



MATEUS RODRIGUES SANTOS

**REDES NEURAIS ARTIFICIAIS APLICADAS À PREVISÃO
DE SÉRIES TEMPORAIS:
ESTUDO SOBRE O CAFÉ ARABICA BRASILEIRO**

LAVRAS – MG

2019

MATEUS RODRIGUES SANTOS

**REDES NEURAS ARTIFICIAIS APLICADAS À PREVISÃO DE SÉRIES
TEMPORAIS:
ESTUDO SOBRE O CAFÉ ARABICA BRASILEIRO**

Trabalho de Conclusão de Curso apresentado
à Universidade Federal de Lavras, como parte
das exigências do Programa de Graduação em
Engenharia de Controle e Automação.

Daniel Furtado Leite
Orientador

**LAVRAS – MG
2019**

**Ficha catalográfica elaborada pela Coordenadoria de Processos Técnicos
da Biblioteca Universitária da UFLA**

Santos, Mateus Rodrigues

Redes Neurais Artificiais aplicadas à Previsão de Séries Temporais : Estudo sobre o Café Arabica Brasileiro / Mateus Rodrigues Santos. 1^a ed. rev., atual. e ampl. – Lavras : UFLA, 2019.

58 p. : il.

Monografia(Conclusão de Curso)–Universidade Federal de Lavras, 2018.

Orientador: Daniel Furtado Leite.

Bibliografia.

1. TCC.

MATEUS RODRIGUES SANTOS

**REDES NEURAS ARTIFICIAIS APLICADAS À PREVISÃO DE SÉRIES
TEMPORAIS: ESTUDO SOBRE O CAFÉ ARABICA BRASILEIRO
ARTIFICIAL NEURAL NETWORKS APPLIED TO COMMODITY PRICE
FORECASTING: A STUDY ON THE BRAZILIAN ARABICA COFFEE**

Trabalho de Conclusão de Curso apresentado
à Universidade Federal de Lavras, como parte
das exigências do Programa de Graduação em
Engenharia de Controle e Automação.

APROVADA em 18 de Novembro de 2019.

Tatiane Carvalho Alvarenga	IFSUDESTE/MG
Daniele Aparecida de Oliveira Silva	UFLA
Belisário Nina Huallpa	UFLA

Daniel Furtado Leite
Orientador

**LAVRAS – MG
2019**

À Heitor, Eliana e Bruno, que tornaram realidade essa formação.

AGRADECIMENTOS

Dedico, aqui, minha eterna gratidão;

Aos brasileiros e brasileiras que pagam seus impostos que se tornam recursos para educação pública de qualidade.

Ao corpo administrativo e docente da Universidade Federal de Lavras por materializar ensino de excelência com estes recursos. Especialmente aos docentes Daniel Leite, pela orientação neste trabalho; e Felipe Silva, pelo companheirismo em outros projetos.

À minha família, especialmente pai, mãe e irmão que exercem força diária no acompanhamento de todo este caminho até aqui. Aos avôs e avós, tios e tias, primos e primas, aqueles estão ao lado nos passos desta trilha.

À Yule, que acompanhou este trabalho em seu lado pessoal, pelo amor, companheirismo e cumplicidade em partes importantes desse crescimento.

Aos amigos da turma de Engenharia de Controle e Automação 2014/1, aos quais no caminho das pedras pudemos provar que o esforço do todo é maior que a soma dos indivíduos. À todos os irmãos e irmãs da equipe de robótica TROIA, por todo aprendizado e crescimento que compartilhamos.

Aos meus amigos João Tavares e Lucas Strazza, partes fundamentais da construção do projeto que originou esta obra, por tornarem todo trabalho leve e engrandecedor. À todos companheiros de trabalho.

I propose to consider the question, 'Can machines think?'
(Alan Turing)

RESUMO

Com o advento da digitalização nos diversos setores da indústria, o gerenciamento e a adoção de uma estratégia correta de utilização de dados se tornou fundamental ao mercado. Os anos atuais marcam a intensificação das pesquisas em inteligência computacional visando a melhoria e o desenvolvimento de novas funcionalidades em ferramentas de auxílio à tomada de decisão por meio de dados. Neste contexto, o presente trabalho apresenta três abordagens baseadas em redes neurais artificiais. As redes têm o papel de aproximar dinâmicas não-lineares observadas a partir de dados indexados no tempo, i.e., a partir de séries temporais. O objetivo dos experimentos é (i) o entendimento de como esses modelos se comportam face à uma série de dados com variações de tendência e alto grau de comportamento estocástico, e (ii) a comparação de seus desempenhos frente à uma abordagem estatística tradicional. Diferentes meta-parâmetros dos algoritmos de treinamento de redes neurais são avaliados. Os modelos neurais considerados neste estudo são o Perceptron Multi-Camadas (MLP), a rede de Função de Base Radial (RBF), e Sistema de Inferência Neuro-Fuzzy Adaptativo (ANFIS). O desempenho dos modelos em estimação um tempo à frente de séries temporais relativas ao café arábica é comparado àquele de um modelo Auto-Regressivo Integrado de Médias Móveis (ARIMA). Os resultados obtidos das redes, implementadas em Python, tiveram desempenho insuficiente frente ao modelo ARIMA para o caso em estudo, e concluiu-se que o aprofundamento nas técnicas de parametrização dos modelos utilizados pode melhorar sua performance para superação deste, além da necessidade de um estudo mais aprofundado sobre as demais variáveis que influenciam no movimento daquela a ser predita.

Palavras-chave: Redes Neurais Artificiais. Predição. ARIMA. ANFIS. Café.

ABSTRACT

With the advent of digitization in a variety of industrial sectors, data management and adoption of an appropriate strategy to exploit the available data have become fundamental issues to the market. Nowadays, a clear increasing of research in computational intelligence aiming at improving and developing new functionalities to assist decision making is noticeable. This paper presents three artificial-neural-network-based approaches, which play the role of approximating the nonlinear dynamics observed from time-indexed data, i.e., from time series. The goal of the experiments is (i) to understand how the neural models behave in dealing with a sequence of data featuring time-varying trends and a high degree of stochastic behavior; and (ii) to compare their performance against that of a traditional statistical approach. Different meta-parameters are considered for the learning algorithms of the neural networks. The neural models used in this study are the Multi Layer Perceptron (MLP), the Radial-Basis-Function neural network (RBF), and the Adaptive Neuro-Fuzzy Inference System (ANFIS). The performance of the neural models in one-step-ahead prediction considering time series related to the arabica coffee is compared to that of an Auto Regressive Integrated Moving Average model (ARIMA). The results obtained from the networks, implemented in Python, underperformed the ARIMA model for the case in study, and it was concluded that the deepening in the parameterization techniques of the models used can improve its performance to overcome it, besides the need for a better performance. further study on the other variables that influence the movement of the one to be predicted.

Keywords: Artificial Neural Networks. Forecast. ARIMA. ANFIS. Coffee.

LISTA DE FIGURAS

Figura 2.1 – Séries de preços do café em fechamentos diários dos anos de 2010 a 2018.	15
Figura 2.2 – Mapa da distribuição de produção do café no Brasil.	17
Figura 2.3 – Fluxograma de desenvolvimento de um modelo ARIMA.	21
Figura 3.1 – Modelo de neurônio artificial Perceptron.	24
Figura 3.2 – Esquemático de Rede Neural MLP.	26
Figura 3.3 – Plotagem de função Gaussiana para diferentes ϵ	29
Figura 3.4 – Esquemático de Rede Neural RBF.	30
Figura 3.5 – Comparação entre os modelos MLP e RBF abordados em um exemplo de classificação de dados bidimensionais.	31
Figura 3.6 – Ilustração de conjuntos de pertinência para a variável linguística 'idade'. As curvas definem graus de pertinência para os rótulos "jovem", "maduro", e "idoso", em função da entrada numérica "idade".	33
Figura 3.7 – Esquema geral de um Sistema de Inferência Fuzzy (FIS) convencional.	34
Figura 3.8 – Arquitetura de uma rede ANFIS.	35
Figura 4.1 – Fluxograma de análise comparativa entre modelo estatístico clássico e 3 modelos de Inteligência Artificial para predição de séries temporais.	41
Figura 4.2 – Matriz de correlação dentre as séries temporais que compõem a base de dados.	43
Figura 5.1 – Gráficos de resultados de predição em dados separados para teste com pa- rametrização padrão das bibliotecas utilizadas.	50
Figura 5.2 – Gráficos de resultados de predição em dados separados para teste com pa- rametrização otimizada com busca em rede.	51
Figura 5.3 – Curvas de desempenho por número de épocas de treinamento para MLP.	52

LISTA DE TABELAS

Tabela 5.1 – Resultados da experimentação com parâmetros padrões.	49
Tabela 5.2 – Resultados da experimentação com parâmetros padrões.	51

SUMÁRIO

1	INTRODUÇÃO	11
1.1	Objetivos	12
1.2	Contribuições	13
1.3	Estrutura do Trabalho	13
2	SÉRIES TEMPORAIS	14
2.1	Mercado Brasileiro do Café Arábica	16
2.2	Modelo Auto Regressivo Integrador de Médias Móveis	18
2.3	Variáveis Relevantes	20
3	INTELIGÊNCIA COMPUTACIONAL	23
3.1	Redes Neurais Artificiais	23
3.1.1	Modelo Neural <i>Perceptron</i> Multicamadas	25
3.1.1.1	Algoritmo de Retro Propagação de Erro	27
3.1.2	Modelo Neural Função de Base Radial	28
3.2	Modelo Neuro- <i>Fuzzy</i> Adaptativo	32
3.2.1	Lógica <i>Fuzzy</i>	32
3.2.2	Sistema de Inferência <i>Fuzzy</i>	33
3.2.3	Sistema de Inferência Neuro- <i>Fuzzy</i> Adaptativo	34
3.2.3.1	Algoritmo de Aprendizado	37
3.3	Predição de Séries Temporais	39
4	METODOLOGIA	40
4.1	Fluxograma Geral	40
4.2	Bases de Dados	40
4.3	Considerações de Software	44
4.3.1	Parametrização	46
4.4	Medição de Desempenho	46
5	RESULTADOS	48
5.1	Análise Comparativa	48
5.1.1	Parametrização Padrão	49
5.1.2	Parametrização Otimizada	49
5.2	Sumário e Discussões	52
6	CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS	54

REFERÊNCIAS 55

1 INTRODUÇÃO

Imagens do mercado financeiro são marcadas pela utilização de grandes montantes de dados; análises de tabelas, gráficos, com a ideia de buscar conhecimento sobre o futuro para realizar transações de capital de forma lucrativa. A obtenção de sucesso neste escopo é intimamente ligada à capacidade de prever corretamente as tendências mercadológicas para realizar as corretas compras e vendas. Os dados em questão nesse montante compõe uma família de problemas chamada de séries temporais, caracterizada por instâncias indexadas no de tempo, com período de amostragem constante em geral.

