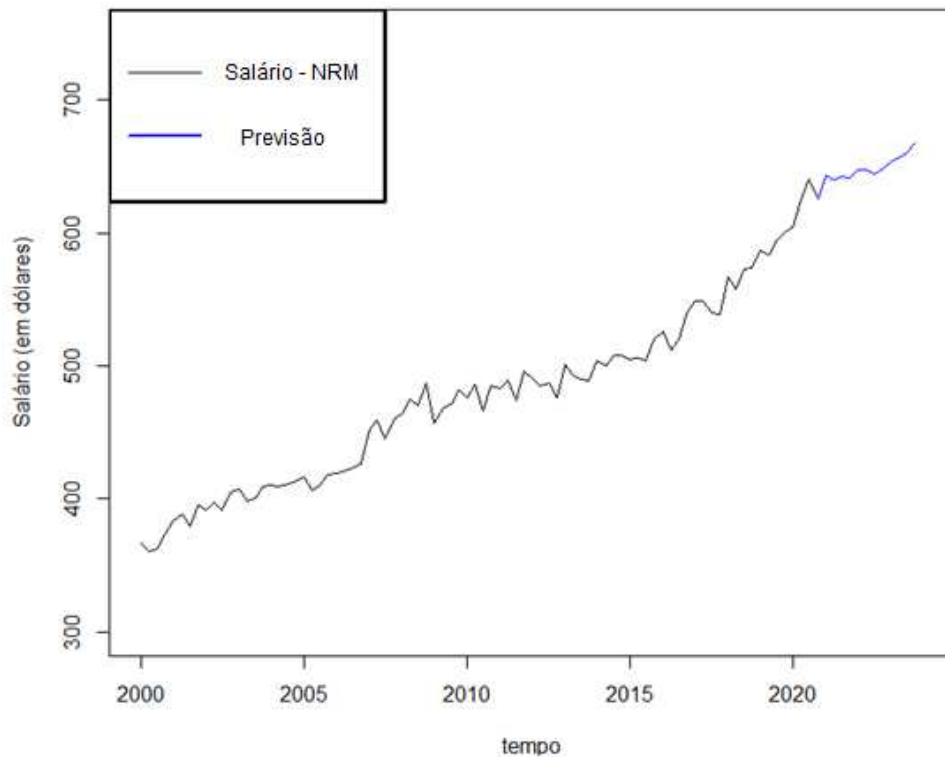


FIGURA 11 – Rede MLP de 150 neurônios na camada interior. Fonte: Elaboração própria pelo software R.

Esta rede apresentou um MAPE de 3,94%, um MSE de 815 e um RMSE de 28,54; sendo todas as medidas computadas com previsões fora da amostra. Sendo assim este modelo foi escolhido para fazer a previsão do salário de tarefas Não-Rotineiras Manuais (NRM).

O GRÁFICO 7 apresenta as previsões da rede neural MLP para os próximos 3 anos. A previsão continua projetando a tendência de alta, sem quebras estruturais na série. O valor do salário previsto para o último trimestre de 2023 é de 668 dólares.

GRÁFICO 7 – Previsão Salários – NRM com rede MLP



Fonte: Elaboração própria pelo software R.

4.2.4 Salário de tarefas Rotineiras Manuais

Novamente, estão apresentadas no QUADRO 8 e no QUADRO 9, as diferentes configurações e arquiteturas dos modelos que foram testados para esta série. O modelo ARIMA que melhor desempenhou fora da amostra foi o ARIMA(1,1,0)(1,0,1)[4], este modelo teve seus parâmetros significantes e teste Ljung-Box indicou ausência de correlação dos termos de erro. A rede neural MLP que melhor ajustou-se aos dados foi uma rede com 15 neurônios na camada interior de processamento, ela foi treinada 20 vezes e sem taxa de *dropout*. Dentre diversas arquiteturas de redes LSTM testadas, a que obteve melhores resultados foi uma LSTM com 1000 *epochs* no treinamento, 50 unidades de neurônios por camada e sua taxa de *dropout* é de 50%.