Cada estimacão futura, neste contexto é dada por uma função dependente das instâncias em instantes anteriores, variações imprevisíveis, e influências de variáveis externas. Quanto menor a dependência dessas variáveis externas mais comum o uso de estimadores estatísticos comuns, e mais facilitada é a modelagem e reprodução da função que determina o comportamento da série. São de uso comum, e relevante de serem citados os processos estocásticos para as análises preditivas nesses tipo de dados que envolvem variáveis financeiras (NAUCK; KLAWONN; KRUSE, 1997), (HAYKIN, 1994).

Entretanto, quando se possui uma base de dados reais, observa-se que essas funções determinantes das variações dos valores no tempo são complexas, e se torna factível afirmar que sua predição com uma precisão arbitrariamente pequena é complicada. Isso se dá pelo espectro de variáveis, não apenas quantitativas sobre as quais o mercado financeiro está sujeito. Sendo essas de cunho econômico, político, relações internacionais, taxas de cambio variáveis, especificidades de produção (quando o ativo negociado está vinculado a um produto físico), dentre outros fatores. Em contrapartida, a aplicabilidade também se faz forte em diversos problemas dentro e fora do mercado financeiro, como estatísticas de trânsito, previsões do tempo, etc. (SANDMANN, 2006).

Com o advento da computação, técnicas de ciência de dados melhoraram cada vez mais as perspectivas no campo da predição de valores de séries temporais. A princípio com o simples fato dos computadores serem capazes de processar mais dados de forma extremamente eficiente, e mais recentemente com a ascensão de técnicas de aprendizado de máquina, as quais permitem a criação de algoritmos capazes de trabalhar de forma similar à mente humana, processando quantidades imensas de dados históricos para gerar modelos capazes de predizer o futuro com base nos dados presentes (WALLER; FAWCETT, 2013). A base da chamada neuro computação são as Redes Neurais Artificiais (RNA). Elas fornecem inspiração para o desenvolvimento de

modelos e algoritmos de aprendizado. Estas são técnicas de Inteligência artificial que têm contribuído de forma ampla para a resolução dos problemas de análise de dados, e que se fazem objeto de estudo deste trabalho.

Diversas contribuições no escopo de aplicação de RNAs para as séries temporais no mercado financeiro já se encontram na literatura, apesar de serem raras as aplicações em produção ajudando negociadores a tomarem suas decisões. Contudo, sabe-se que no mercado de commodities agrícolas, as variações indefinidas de preço afetam a cadeia produtiva como um todo (GUTIERREZ; ALMEIDA, 2013). Este fato foi responsável pela idealização dos contratos futuros na metodologia de negociação de ativos (RIBEIRO; SOUSA; ROGERS, 2006b). O retorno sobre investimento dos produtores, e conseqüentemente o fluxo completo de fornecimento dos produtos fica à deriva de preços que não são previstos hodiernamente; e que sofrem com variações mais acentuadas em comparação com outros ativos, principalmente pelo fato de haverem variáveis não econométricas envolvidas na equação global tais como condições climáticas em diferentes regiões do mundo (MESQUITA et al., 2000).

Tendo em vista o cenário de dificuldade dos produtores e negociadores de commodities, e a relevância, histórica e hodierna, das atividades de cafeicultura para a economia brasileira (GUTIERREZ; ALMEIDA, 2013), faz-se base para as experimentações deste trabalho os dados relacionados ao setor, visando o desenvolvimento de análises da aplicabilidade das tecnologias ascendentes em ciência de dados para o auxílio de tomada de decisões de produtores e negociadores do Café Arábica brasileiro. É importante ressaltar que a aquisição e entendimento dos dados fora feita em parceria com a Agência de Inovação do Café da Universidade Federal de Lavras (Inova Café - UFLA).

1.1 Objetivos

Este trabalho tem por objetivo comparar resultados de previsões de séries temporais dos preços de fechamento de Contratos Futuros do Café Arábica tipo 4/5 na Bolsa de Mercadorias e Futuros da BOVESPA (BM&F). São consideradas três técnicas de Inteligência Artificial, especificamente, redes neurais *feed-forward* e rede neuro-fuzzy; e uma técnica linear estatística, que será usada como padrão de comparação. As técnicas foram escolhidas dado seu contexto histórico como introdutoras de determinados conceitos matemáticos para inteligência artificial.

O propósito é concluir sobre a aplicabilidade das técnicas inteligentes em estudo, ou sobre a necessidade de se buscar algoritmos de aprendizado de máquina sofisticados para a

obtenção de uma parametrização que leve os modelos inteligentes a proverem predições mais adequadas do que aquelas oferecidas por técnicas estatísticas. Espera-se que os resultados contribuam para tomada de decisões nas negociações de commodity.

1.2 Contribuições

Espera-se, como contribuição deste trabalho, o desenvolvimento e aprimoramento de campos de estudo de ciência de dados no agronegócio; a coleta de resultados relevantes e construção de modelos preditores de valores futuros no contexto de cafeicultura; a aquisição de conclusões fundamentadas na análise comparativa de técnicas potenciais para a resolução do problema. Busca-se, assim, a integração entre campos diversos para a aquisição de conhecimentos relevantes para a ciência, tecnologia e desenvolvimento econômico.

1.3 Estrutura do Trabalho

O presente documento encontra-se estruturado, a partir do próximo capítulo, sobre a seguinte organização:

- O **Capítulo 2** remonta os conceitos básicos de séries temporais, aprofunda no histórico da série temporal objeto de estudo do trabalho, e esclarece sobre a técnica de análise preditiva estatística padrão, concluindo-se com o levantamento das variáveis que irão compor a base de dados;
- O **Capítulo 3** apresenta a fundamentação teórica das técnicas de Inteligência Computacional a serem aplicadas nos experimentos, e retoma um histórico do estado da arte desse campo de estudos;
- No **Capítulo 4** encontram-se as considerações sobre as metodologias utilizadas para construção e avaliação dos experimentos;
- O **Capítulo 5** expõe os resultados da experimentação e levanta as devidas discussões baseadas nas análises comparativas;
- Por fim, o **Capítulo 6** apresenta as conclusões, e observa holisticamente os futuros possíveis para a pesquisa.

2 SÉRIES TEMPORAIS

Ciência de dados se tornando principal foco de estudo nas áreas relacionadas à tecnologias da informação, com aplicabilidade diversa e fundamental para tomada de decisão estratégica em ambiente corporativo, industrial, governamental, etc. Essa tecnologia tem se consolidado como ambiente computacional para tratamento de informações, aprendizado e modelagem. A utilização dos registros organizados de determinada atividade, chamados de banco de dados, quando devidamente manipulados, promovem a observação de comportamentos que guiam as organizações para cumprimento de seus propósitos (SILBERSCHATZ; SUNDARSHAN; KORTH, 2016).

São diversas as manipulações que podem ser executadas em uma base de dados para diferentes objetivos. Neste trabalho, tem-se o intuito de aprofundamento nas técnicas de predição de valores futuros com base na interpretação de dados passados. Será abordado no capítulo 3 o estado da arte dessas técnicas.

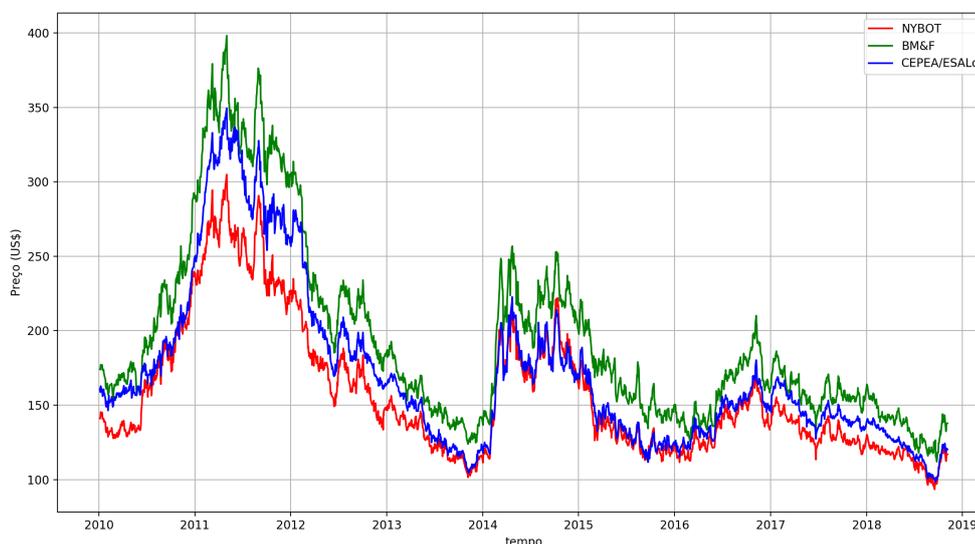
A predição de valores em uma base de dados se caracteriza pela aplicação de técnicas de predição propriamente ditas (comumente chamada de seu termo em inglês *forecasting methods*), ou de classificação, contendo diferença fundamental. As técnicas de classificação se baseiam no agrupamento das instâncias em classes, determinadas por seu conjunto de características. Dado um conjunto de instâncias, com a classe pré determinada, os algoritmos que cumprem este papel são capazes de compreender a relação dos atributos para a diferenciação entre as classes, e dessa forma "prever" a classificação de uma nova instância que é inserida naquele contexto. Esta metodologia é capaz de prever apenas se a instância faz ou não parte de um determinado grupo, levantando um exemplo ao escopo da econometria, pode-se classificar uma série de preços de uma commodity em "alta no dia seguinte" ou "baixa no dia seguinte", e classificar novas instâncias fazendo a predição da tendência dos valores de preço.

Divergente a essa abordagem, a predição de valores consiste na aproximação do valor de um determinado atributo das instâncias de um banco de dados com base em seu escopo. Em um cenário que a saída não é binária como na classificação, mas indica literalmente o valor predito para aquela variável. Esta segunda abordagem de predição de valores é o foco principal do trabalho, e o escopo de aplicação consiste em um tipo específico de bases de dados, chamado de séries temporais.

Uma sequência de observações de determinada variável ao longo do tempo, em ordem sucessiva, geralmente em períodos regulares (intervalo constante) é chamada de série temporal.

Dado exemplo do objeto de estudo deste trabalho, a série de preços e demais variáveis a serem discutidas na Seção 2.3, coletados como fechamento diário da bolsa de valores. Se compõe da medida dos dados numéricos "preço" ao longo dos dias. Neste exemplo, não se pode afirmar que os períodos são uniformes devido ao condicional de serem coletados dados apenas em dias úteis, entretanto a aproximação é aceita visto que todo o escopo do sistema se encontra na mesma condição. A Figura 2.1 seguinte demonstra as séries temporais para o fechamento do café na BM&F em reais brasileiros (R\$) nos períodos de 2010 a 2018.

Figura 2.1 – Séries de preços do café em fechamentos diários dos anos de 2010 a 2018.



Fonte: O Autor, 2019

A literatura sobre o tema, observada em Keiel e Bender (2018), identifica as series de ativos¹ como detentoras de baixa correlação linear. Possuem, de forma intrínseca, um componente estocástico² que soma grande dificuldade à obtenção de modelos confiáveis para as mesmas.

Desde a década de 70, a metodologia postulada por Box e Jenkins (1990), sumariza para a formulação dessas séries como equações matemáticas o uso de modelo Auto Regressivo Inte-

¹ É caracterizado ativo os bens, valores, créditos, direitos e assemelhados que formam o patrimônio de uma pessoa física ou jurídica, sendo avaliados pelos respectivos custos e que detém possibilidade de valorização em termos monetários (ATIVIVO, 2010)

² Componente aleatório gerado por processo estocástico, característico de funções que modelam séries temporais. Estas comumente contém um componente determinístico, facilmente modelável, linear ou não, somado ao componente estocástico.

grador de Média Móvel (ARIMA³). Esta abordagem trouxe um padrão aceitável para o estudo das movimentações macroeconômicas, e foi tomada amplamente no campo da econometria. Este modelo pode ser definido como a forma tradicional utilizada na econometria para a execução dos experimentos propostos neste trabalho. Portanto, os resultados das aplicações dos modelos de RNAs propostos serão comparados, além de mutualmente, à essa abordagem. Na Seção 2.2 aprofundar-se-á à técnica matemática.

2.1 Mercado Brasileiro do Café Arábica

Como líder na produção e exportação mundial de café, sendo responsável por mais de 30 por cento do montante mundial de quase 160 milhões de sacas de 60 Kg produzidas em 2018, o Brasil se encontra no centro do radar de importância para as movimentações relacionadas à commodity (ABIC, 2018). O faturamento bruto de 25 bilhões de reais no ano de 2018 colocou o café como o quinto ativo agrícola de maior relevância para o PIB nacional neste ano (MERCADO, 2019). Especificamente a espécie chamada Arábica⁴ ocupa 80% deste montante faturado, como pode ser visto na Figura 2.2 que demonstra a distribuição do plantio dessas culturas. Esta relevância justifica o objeto de estudo, junto ao escopo da Universidade Federal de Lavras, como polo de inovação na cafeicultura.

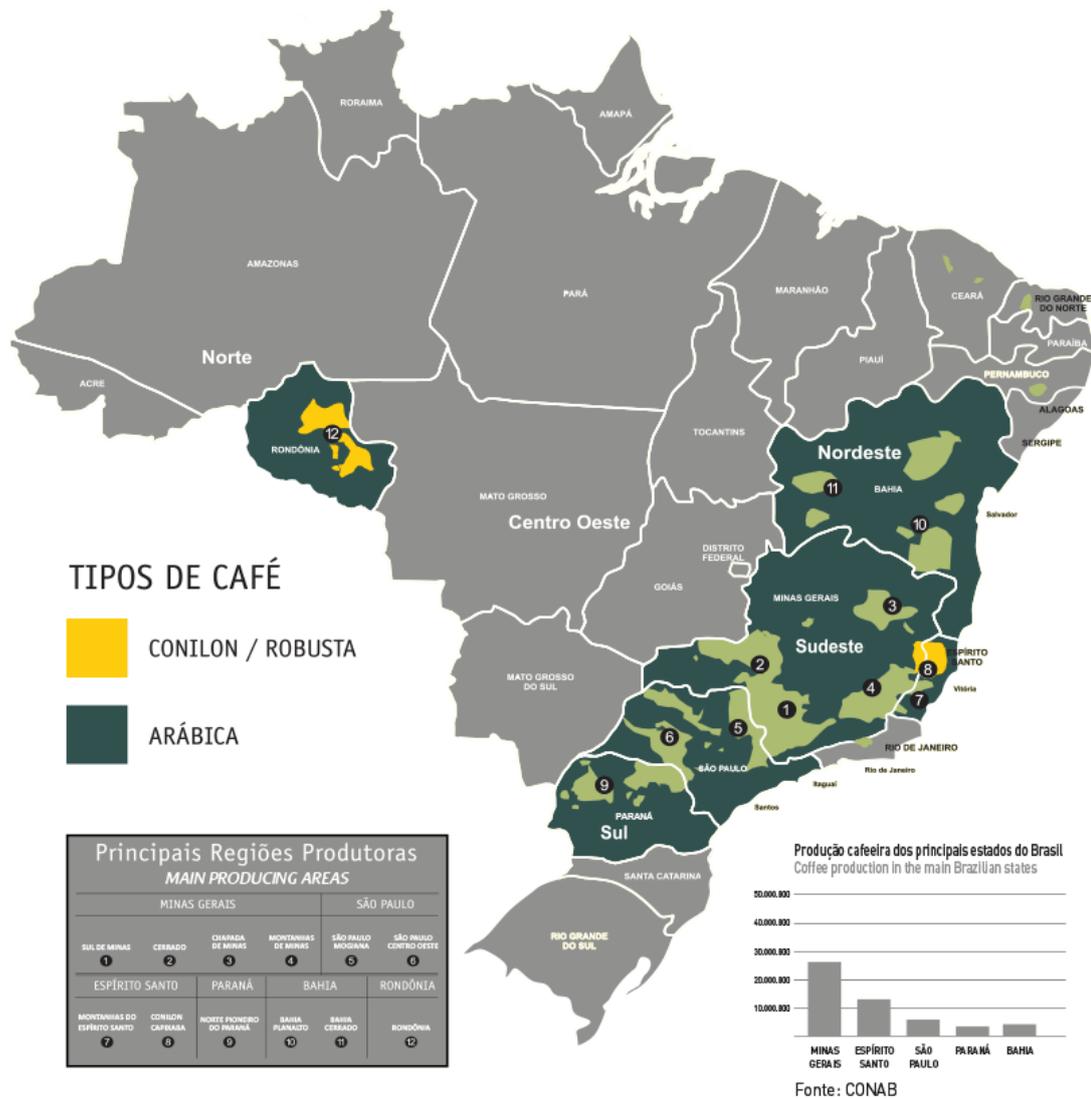
A prosperidade financeira do produto em terras brasileiras é componente fundamental da história político-econômica do país, e suas origens são descritas por Furtado (2006), que discorre como o produto se tornou a opção mais rentável de investimentos no final do século XIX, devido a fatores como:

- Pragas destruidoras de cafezais no Ceilão, África, no momento principal produtor;
- Independência dos Estados Unidos, responsável por seu desenvolvimento econômico e posicionamento como grande importador;
- Grande oferta de crédito no Brasil, propiciando financiamento de lavouras;
- Depreciação do câmbio, beneficiando a rentabilidade de exportadores;

³ Do inglês *Autoregressive Integrated Moving-Average*.

⁴ O café commodity é comercializado se diferenciando em duas espécies, Arábica e Conilon (este também chamado no mercado de Robusta). Os grãos que compõem cada espécie são oriundos de diferentes variações biológicas da planta. O café Arábica possui maior valor agregado, por ter um custo de produção maior e ser responsável pela característica de sabor na industrialização do produto final.

Figura 2.2 – Mapa da distribuição de produção do café no Brasil.



Fonte: (CONAB, 2019)

- Disponibilidade de mão de obra oriunda da grande imigração que, no momento, ocorria na América do Sul;
- E, factivelmente, as condições de solo e clima do Brasil que propiciam a prosperidade biológica das lavouras.

Entretanto, o mal funcionamento da regulação de mercado, e a dependência dos mecanismos mercadológicos externos, segundo Bacci (2007), somavam instabilidade às previsões de retorno sobre investimentos no setor. Apesar de se apresentar como base consolidada da economia nacional, a comercialização do produto sofre, historicamente, com o comportamento

volátil dos preços, característica comum no mercado de commodities no geral. Para resolução das flutuações, e conseqüente incentivo aos investidores e produtores, o governo brasileiro passou por uma série de regulamentações ao longo do século XIX, além de acordos com outros países produtores tais como a criação da OIC (Organização Internacional do Café), e a mais recente APPC (Associação de Países Produtores de Café), que objetivam buscar a elevação dos preços no mercado. O caráter histórico das políticas mundiais do mercado cafeeiro fora bem documentado por (BACCI, 2007), e demonstra a dificuldade da análise de correlações das variáveis envolvidas neste mercado, uma vez que as flutuações estão sujeitas à distúrbios políticos, climáticos, mercadológicos, que nem sempre podem ser modelados de forma intrínseca às equações matemáticas.

Nesse cenário, o objetivo deste trabalho é a comparação dos resultados da aplicação de técnicas estatísticas e inteligentes na predição dos preços do café. Especificamente o valor que deseja-se prever diz respeito ao preço da saca de 60Kg de Café Arábica 4/5 no mercado futuro⁵ da BM&F. A escolha dessa variável como objeto de estudo deste trabalho se dá por ser o valor de maior importância para as movimentações do setor no contexto de economia brasileira.

2.2 Modelo Auto Regressivo Integrador de Médias Móveis

Oriundo de (BOX; JENKINS, 1990), o modelo ARIMA se tornou a metodologia econométrica mais importante, em número de citações na literatura, para a modelagem linear e predição de séries temporais (KIA, 2012). Entendendo em partes o nome do modelo, trata-se de uma metodologia de auto regressão (AR), o que indica que a variável de interesse é analisada sobre seus estados atrasados; com médias móveis (MA), sendo sua regressão baseada na média ponderada de alguns últimos valores da série; e integrador (I), significando que os valores da série são substituídos pela diferença entre seu valor e o valor de um atraso, podendo, inclusive, ocorrer o processo de diferenciação em mais de um grau. A cada um desses termos que compõe o modelo ARIMA, são atribuídos parâmetros de adaptação *offline* para o atingimento da melhor adaptação ao escopo (LI et al., 2012).

⁵ As negociações no mercado de café são feitas basicamente de duas maneiras, a chamada venda física, onde os contratos são fechados para sacas em estoque, e o produto é entregue ao comprador no momento negociado na compra. Ou uma segunda opção chamada de contratos futuros, onde os compradores negociam sacas de café que ainda serão produzidas a serem entregues em datas futuras. O preço do café commodity na bolsa de valores é dado pelas negociações correntes em mercado futuro.

O processo em questão leva em conta que a série temporal é um processo estocástico composto de variáveis estocásticas que dependem do tempo. Nesse sentido, uma única variável da série é incerta e impossível de se modelar, enquanto que a série como um todo pode ser descrita aproximadamente por um modelo matemático correspondente. É utilizado como uma generalização do método ARMA, sua versão sem o componente integrador. Este segundo é utilizado quando a série é considerada estacionária, enquanto que no caso ARIMA, o grau do componente integrador indica o número de diferenciações necessárias para levar a série a se equiparar a um ruído branco⁶, onde é aplicado o método ARMA. Isto porque, a aplicação dos processos AR e MA dependem da característica estacionária da série, que significa, em um processo estocástico, que a distribuição de probabilidade não é alterada com o tempo. Consequentemente, a média e variância da série é a mesma para qualquer parcela de tempo ao longo da existência da série (BACCI, 2007).