QUADRO 8 – Sarimas Testados Para a Série RM

Modelo	Equação	RMSE	MAPE	MSE
SARIMA(1,1,0)(1,0,1)[4]	$(1 - 0,99B^4)(1 + 0,45B)(1 - B)Z_t = (1 + 0,97B^4)a_t$	33,1	3,50%	1095,3
SARIMA(1,1,2)(1,0,1)[4]	$(1 + 0,45B)(1 - 0,99B^4)(1 - B)Z_t = (1 + 0,1B)(1 + 0,18B)(1 + 0,97B^4)a_t$	34,7	3,56%	1135,6

QUADRO 9 - RNAs Testados Para a Série RM

Metodologia	Neurônios	Etapas de treinamento	Taxa <i>dropout</i>	RMSE	MAPE	MSE
LSTM	15/camada	1000	0%	32,5	3,27%	1054,8
	50/camada	1000	50%	20,46	2,17%	419
	50/camada	500	50%	40,4	4,56%	1627,3
	50/camada	50	50%	28,87	3%	833,6
	15/camada	1000	50%	30,8	3,47%	948,5
	15/camada	500	50%	27,9	3,11%	780,25
	15/camada	50	50%	31,11	3,33%	967,9
	100/camada	1000	50%	26	2,58%	678,2
	100/camada	500	50%	29,1	3,16%	846,5
	100/camada	50	50%	32,2	3,41%	1037
	50/camada	1000	5%	28	2,81%	786,6
	50/camada	500	5%	33	3,23%	1086,2
	50/camada	50	5%	38,75	4,18%	1502
	15/camada	1000	5%	27,85	3,03%	775,9
	15/camada	500	5%	31,1	3,50%	969,8
15/camada	50	5%	36,9	3,95%	1362,2	
MLP	100/camada	20	0%	16,3	1,67%	264,5
	5/camada	20	0%	18,65	1,88%	348,4
	15/camada	20	0%	12,6	1,30%	158,7

Considerando todos estes modelos e as três medidas de desempenho (MSE, MAPE e RMSE), o modelo que melhor desempenhou fora da amostra foi a rede MLP de 15 neurônios na camada interior. A rede considerou como dado de entrada o *lag* 1 da série temporal e também a sazonalidade determinística presente na série. Inicialmente o valor dos parâmetros foram dados aleatoriamente e a rede precisou de 20 etapas de treinamento para ajustá-los até ter um bom ajuste. A FIGURA 12 apresenta a estrutura da rede neural.