Cada série temporal, de acordo com suas características, será descrita com maior fidelidade por uma parametrização diferente do ARIMA, e esta leva em conta três argumentos: q sendo a ordem da média móvel, em que uma observação de Y em um determinado instante t , Y_t , é gerada pela média ponderada dos erros de regressão por q períodos. Esses erros são, na prática, perturbações aleatórias normalmente distribuídas. O modelo pode ser descrito por

$$Y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q};$$

onde μ é a média da série, θ são os parâmetros de ponderação, e ε os termos de perturbação. Dessa forma, Y_t é influenciado por q eventos passados. Em segundo lugar, o parâmetro p é a ordem de auto regressão, que indica o número de observações passadas a se obter uma média ponderada que somada a uma perturbação aleatória no período corrente gera a observação Y_t , tal como

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \sigma + \varepsilon_t;$$

em que ϕ são os parâmetros de ponderação, e σ uma constante relacionada à média geral do processo. Por fim, d representa a ordem de integração, com significado supracitado, e uma série

⁶ Série temporal que pode ser descrita por uma distribuição normal, em que toda função amostral se aproxime de zero.

W , de diferenciação de Y tem suas observações W_t obtidas por

$$W_t = \delta^d Y_t.$$

No desmonte dessa arquitetura, modela-se um processo ARMA (p, q) , aquele de fato responsável pela descrição da série, como

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \sigma + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q},$$

e por fim, o processo ARIMA pode ser descrito pela seguinte equação:

$$\phi(B)W_t = \sigma_\theta(B)\varepsilon_t,$$

sendo B um operador de defasagem de tempo (PINDYCK; RUBINFELD, 2001).

A adequação de um ARIMA a um modelo se baseia em um fluxo conforme mostra a Figura 2.3, que passa pela identificação geral dos modelos a serem analisados; especificação do modelo, baseada em análise de autocorrelação e outros critérios; estimação dos parâmetros p , q e d por mínimos quadrados; verificação do modelo ajustado, realizada por meio da análise de resíduos para medição de sua adequação aos valores preditos; repetição no caso de não satisfação do modelo.

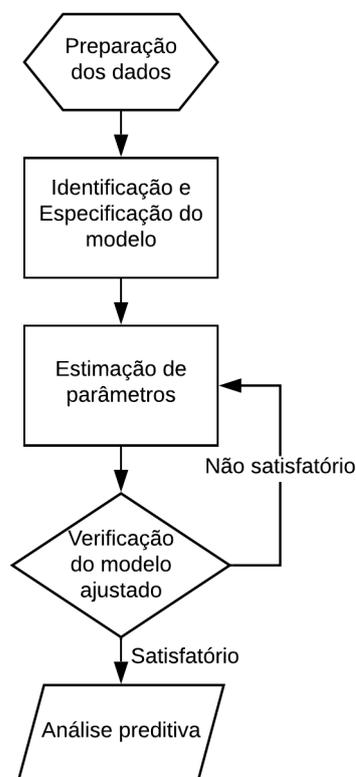
2.3 Variáveis Relevantes

Retomando o objeto de estudo central do trabalho, e visando contratos futuros de café arábica tipo 4/5 na BM&F, faz-se necessário analisar a relevância de variáveis exógenas⁷ para seu comportamento.

A análise mercadológica de commodities aponta a importância para as variações de volumes negociados nos mercados de títulos da bolsa de Nova York, relação de estoques (EST) e exportações (EXP), para o acompanhamento do preço. Tais variáveis, e seu histórico de

⁷ Por definição, são variáveis definidas por decorrências exteriores à teoria causa. Seu comportamento é exterior ao modelo estudado, entretanto seu valor possui correlação com as endógenas. Sendo estas as variáveis cujo equilíbrio se dá do próprio modelo.

Figura 2.3 – Fluxograma de desenvolvimento de um modelo ARIMA.



Fonte: O Autor

dados podem ser encontradas no repositório público CFTC⁸, o qual reporta semanalmente as posições volumétricas de negociação da bolsa de Nova York, assim como os estoques para cada commodity e as relações entre os contratos (U.S COMMODITY FUTURES TRADING COMMISSION, 2019).

Além deste relatório, a base de dados de análise do mercado cafeeiro pode ser composta pelos dados da CEPEA/ESALQ⁹, que reporta diariamente a relação de preço em dólar do café futuro na bolsa, o preço da saca de 60 Kg de café a vista no mercado brasileiro, e a variação de câmbio para o período. Dados fundamentais para a composição das análises de mercado. Na Figura 2.1 pode ser vista a variação desde o ano 2010 dos preços do café citados, e sua relação com a taxa de câmbio para os preços dados em moeda brasileira.

⁸ Comissão de negociações futuras de commodities dos Estados Unidos, do inglês *U.S. Commodities Future Trading Comission*.

⁹ Centro de Estudos Avançados em Economia Aplicada da Escola Superior de Agricultura Luiz de Queiroz.

No desmonte dessa arquitetura descrita, Minga, Alves e Parré (2016) demonstram, utilizando vetores auto regressivos, e decomposição de variâncias¹⁰ que as variáveis especulativas (aquelas referentes à volumes de negociações) e seus movimentos não estão suficientemente ligados às variações de preços nas commodities. Os autores ainda conclui que a dinamicidade dos preços e suas respectivas análises devem estar fundamentalmente baseadas nos próprios mercados e não nas relações políticas monetárias. Tem-se ainda, neste escopo, a análise de Ribeiro (2006) que buscou relações entre os preços à vista, contratos futuros, estoques e exportações, e concluiu em seu trabalho (no qual utilizou a mesma técnica de vetores auto regressivos que Minga) que as quantidades estocadas do produto, e os volumes de exportação possuem alta correlação linear com o preço a vista no mercado brasileiro, reportado pelas fontes da CEPEA. Conclusão convergente às leis de oferta e demanda, entretanto, para os contratos futuros a correlação não se aplica da mesma forma.

Dadas considerações e revisões do estado da arte, opta-se por utilizar, neste trabalho, como demais variáveis de entrada, as séries de preço de fechamento diário da bolsa de Nova York para o Café Arábica, e a cotação diária do dólar. Ambas como variáveis exógenas para compor, juntamente aos atrasos da própria série, a predição futura do objeto de estudo.

¹⁰ Metodologias econométricas de análise de séries temporais, capazes de mensurar sua co integração, ou seja, o nível em que duas séries fazem movimentos conjuntos.

3 INTELIGÊNCIA COMPUTACIONAL

3.1 Redes Neurais Artificiais

Oriundas de estudos da neuro computação, as RNAs tiveram sua primeira aparição em documentos datados de 1943, quando MacCulloch e Pitts representaram um algoritmo para um modelo computacional que simulava o funcionamento dos neurônios biológicos, criando a primeira abordagem neural artificial (AIHARA; TAKABE; TOYODA, 1990). Sua abordagem era simplista, onde se obtinha uma única saída composta do somatório das entradas; e fora construída com processos eletrônicos, utilizando-se resistores variáveis e amplificadores para simulação das conexões sinápticas.

Este campo de estudo encontrou grande ascensão ao longo das três décadas seguintes, onde apareceram conceitos de aprendizado computacional com o algoritmo de redes de Hebb, que prevê a atribuição de pesos às conexões sinápticas. O modelo de Hebb serviu de base para os avanços de sistemas adaptativos e de aprendizagem, levantando a utilização do termo inteligência artificial, associado à capacidade dos algoritmos de se adaptarem a partir de seus erros. É interessante ressaltar que o conceito de aprendizagem neste momento era simplesmente ligado ao fato da atribuição de pesos às conexões diferentes na rede neural, como na formulação da aprendizagem Hebbiana $w_{ij} = x_i x_j$, onde w_{ij} é o peso da conexão entre os neurônios i e j , e $x_i x_j$ as entradas associadas a cada um deles (SOMPOLINSKY, 1987).

Em seguida, a criação do conceito de Perceptron por Rosenblatt (1958), primeiros algoritmos para construção de classificadores lineares¹, cujo objetivo é o reconhecimento de padrões em dados. O perceptron é capaz de executar uma classificação binária, prevendo se uma determinada entrada pertence ou não à uma determinada classe, a partir do treinamento com entradas manualmente e corretamente classificadas. Este algoritmo também é baseado na atribuição de pesos às conexões da rede, seguindo a formulação que mapeia um vetor de entradas x , para um valor de saída y a partir de uma função $f(\cdot)$. Logo,

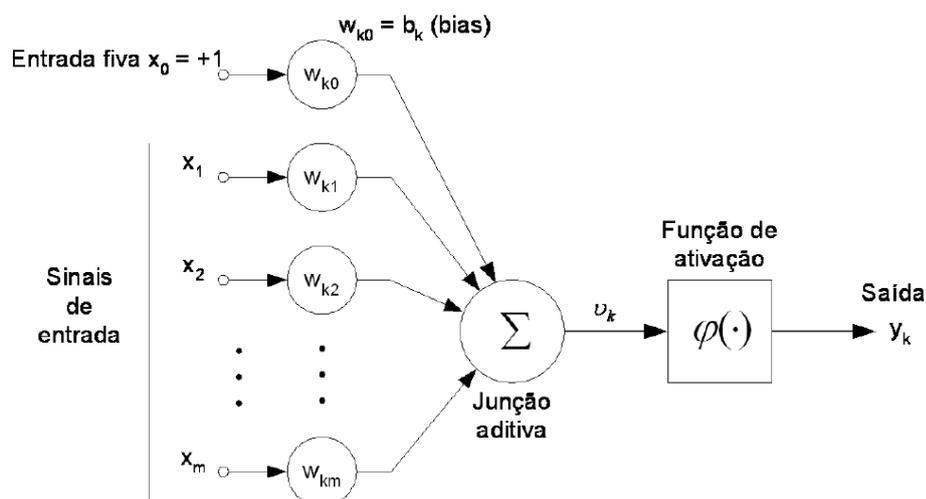
$$f(x) = \begin{cases} 1 : wx + b \geq 0 \\ 0 : \text{caso contrário} \end{cases}$$

¹ Classificadores lineares são aqueles capazes de obter resultados em padrões linearmente separáveis, que por definição, são aqueles que se encontram em lados opostos de um hiperplano.

sendo w o vetor de pesos para as entradas, e wx o produto escalar entre os vetores. O termo b é chamado de viés ². Ele é uma constante que deve ser ajustada como um parâmetro do modelo.

A Figura 3.1 mostra o modelo de um Perceptron isolado, com m entradas x_i , e seus respectivos pesos aplicados a cada entrada w_{ki} (sendo estes os parâmetros ajustáveis por algoritmos de aprendizagem); uma junção aditiva que fornece uma entrada única v_k para a função de ativação. A função de ativação modela o limiar de disparo do neurônio (SANDMANN, 2006). As funções de ativação possuem finalidade diversa, e podem ser de diferentes tipos. O aprofundamento em suas características e influência nas redes é dado nas subseções dedicadas a cada uma delas.