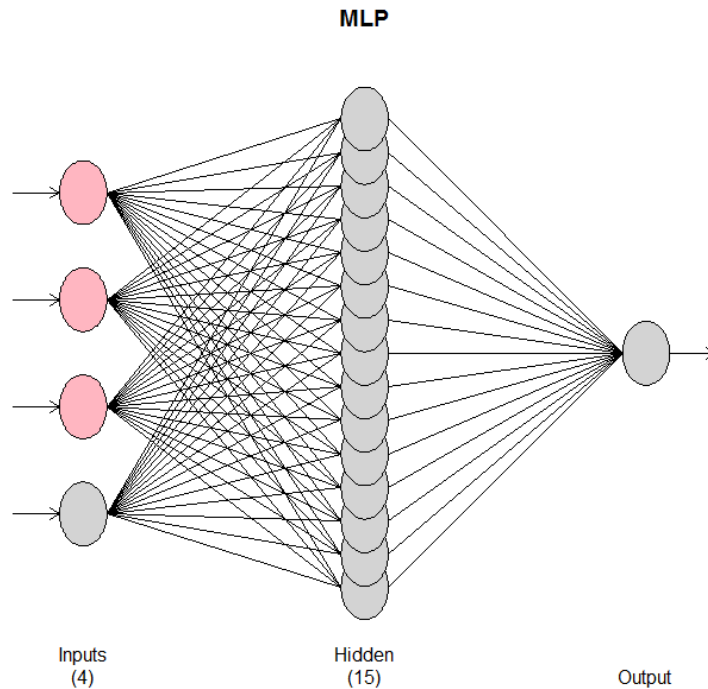
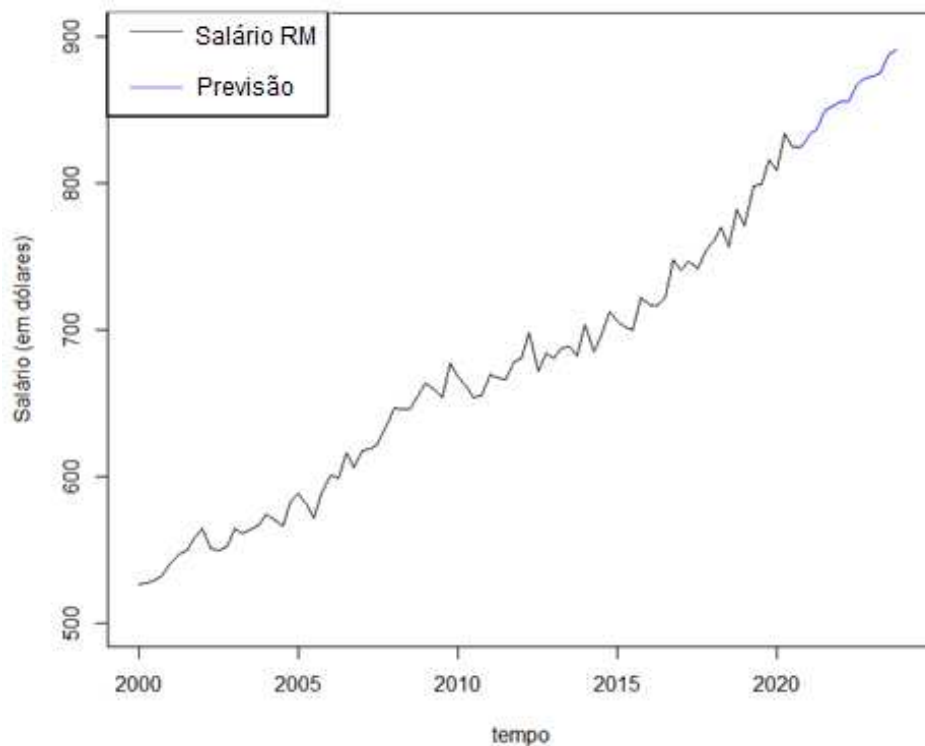


FIGURA 12 – Rede MLP com 15 neurônios de processamento e 4 de entrada (os neurônios em rosa representam a sazonalidade da série que a rede optou por tomar como entrada). Fonte: Elaboração própria pelo software R.

Esta rede apresentou um MAPE de 1,3%, um MSE de 158,7 e um RMSE de 12,6; sendo todas as medidas computadas com previsões fora da amostra. Deste modo, este modelo foi escolhido para fazer a previsão do salário de tarefas Rotineiras Manuais(RM).

O GRÁFICO 8 apresenta as previsões da rede neural MLP para os próximos 3 anos. A estimativa mantém tendência de alta, não apresenta quebras estruturais e projeta um aumento do valor do salário atingindo pico de 890 dólares no último trimestre de 2023.

GRÁFICO 8 – Previsão salários - RM com rede MLP



Fonte: Elaboração própria pelo software R.

4.2.5 Resultados da Modelagem

Os dados disponibilizados na *Current Population Survey* mostram que os salários de todas as quatro categorias estudadas tiveram aumentos expressivos em termos absolutos nos últimos 20 anos. Os primeiros dados disponíveis após a pandemia da COVID-19 também mostram que esta crise não afetou negativamente o salário semanal médio, uma possível razão para isto é a conhecida rigidez salarial. Além disto, a série agregada dos salários médios semanais divididos pelo PIB apresenta um comportamento decrescente que persistiu nos últimos 20 anos, o que indica a necessidade cada vez menor de pagar salários para produzir.