Figura 3.1 – Modelo de neurônio artificial Perceptron.



Fonte:(HAYKIN, 1994)

As pesquisas em cima dos modelos de Perceptron levaram às primeiras utilizações do termo aprendizado de máquina ³, na obra de (MINSKY; PAPERT, 1969), que também descobriram os principais problemas de processamento dessas redes para a época em que se encontravam. O primeiro, e mais trivial historicamente, diz respeito à incapacidade das máquinas daquele tempo processarem redes de grande escala; e em segundo lugar, a percepção de que os perceptron simples não se adequavam à operação ou-exclusivo⁴ em seu processamento.

O contexto destas limitações ocasionou desaceleração nos avanços, que voltaram à tona apenas no começo dos anos 80, primeiramente com o surgimento dos algoritmos de retro pro-

² Comumente chamado na literatura pelo seu termo em inglês, *bias*.

³ Do inglês *machine learning*, conceito amplamente utilizado hodiernamente.

⁴ Operação binária conhecida como XOR, onde o valor de retorno só é verdadeiro caso os binários comparados sejam diferentes entre si.

pagação⁵, que permitiam a execução das operações XOR, especialmente considerando treinamento em camadas múltiplas da rede de forma eficiente. Este algoritmo era capaz de distribuir o termo de erro pelas camadas da rede, uma vez que este é computado no valor de saída, e propagado para a redistribuição de pesos em cada conexão.

Os anos seguintes foram altamente produtivos para o campo, de acordo com o crescimento da capacidade computacional, em memória e processamento, novos algoritmos foram tomando seu lugar, com a utilização de técnicas diferentes para as diversas aplicabilidades que foram surgindo como demanda para os modelos e algoritmos de aprendizagem de máquina. Estruturas como as redes neurais recorrentes e convolucionais instauraram crescimento no campo com o surgimento do conceito de aprendizagem profunda⁶, entretanto a abordagem de seu funcionamento foge ao escopo deste trabalho.

Nos subtópicos seguintes a abordagem será aprofundada nos três modelos utilizados nos experimentos deste trabalho. Será discutida uma abordagem teórica para os *Perceptrons* Multicamadas (MLP⁷), uma rede neural com neurônios dotados de Funções de base Radial (RBF⁸), e uma abordagem baseada em Rede Neuro-Fuzzy Adaptativa (ANFIS⁹), fornecendo a base para o entendimento dos resultados obtidos na aplicação das técnicas para previsão de série temporal. A seleção desses três métodos se deu a partir do objetivo de uma abordagem introdutória ao universo das RNAs, visto que as redes MLP e RBF compõem as primeiras estruturas criadas nesse contexto, diferindo entre si pela sua abordagem matemática, o que enriquece as análises deste trabalho com possibilidades de aproximação a partir de funções distintas. O modelo ANFIS foi selecionado para contextualizar uma metodologia que aborda lógica *fuzzy*, de forma introdutória e matematicamente didática.

3.1.1 Modelo Neural *Perceptron* Multicamadas

A evolução dos algoritmos classificadores, apoiando-se sobre a nova perspectiva possibilitada pela técnica de retro propagação, resultou no pontapé aos estudos de aprendizagem de máquina, com a popularização do MLP. Muito popular na década de 80, estas estruturas de

⁵ Comumente chamado na literatura pelo termo em inglês *backpropagation*, o algoritmo foi proposto primeiramente por (WERBOS, 1988).

⁶ Comumente citado na literatura pelo termo em inglês *deep learning*.

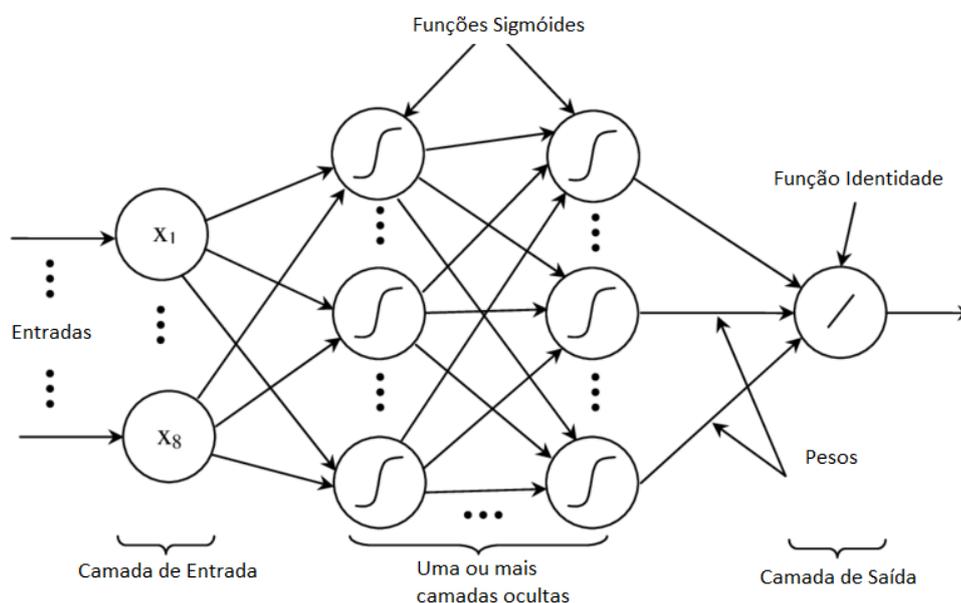
⁷ Do inglês *Multi Layer Perceptron*.

⁸ Do inglês *Radial Basis Function*.

⁹ Do inglês *Adaptive Neuro-Fuzzy Interference System*.

rede consistem em RNAs em alimentação avante¹⁰ compostas por ao menos três camadas de neurônios ligadas entre si, contendo parâmetros que ponderam as conexões. O aprendizado, nesta abordagem, refere-se à dinâmica da atribuição de pesos, permitida pelo uso do algoritmo de retro propagação de erros de estimação ao longo do treinamento. Este tipo de aprendizado é denominado aprendizado supervisionado¹¹. Devido à presença de função de ativação não-linear¹² para cada neurônio, a rede MLP é um classificador não linear de padrões. É comum assumir funções sigmóides para os neurônios das camadas intermediárias deste tipo de rede, conforme mostra a Figura 3.2.

Figura 3.2 – Esquemático de Rede Neural MLP.



Fonte: O Autor, 2019.

São camadas de neurônios deste tipo de rede, ao menos, a camada de entrada, uma camada oculta, e a camada de saída. Cada qual, composta por um número de neurônios *Perceptrons* com próprio peso e função de ativação que são constantemente adaptados ao ambiente durante o treinamento, no processo de aprendizagem. Na grande maioria das aplicações não há a necessidade do uso de inúmeras camadas ocultas, visto que uma única camada pode funcio-

¹⁰ Do inglês *feedforward*, é a técnica aplicada à RNAs sobre a qual o fluxo do sinal de entrada se propaga sequencialmente entre as camadas, e o sistema reage a um distúrbio na saída de maneira pré determinada.

¹¹ Técnica de aprendizado em que a resposta desejada é constantemente comparada à resposta observada para geração do sinal de erro. E o ajuste de parâmetros acontece sempre prezando pela diminuição desse erro.

¹² função de ativação refere-se àquela que define a saída do neurônio com base em sua ativação interna, fazendo com que a saída seja igual a $g(\sum(wx) + b)$.

nar como aproximador universal de funções (HAYKIN, 1994). Entretanto, vale-se ressaltar que a utilização de uma única camada pode ocasionar um problema de superdimensionamento na quantidade de nós desta, prejudicando a performance computacional (SANDMANN, 2006). A capacidade de aproximação universal não significa, na prática, que uma precisão arbitrária do erro de estimação será obtida, e o uso de apenas uma camada pode dificultar sua obtenção.

Na topologia ilustrada na Figura 3.2, tem-se funções sigmóides como ativadoras dos neurônios. As funções de ativação modelam o disparo do neurônio em função do somatório de suas entradas e o sinal de viés. O uso desta função não linear no tipo de rede em questão é devido à seu comportamento inspirado no processo químico-elétrico associado ao disparo de um neurônio biológico, que acrescenta maior dinamicidade e justifica o uso das múltiplas camadas, uma função linear poderia ser reproduzida por um *perceptron* único (HAYKIN, 1994). As sigmóides podem ser expressas por

$$f(v_j) = \frac{1}{1 + e^{v_j}}$$

sendo v_j o a soma ponderada das entradas mais o viés, e $f(v_j)$ a saída referente ao neurônio j .

3.1.1.1 Algoritmo de Retro Propagação de Erro

Nas redes MLPs, o sinal de entrada tem um fluxo "da esquerda para a direita", onde os valores são propagados para os neurônios da próxima camada até a saída. Sendo X o vetor de entradas e cada um de seus elementos indexados por i , o processo de propagação se inicia na camada de entrada, os sinais são propagados para a primeira camada oculta, onde cada neurônio, atuando como um *perceptron*, realiza a soma ponderada das entradas pelos seus pesos e aplica sua função de ativação. A saída de cada neurônio alimenta os neurônios da camada seguinte por N camadas ocultas que compõem a rede, até a chegada na camada de saída. Sendo cada neurônio de uma camada indexado por j , a propagação em uma camada oculta pode ser descrita por

$$Y_k = \varphi\left(\sum_{j=0}^{N_j} v_{kj} \varphi\left(\sum_{i=0}^{N_k} w_{ji} x_i\right)\right),$$

onde Y é o vetor de saída da camada, indexado por k elementos, v e w as matrizes que correlaci-

onam os pesos de cada neurônio à saída e à entrada respectivamente (DUDA; HART; STORK, 2000), (HAYKIN, 1994).

Baseado na aprendizagem por correção de erro, o algoritmo de retro propagação utilizado no modelo MLP ajusta seus pesos a partir de um sinal de diferença entre a saída de cada neurônio y_k e o valor desejado d_k , sendo

$$e_k = d_k - y_k.$$

A retro propagação, então, atua na rede no sentido "da direita para a esquerda", a partir de um vetor de entrada dado pelo erro da amostra, que consiste no erro médio quadrático dos neurônios de saída. Sendo assim, a função objetivo a ser minimizada pelo algoritmo de retro propagação é dada por

$$J = \frac{1}{N_k} \sum_{k=1}^{N_k} e^2,$$

sobre a qual é possível punir parâmetros individuais da rede de forma proporcional à sua contribuição ao erro de estimação. O erro para cada neurônio é definido por um fator de sensibilidade, que determina a direção vetorial da busca no espaço dos pesos. Este fator, para uma determinada amostra n é definido pela derivada parcial $\partial J(n)/\partial w_{ji}(n)$. Nesse escopo, defini-se a correção a ser aplicada no peso w_{ji} como Δw_{ji} , calculado por

$$\Delta w_{ji} = -\eta \frac{\partial J(n)}{\partial w_{ji}(n)},$$

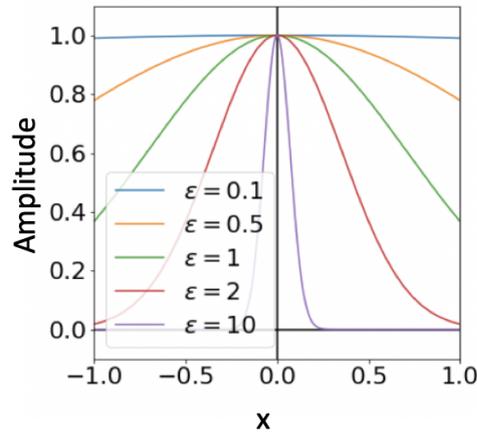
que tem η definido como a taxa de aprendizagem do modelo MLP. Em outras palavras, os parâmetros da rede neural caminham na direção do gradiente de erro até convergirem à um mínimo local da superfície de erro no espaço dos parâmetros.

3.1.2 Modelo Neural Função de Base Radial

Com relevante aplicabilidade em previsões de séries temporais e classificações, foi formulada por (BROOMHEAD; LOWE, 1988) uma RNA cuja função de ativação é uma RBF. Caracteriza RBF aquela função que depende unicamente da distância (normalmente euclidiana) a partir de uma determinada origem, sendo $\varphi(x) = \varphi(\|x\|)$. Como exemplo comum de RBF

pode-se citar a função Gaussiana $\varphi(r) = e^{-\varepsilon r^2}$, que é unicamente dependente da distância de seu centro dada por r .

Figura 3.3 – Plotagem de função Gaussiana para diferentes ε .



Fonte: O Autor, 2019.

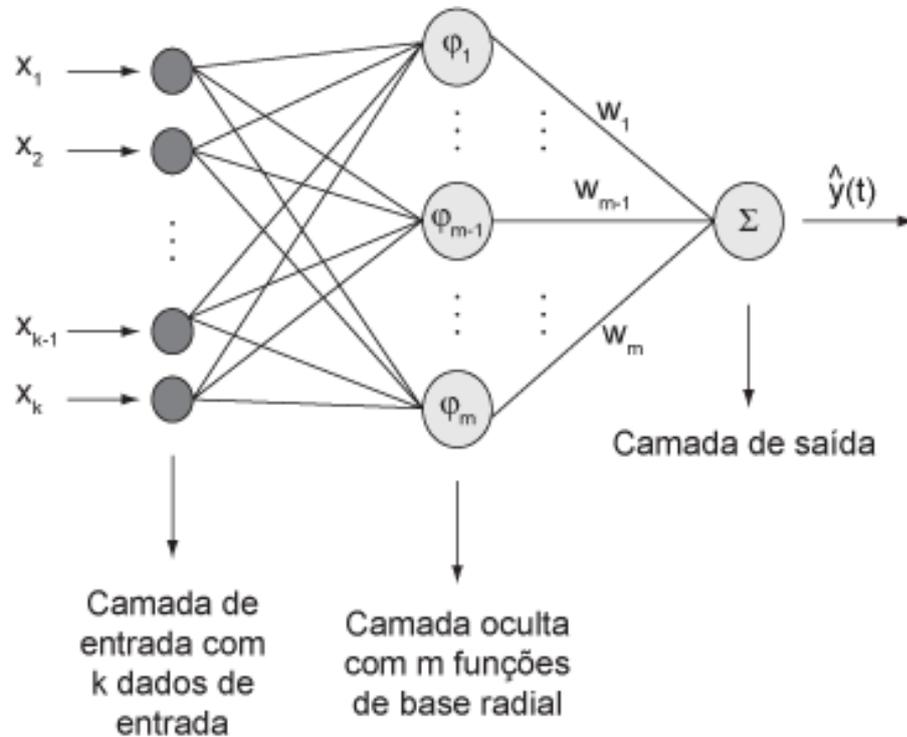
Neste modelo, o valor de saída se dá pela combinação linear das RBFs dos valores de entrada, aplicados os devidos pesos. São modelos que possuem apenas uma camada oculta (como pode ser visto na Figura 3.4, que ilustra a estrutura da rede), o que torna a convergência mais rápida que o usual em outras redes para o processamento das mesmas entradas. Apesar de possuírem apenas uma camada, é possível provar que estas funções são aproximadores universais, que são aquelas redes neurais que podem ser aproximadas por funções contínuas (CASTRO; CASTRO, 2001).

Tendo como primeira camada, a entrada representada pelo vetor x , indexado por cada elemento k . A segunda camada é composta por uma RBF φ em cada neurônio, sendo sua função de ativação. Por fim, a saída da rede é igual à soma das saídas da camada oculta, com m neurônios, ponderada pelos pesos w . Matematicamente, pode-se expressar a saída como

$$\hat{y}(x) = \sum_{i=1}^m w_i \varphi(\|x - x_i\|),$$

sendo a função de base radial uma função arbitrária que tem como termo a distância Euclidiana entre os pontos x e x_i , sendo o último o centro da respectiva função. Estes podem ser calculados com base nos dados e algum algoritmo de clusterização ou via algoritmo determinístico de primeira ou segunda ordem baseado na retro-propagação de erros.

Figura 3.4 – Esquemático de Rede Neural RBF.



Fonte: (COELHO; SANTOS; JR., 2008)

De forma matricial, a formulação da saída pode ser expressa sob a seguinte notação $y = w\Phi$, que expandida se torna

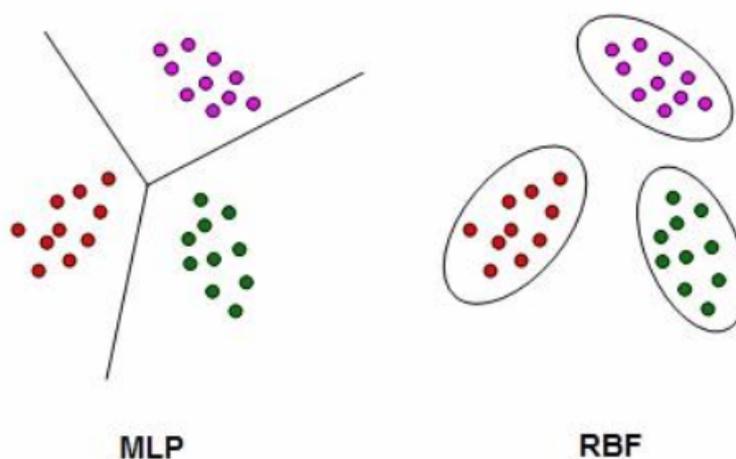
$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_m \end{bmatrix} \begin{bmatrix} \phi_{11} & \phi_{12} & \cdots & \phi_{1m} \\ \phi_{21} & \phi_{22} & \cdots & \phi_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ \phi_{m1} & \phi_{.m} & \cdots & \phi_{mm} \end{bmatrix},$$

onde o valor de m pode ser grande dependendo do modelo a ser treinado, tendendo a se aproximar ao tamanho do conjunto de amostras de treinamento. No desmonte dessa arquitetura, o aprendizado nesse tipo de RNA é feito pelo cálculo da matriz de pesos $w = \Phi^{-1}d$, desde que a matriz de interpolação Φ possua uma inversa, ou seja, seja não-singular (HAYKIN, 1994).

Nesse ínterim, é relevante ressaltar as principais diferenças funcionais entre esta metodologia e a técnica de MLP supracitada. É importante perceber a capacidade exploratória e consequentemente de filtragem de características que a MLP possui, principalmente contando com sua versatilidade no acréscimo de quantas camadas ocultas forem necessárias para o bom

funcionamento do modelo. Dado esse escopo, as MLPs se sobressaem em performance¹³ em cenários onde as características críticas dos dados estão embarcadas em grandes vetores de entrada, tais como em modelos de reconhecimento de imagem. Por outro lado, as RFBs, com seu processamento mais enxuto, possuem convergência mais veloz em sua única camada oculta. Porém sua aplicabilidade com performance satisfatória se encontra em vetores de entrada menores, cuja correlação com a saída seja mais direta possível (SOARES; TEIVE, 2015). A Figura 3.5 demonstra de forma superficial como uma tarefa de classificação é realizada de forma diferente nas redes MLP e RBF. É possível perceber o efeito da função radial definindo curvas de nível em torno de um ponto focal, em contrapartida à definição de bordas das regiões por semi retas no plano, no caso da representação bidimensional, ocasionada pelos *perceptrons*.

Figura 3.5 – Comparação entre os modelos MLP e RBF abordados em um exemplo de classificação de dados bidimensionais.



Fonte: (INC, 2017)

Foram então apresentadas duas metodologias clássicas para a resolução de problemas de predição em séries temporais. Na próxima subsecção será dissertado sobre uma abordagem mais contemporânea, para enriquecimento da análise de diferentes resultados no propósito deste trabalho.

¹³ Afirmação oriunda de análise de diferentes resultados comparativos dentre as duas tecnologias (SOARES; TEIVE, 2015) orientada ao contexto da aplicação, não sendo tomada como regra geral

3.2 Modelo Neuro-Fuzzy Adaptativo

3.2.1 Lógica Fuzzy

O conceito dos conjuntos *fuzzy*¹⁴ se relaciona à capacidade de classificação de uma instância em um contexto. Para tal, são necessárias informações discriminantes da instância, representadas matematicamente por um vetor de dimensionalidade, que são chamados nesse escopo de variáveis *fuzzy*, e uma instância utiliza uma ou mais variáveis fuzzy para ser classificada. Uma variável fuzzy contém um ou mais graus de pertinência, que são seus conjuntos domínio, definidos de acordo com sua natureza e necessidade de granularidade da aplicação. Tais graus de pertinência são definidos por funções de pertinência, que representam o valor de inferência da instância a ser classificada. Sendo uma variável *fuzzy* F descrita como

$$F = \{(x, \mu_F(x)) | x \in X\}$$

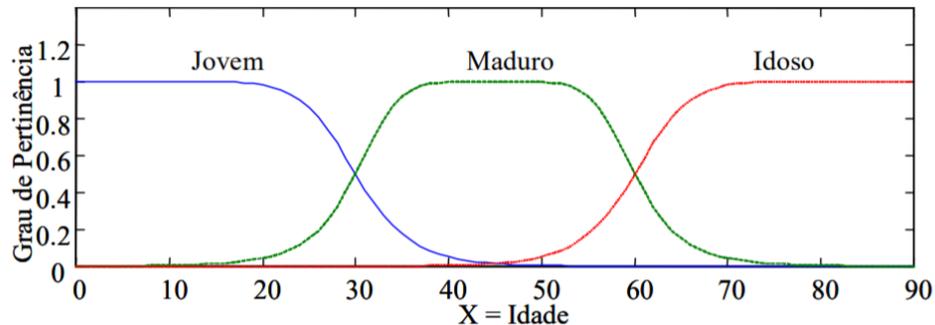
sendo $\mu_F(x)$ uma função pertinência do conjunto F para um x atributo do conjunto X de atributos da instância (JANG; SUN; MIZUTANI, 1997), (SANDMANN, 2006).