Diante do conjunto de dados que analisamos e considerando as previsões dos modelos, espera-se que valor médio semanal dos salários (em termos absolutos) cresça para todas as classes de tarefas de ocupações até 2023. O QUADRO 10 apresenta a análise em termos de variação percentual. A maior variação salarial em termos percentuais foi projetada para a série de salários de tarefas rotineiras cognitivas, a previsão foi realizada pela rede LSTM que melhor desempenhou fora da amostra conforme o QUADRO 5.

QUADRO 10 – Previsões dos Melhores Modelos

Classe de tarefas	Salário observado 4ª trimestre 2020	Salário projetado 4ª trimestre 2023	Varição percentual
Rotineiras Manuais	834,5	890	6,65%
Não-Rotineiras Manuais	626	682	8,95%
Rotineiras Cognitivas	799	875	9,51%
Não-Rotineiras Cognitivas	1353	1425	5,32%

Neste trabalho a escolha pela rede LSTM aconteceu pela característica deste modelo de gerir melhor memória quando comparado a rede MLP. Isto torna a rede LSTM mais apropriada na modelagem de problemas cuja natureza é sequencial, como a previsão de séries temporais. Curiosamente, a rede MLP desempenhou melhor do que a rede LSTM em três das quatro vezes. O motivo mais plausível para este resultado é que as séries são muito pequenas, com apenas 83 observações, isto faz com que a memória maior da rede LSTM não seja muito útil na modelagem.

A visão geral que a rede MLP teve da série durante o treinamento (por mais que seja menor do que a visão da LSTM) foi o suficiente para projetar resultados fora da amostra. A rede LSTM não teve seu potencial explorado pela pouca quantidade de dados e por isso foi superada em desempenho 75% das vezes pela rede MLP.

5 CONCLUSÃO

O presente estudo buscou estudar como o mercado de trabalho americano veio se comportando nos últimos 20 anos, época caracterizada pelo número crescente de inovações tecnológicas e cada vez mais influenciada pelos avanços em IA. Além disto, um segundo objetivo foi configurar diversos modelos e escolher o melhor deles para fazer a previsão dos salários semanais médios para os próximos 3 anos. Diante do que foi exposto nas seções anteriores e com posse dos resultados dos modelos de previsão, foram pontuadas as seguintes conclusões.

Os salários de todas as quatro categorias estudadas tiveram aumentos expressivos nos últimos 20 anos. Atualmente, os salários de ocupações cujas tarefas são não rotineiras cognitivas representam a classe que os salários são maiores na economia norte americana, estas ocupações são notoriamente reconhecidas pelo alto nível de qualificação profissional. As crises econômicas aparentemente não afetaram de forma severa a média salarial semanal dos trabalhadores. Apesar da crise econômica de 2007-2009 ter impactado o mercado, o crescimento dos salários pós-crise seguiu a tendência de alta pré-crise. Os primeiros dados disponíveis após a pandemia da COVID-19 também mostram que esta crise não afetou negativamente o salário mensal médio. Uma das possíveis causas deste fato é a conhecida rigidez salarial.

Se por um lado a série agregada de salários médios mensais apresenta a tendência de alta nas últimas duas décadas, a série agregada dos salários médios semanais divididos pelo PIB apresenta um comportamento oposto: um padrão decrescente que indica a necessidade cada vez menor de pagar salários para produzir. Este resultado corrobora alguns estudos empíricos desta área. Com os avanços tecnológicos cada vez maiores, a entrada de robôs no mercado de trabalho e a crescente mecanização das cadeias produtivas, a massa salarial que os trabalhadores recebem tem se tornado cada vez menor. Se por um lado isto é algo bom e representa aumento de produtividade, por outro lado representa perda persistente de renda com o passar dos anos. O claro comportamento descendente desta série é interrompido bruscamente durante a pandemia da COVID-19 e retorna para o nível de 2010. Isto aconteceu pelo fato de o PIB trimestral americano ter caído mais de 30% no segundo trimestre de 2020, além da rigidez salarial que causou um expressivo aumento na razão salário/PIB.

Diante do conjunto de dados que analisamos, espera-se que valor médio semanal dos salários (em termos absolutos) cresça para todas as classes de ocupações. Todas as previsões realizadas pelos modelos confirmaram a continuação da tendência de alta. Durante o cálculo

das previsões, a rede MLP foi o modelo que melhor desempenhou fora da amostra, curiosamente tendo desempenho melhor que a rede LSTM – própria para modelagem de séries temporais. Uma possível explicação para isto é o fato de a série temporal ser pequena e com apenas 84 observações. A rede LSTM se diferencia da rede MLP pelo fato de ter uma capacidade de memória maior, porém em séries temporais pequenas, a visão global da rede MLP (mesmo que curta) foi suficiente para modelar e projetar bem o futuro. É possível que se a série fosse maior, contendo centenas ou milhares de dados, as características peculiares da rede LSTM a tornariam expressivamente melhor do que a rede MLP nesta tarefa de previsão.

6 REFERÊNCIAS

ACEMOGLU, Daron. 1998. “Why do New Technologies Complement Skills? Directed Technical Change and Wage Inequality.” *The Quarterly Journal of Economics* 113(4): 1055–1089.

ACEMOGLU, Daron, and AUTOR, David. 2011. “Chapter 12—Skills, Tasks and Technologies: Implications for Employment and Earnings,” *Handbook of Labor Economics*. Elsevier, vol. 4b, pp. 1043–1171.

ACEMOGLU, Daron, and GUERRIERI, Veronica. 2008. “Capital Deepening and Nonbalanced Economic Growth.” *Journal of Political Economy* 116(3): 467–498.

ACEMOGLU, Daron, and RESTREPO, Pascual. 2018a. “The Race Between Machine and Man: Implications of Technology for Growth, Factor Shares and Employment.” *American Economic Review* 108(6): 1488–1522.

AUTOR, D.; LEVY, F.; MURNANE, R. The Skill Content of Recent Technological Change: An Empirical Exploration. **The Quarterly Journal of Economics**, novembro de 2003, vol. 118, issue 4, p. 1279–1333.

ACEMOGLU, Daron, and RESTREPO, Pascual. 2018b. “Demographics and Automation.” National Bureau of Economic Research w24421.

BESSEN, J. How Computer Automation Affects Occupations: Technology, Jobs, and Skills. **Federal Reserve Bank of Boston**, dezembro de 2014, no. 14 – 16.

BOX, G. E. P.; JENKINS, G. M. Time series analysis, control, and forecasting. San Francisco, CA: Holden Day, v. 3226, n. 3228, p. 10, 1976.

BISHOP, C. A. Pattern Recognition and Machine Learning. Information Science and Statistics. Springer, New York, 2006.

CIABURRO, G. and VENKATESWARAN, B. (2017) *Neural Networks with R: Smart models using CNN, RNN, deep learning, and artificial intelligence principles*. Packt Publishing Ltd, 2017

CHOLLET & ALLAIRE (2017, Dec. 20). TensorFlow for R: Time Series Forecasting with Recurrent Neural Networks. Retrieved from <https://blogs.rstudio.com/tensorflow/posts/2017-12-20-time-series-forecasting-with-recurrent-neural-networks/>

CHIACCHIO, Francesco; PETROPOULOS, Georgios; PICHLER, David (2018) : The impact of industrial robots on EU employment and wages: A local labour market approach, Bruegel Working Paper, No. 2018/02, Bruegel, Brussels.

CINAR, Y.G., MIRISAEE, H., GOSWAMI, P., GAUSSIÉ, É., AÏT-BACHIR, A., & STRIJOV, V.V. (2017). Time Series Forecasting using RNNs: an Extended Attention Mechanism to Model Periods and Handle Missing Values. *ArXiv, abs/1703.10089*.