Na teoria clássica de conjuntos, um elemento pertence ou não a um determinado conjunto. Já na teoria *fuzzy*, um elemento possui grau de pertinência em um conjunto, podendo possuir graus de pertinência diferentes para diversos conjuntos como ilustra a Figura 3.6. Cada função de inferência possui um valor linguístico associado, ou seja, um rótulo que auxilia na representação de um conceito.

Operações lógicas também podem ser executadas nos conjuntos *fuzzy* assim como nos conjuntos clássicos, tais como união, intersecção e complemento. Em (ZADEH, 1996) pode ser encontrada as descrições detalhadas das operações e suas discrepâncias com os conjuntos clássicos, cujo aprofundamento foge do escopo deste trabalho. Zadeh ainda propõe o conceito de variáveis linguísticas, para aproximar o modelo de pensamento humano, facilitando a manipulação de informações que possuem características de classificação qualitativas. Essas variáveis, conforme supracitado, são "rótulos" que contribuem na expressão de informações, no lugar de números discretos precisos.

¹⁴ Será utilizado o termo em inglês *fuzzy* devido à sua comum utilização mesmo em trabalhos na língua portuguesa, facilitando indexação do termo, e reduzindo a indefinição entre conjuntos nebulosos e conjuntos difusos existentes no português.

Figura 3.6 – Ilustração de conjuntos de pertinência para a variável linguística 'idade'. As curvas definem graus de pertinência para os rótulos "jovem", "maduro", e "idoso", em função da entrada numérica "idade".



Fonte: (SAMPAIO; OLIVEIRA; IGNACIO, 2007)

A possibilidade de manipulação desses rótulos leva à criação de regras lógicas para seu processamento, tal como a regra **se-então fuzzy**, a qual atua juntamente às relações dos conjuntos *fuzzy* para gerar sistemas de processamento de variáveis tais como o Sistema de Inferência Fuzzy (FIS¹⁵) abordado na seção 3.2.2, que é base para o modelo neural a ser utilizado nas experimentações. A regra se-então *fuzzy* é fundamentada pela expressão

se x é A então y é B ,

onde " x é A " é o antecedente, " y é B " é sua conclusão, x e y são inferências dos elementos no conjunto, e A e B são rótulos de funções de pertinência.

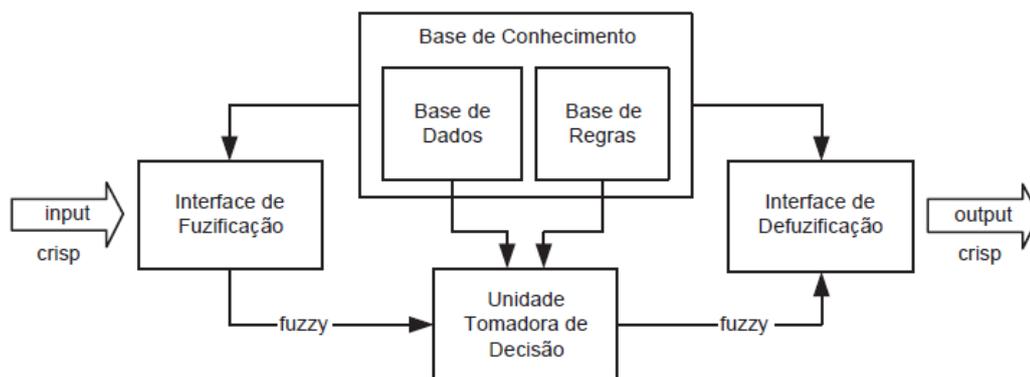
3.2.2 Sistema de Inferência Fuzzy

Visando permitir a implementação de aplicações que consumem as regras de relações dos conjuntos e regras se-então *fuzzy*, um FIS é um sistema de controle que tem como entrada valores numéricos exatos (*crisp*), e utiliza-se de funções de pertinência presentes em cada uma de suas variáveis *fuzzy* para transformar essas entradas em valores de inferência de variáveis linguísticas. Essa operação é chamada de **fuzificação**. A partir de então, é possível realizar operações com os valores de entrada baseadas em regras se-então *fuzzy*, manipulando os dados em um processamento linguístico que é baseado em uma base de dados e regras que representam o "conhecimento" do sistema. Por fim, um processo de **defuzificação**, onde os valores das variáveis linguísticas são interpretados para valores *crisp*, é aplicado para o retorno da saída do

¹⁵ Da sigla em inglês *Fuzzy Inference System*.

sistema. O FIS pode ser interpretado como um *framework* básico para implementação de soluções com lógica *fuzzy*, tais como a previsão de séries temporais, entre outras (SANDMANN, 2006). A Figura 3.7 ilustra o fluxo explicado.

Figura 3.7 – Esquema geral de um Sistema de Inferência Fuzzy (FIS) convencional.



Fonte: (SANDMANN, 2006)

3.2.3 Sistema de Inferência Neuro-Fuzzy Adaptativo

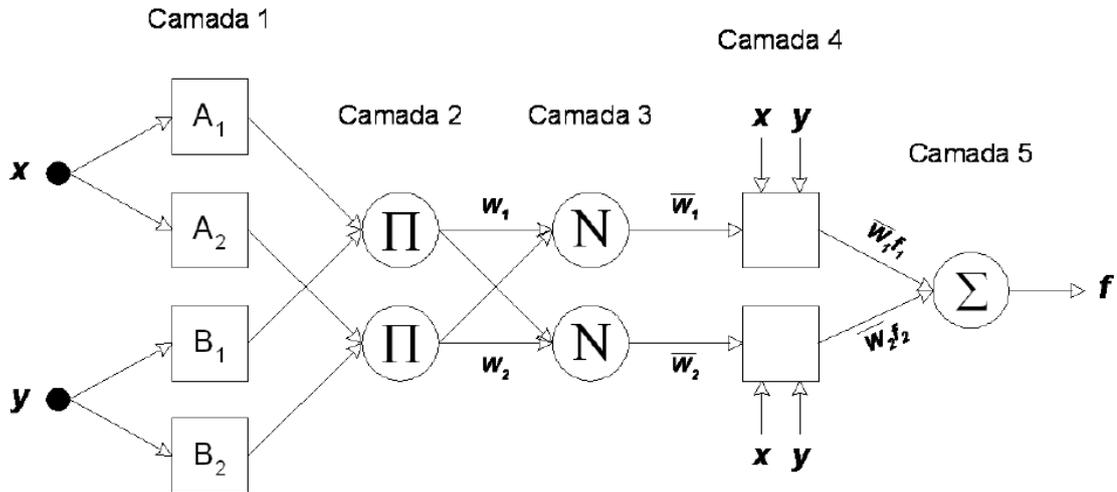
Com ampla aplicação nas áreas de controle de sistemas dinâmicos e previsão de funções não lineares, o ANFIS foi pioneiro nas arquiteturas neuro-*fuzzy*, proposto por (JANG, 1993). Esses sistemas funcionam com o processamento de variáveis linguísticas como os demais de engenharia *fuzzy*, entretanto o ANFIS é implementado como um sistema híbrido incorporado¹⁶ sob o domínio da arquitetura de RNAs. Nesse escopo, ANFIS podem processar variáveis linguísticas com capacidade de se ajustar e otimizar uma solução para o ambiente em que está inserido, ou seja, incorpora as capacidades de aprendizado das RNAs, o que justifica suas aplicações supracitadas.

O fluxo de informações em uma arquitetura neuro-*fuzzy* adaptativa pode ser ilustrada como em uma RNA, por camadas, cada camada indexada por l , cada camada contendo seus respectivos nós indexados por i , que são neurônios que possuem suas funções de ativação de acordo com a funcionalidade da camada a qual ele pertence. Essa rede recebe como entrada variáveis *crisp* do mundo real, que são convertidas em variáveis linguísticas e retomadas ao final ao seu formato nos conjuntos clássicos. A dissertação sobre o funcionamento do fluxo se

¹⁶ Modelos híbridos incorporados são aqueles que possuem alto grau de acoplamento entre os paradigmas utilizados, utilizado no projeto de sistemas que suprem falhas de um paradigma com funcionalidades de outros. Tal grau de acoplamento leva a fazer com que seja indistinguível a atuação de um paradigma sobre outro dentro do sistema.

dará com base na Figura 3.8, onde as entradas são as variáveis x e y , e a saída é a função $f(x,y)$, a seguir.

Figura 3.8 – Arquitetura de uma rede ANFIS.



Fonte: (JANG, 1993)

- **1ª Camada:** As entradas da rede passam a princípio pela camada de fuzzificação, onde os sinais *crisp* alimentam neurônios associados a variáveis linguísticas. As funções de pertinência $\mu_{A_i}(x)$ e $\mu_{B_{i-2}}(y)$ são as funções de ativação dos neurônios dessa camada, e determinam o grau de pertinência para as entradas para cada rótulo linguístico associado. Essas funções podem ser de tipos variados, tais como trapezoidais, por exemplo,

$$\mu(x) = \frac{1}{1 + \left(\frac{x-c}{a}\right)^2)^b},$$

onde a , b e c são parâmetros ajustáveis durante o processo de aprendizagem da rede. A saída de cada neurônio associa a pertinência daquela entrada àquela variável linguística (MATHUR; GLESK; BUIS, 2016)¹⁷.

- **2ª Camada:** Chamada de camada de pertinência, é nesta onde são aplicadas as operações da base de conhecimento FIS. Sendo assim, aqui se encontram as regras e conhecimento para o sistema, e cada nó é equivalente a aplicação de uma regra se-então *fuzzy*. A operação matemática equivalente para cada regra é o produto dos graus de pertinência de

¹⁷ Referência bibliográfica para este e demais tópicos desta lista.

entrada, resultando nas saídas ω_i , que são os parâmetros de consequência do FIS. A operação é feita tal como:

$$\mu_{A_i}(x)\mu_{B_i}(y).$$

Esta é a camada mais crítica nos problemas de dimensionalidade dos ANFIS, visto a necessidade de um nó para cada parâmetro de premissa, para cada regra associada, por isso é importante um adequado tratamento dos dados e da base de conhecimento para a viabilidade de processamento destes sistemas (SANDMANN, 2006).

- **3ª Camada:** Cada nó desta camada normaliza os valores dos parâmetros de consequência, evitando uma distorção destes para uma posterior ponderação. A saída de cada nó pode ser modelada como a taxa da consequência daquela entrada sobre a soma de todas as consequências, tal como

$$\bar{\omega}_i = \frac{\omega_i}{\sum_{k=1}^N \omega_k},$$

sendo N o número de regras se-então *fuzzy* aplicados na 2ª camada. A partir de então, as saídas dessa camada são os sinais normalizados de cada parâmetro de consequência.