DAMETTO, R, CÉSAR. Estudo da aplicação de redes neurais artificiais para predição de séries temporais financeiras. 2018. Disponível em: <<http://hdl.handle.net/11449/157058>> . Acesso em: 23/12/2019

FREY, C.; OSBORNE, M. The Future of Employment: How Susceptible are Jobs to Computerisation? **Technological Forecasting and Social Change**, 2017, vol. 114, issue C, p. 254-280.

G. JAMES, D. WITTEN, T. HASTIE, and R. TIBSHIRANI. *An Introduction to Statistical Learning: with Applications in R*. Springer Texts in Statistics, 2013.

HASTIE, T., TIBSHIRANI, R. & FRIENDMAN, J. (2009) "The Elements of Statistical Learning: Data Mining, Inference and Prediction", Springer, Stanford, CA, USA, ISBN: 978-0-387-84858-7.

HYNDMAN, RJ e ATHANASOPOULOS, G. (2012) *Forecasting: Principles and Practice*. OTexts. <https://www.otexts.org/fpp>

HOCHREITER, Sepp & SCHMIDHUBER, Jürgen. (1997). *Long Short-Term Memory*. *Neural computation*. 9. 1735-80. 10.1162/neco.1997.9.8.1735.

KIM, S., & KANG, M. (2019). Financial series prediction using Attention LSTM. *ArXiv, abs/1902.10877*.

KROLLNER, B., VANSTONE, B., & FINNIE, G. (2010). Financial time series forecasting with machine learning techniques: A survey. In *Proceedings of the 18th European Symposium on Artificial Neural Networks (ESANN 2010): Computational Intelligence and Machine Learning* (pp. 25-30)

LOPES, Lucas. (2018). Predição do preço do café Naturais Brasileiro por meio de modelos de statistical machine learning - *Revista Sigmae*. 7. 1-16.

MACIEL, L., e BALLINI, R. (2009). Design a neural network for time series financial forecasting: accuracy and robustness analisys.

MATTOS, R. Close Enough to a Jobless Society: Reflections on Historical Materialism and Artificial Intelligence. **Discussion Paper**, 2018, Programa de Pós-Graduação em Economia – Universidade Federal de Juiz de Fora, n. 07/2018

MATTOS, R. Modelos Arima (Metodologia de Box & Jenkins). *Notas de Aula*, 2017, Faculdade de Economia – Universidade Federal de Juiz de Fora. p. 1-36.

PERLIN, Marcelo and RAMOS, Henrique, GetHFData: A R Package for Downloading and Aggregating High Frequency Trading Data from Bovespa (November 20, 2016). Available at SSRN: <https://ssrn.com/abstract/=2824058> or <http://dx.doi.org/10.2139/ssrn.2824058>

PETNEHÁZI, GÁBOR. (2018). Recurrent Neural Networks for Time Series Forecasting.

PORTUGAL, M. S. (1995), "Neural Networks Versus Time Series Methods: A Forecasting Exercise", *Revista Brasileira de Economia*, vol. 49, n. 4, p. 611-629.

ROSENBLATT, F. (1958). The *Perceptron*: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6), 386–408. <https://doi.org/10.1037/h0042519>

SAMUEL, A. L. (1959). Some studies in machine learning using the game of checkers. *IBM Journal of research and development*, 3(3), 210-229.

TURING, Alan M. (1950). Computing machinery and intelligence. *Mind* 59 (October):433-60.

CHUNG, Junyoung, et al (2014). "Empirical evaluation of gated recurrent neural networks on sequence modeling." (2014)

SIAMI-NAMINI, Sima; TAVAKOLI, Neda; NAMIN, Akbar Siami. (2018) A comparison of ARIMA and LSTM in forecasting time series. In: 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA). IEEE, 2018. p. 1394-1401.

ZANINI, A. Regulação econômica no setor elétrico brasileiro: uma metodologia para definição de fronteiras de eficiência e cálculo do fator X para empresas distribuidoras de energia elétrica. Rio de Janeiro, 2004. 147p. Tese de Doutorado - Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro, PUC-Rio.