- **4ª Camada:** Sendo esta a camada de defuzificação, seu papel é calcular o valor individual de saída a partir das regras de inferência, matematicamente por meio dos parâmetros de consequência normalizados. Cada nó desta camada recebe seu respectivo $\bar{\omega}_i$ normalizado, além da re alimentação das entradas, executando um cálculo equivalente a

$$f(\bar{\omega}_i, x, y) = \bar{\omega}_i(p_i x + q_i y + r_i),$$

de forma que p_i , q_i e r_i são parâmetros adaptativos, fazendo desta, a camada adaptativa do ANFIS.

- **5ª Camada:** Por meio do somatório de todos os valores obtidos na 4ª camada, a camada de saída realiza o processo de defuzificação, obtendo um valor único no universo *crisp* correspondente às entradas. Seu modelo é simplesmente

$$\sum_i \bar{\omega}_i f_i.$$

Na estrutura apresentada podem haver variações relevantes, tais como uma rede de 4 camadas onde a 3ª e a 4ª camada descritas são reduzidas a uma única ponderação média (SANDMANN, 2006). Por fim, é importante salientar que a arquitetura apresentada se baseia no modelo FIS Takagi-Sugeno (TAKAGI; SUGENO, 1983), apesar de ser possível sua implementação a partir de outros modelos FIS encontrados na literatura. A diferença será principalmente na operação de parâmetros consequentes nas regras se-então *fuzzy* e no processo de desfuzificação (JANG; SUN; MIZUTANI, 1997).

Sistemas ANFIS tem processo de aprendizagem lento, e o número de neurônios para tornar a rede viável cresce exponencialmente ao número de variáveis de entrada e de regras se-então *fuzzy* a serem consideradas. A superação desse problema é proposta em forma de pré processamentos, como pode ser estudado a partir de (MITRA; HAYASHI, 2000), porém sua aplicação não é recomendada para aplicações com grande número de variáveis de entrada. A seguir será apresentado o processo de aprendizagem que traz a esses sistemas interface com as RNAs.

3.2.3.1 Algoritmo de Aprendizado

A customização de um ANFIS para o escopo de aplicação se da por meio de sua capacidade de ajustar pesos otimizando o resultado em convergência a uma base de dados de treinamento. Por isso seu paradigma é aproximado ao das RNAs, e isso é feito por meio de um processo de aprendizagem. Entretanto, de forma divergente à estas, os ANFIS atuam sobre os parâmetros ajustáveis das funções de pertinência, chamados parâmetros de premissa, e não sobre os pesos das interligações de seus neurônios. A modificação desses parâmetros irá, na prática, aferir as curvas de pertinência para a classificação das variáveis *fuzzy*, com base nas regras se-então *fuzzy* aplicadas aos dados de treinamento (SANDMANN, 2006).

Predominantemente utilizado nesse escopo, o algoritmo de aprendizado híbrido (JANG, 1993) é composto da realização de um aprendizado matricial para atingir os valores de ajuste até determinada camada, método similar ao utilizados nas RBFs, dissertado na subseção 3.1.2. Além de realizar também um passo de retropropagação para ajuste dos parâmetros de premissa, explicado na subseção 3.1.1.1.

Para o ajuste dos pesos, inicia-se pelo entendimento da relação entre a matriz de pesos W e os parâmetros de consequência em uma matriz X , para as quais o valor de saída do ANFIS f é descrito por

$$f = XW = (\bar{w}_1x)p_1 + (\bar{w}_1y)q_1 + (\bar{w}_1)r_1 + (\bar{w}_2x)p_2 + (\bar{w}_2y)q_2 + (\bar{w}_2)r_2,$$

para a rede descrita na Figura 3.8. Então, para o cálculo dos valores desejados de w , e a obtenção da diferença do desejado para o obtido, é utilizada a equação acima, transformada com base em X ser uma matriz inversível, resolvendo W para uma pseudo-inversa, tal como:

$$W = (X^T X)^{-1} X^T f$$

Cada elemento do vetor da Camada 4, dado por w_i , é obtido então em W . A partir de então o erro pode ser retropropagado para as demais camadas, e obtido valor de correção de cada parâmetro das curvas de pertinência das variáveis *fuzzy*, que são dados, de forma genérica representados por α , por

$$\Delta\alpha = \eta \frac{\partial E}{\partial \alpha} = \eta \sum_{p=1}^P \frac{\partial E_p}{\partial \alpha},$$

sendo P o número de componentes no vetor de entradas da rede, E de erro geral dado pelo somatório dos erros de f das saídas da camada 4, e η a taxa de aprendizagem do ANFIS, que pode ser expressa por

$$\eta = \frac{k}{\sqrt{\sum \alpha \left(\frac{\partial E}{\partial \alpha}\right)^2}},$$

sendo k o valor de tamanho do "passo", que define o número de transições de gradiente, e na prática, define a velocidade de convergência da rede. A formulação matemática completa de cada passo da aprendizagem é apresentada em (JANG, 1993) e (SANDMANN, 2006).

3.3 Predição de Séries Temporais

Assim como fora citado para as RNAs abordadas nesse estado da arte, MLP e RBF, os ANFIS tem capacidade de atuação como aproximador universal de funções. Com um suficiente número de regras se-então essas redes podem atender ao teorema da aproximação universal e serem exploradas como preditores de funções não lineares (JANG; SUN; MIZUTANI, 1997).

Essa capacidade será explorada para as três metodologias revisadas, visando definir valores futuros para uma série temporal com alto grau de estocasticidade, como fora abordado no capítulo 2, de histórico do preço do café Arábica em contratos futuros no mercado brasileiro. O próximo capítulo, de número 4, aborda a construção dos experimentos para obtenção dos resultados de predições de cada um dos três métodos, visando a conclusão de sua performance como estimadores de funções, no escopo dos dados em estudo.

4 METODOLOGIA

Sob a perspectiva da fundamentação teórica levantada até aqui, o presente capítulo remonta as considerações do ambiente de experimentação proposto, levantando uma ilustração do processo de experimentação utilizado; considerações sobre o banco de dados por completo, e sua preparação para os modelos; análise da montagem do software responsável pelo processamento dos modelos, otimização dos parâmetros, e montagem dos resultados; e revisão das metodologias estatísticas utilizadas para quantificação e comparação intra e entre modelos.

4.1 Fluxograma Geral

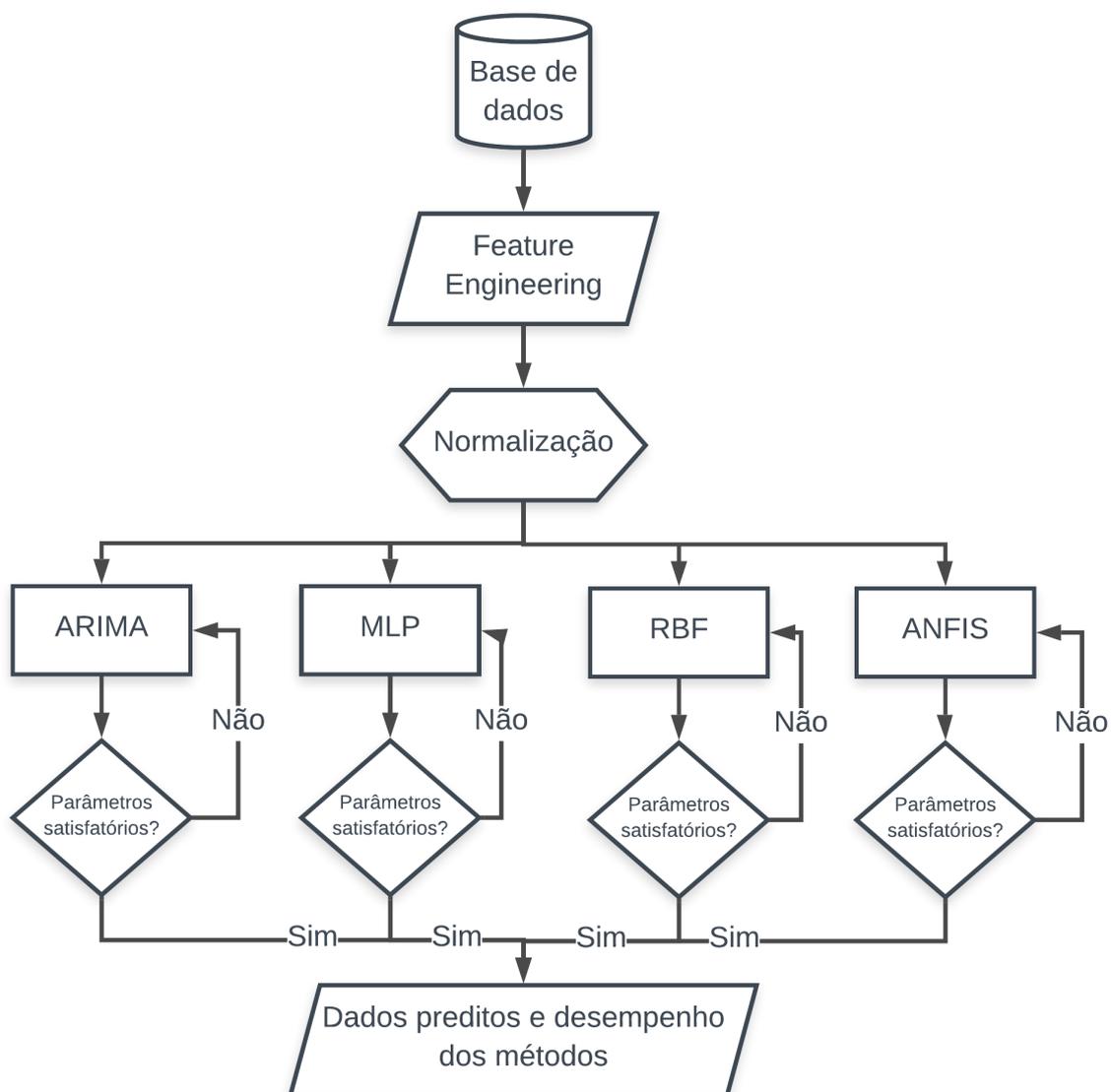
Como ilustrado na Figura 4.1, a partir de uma base de dados públicos (arquitetada em conjunto com a Inova Café - UFLA, e explicada na seção 4.2), é feita uma análise de *feature engineering*¹ baseada na literatura que aborda variáveis de maior influência na cotação da commodity, conforme abordado na seção 2.3. Os dados a serem utilizados na predição são normalizados, e enviados para cada um dos quatro modelos. É então que a camada de treinamento ajusta a rede e retorna a métrica de desempenho para a parametrização. A partir deste passo, a análise compõem-se de uma camada de busca em rede, que buscará em um conjunto de combinações de parametrizações a combinação que retorna melhor desempenho segundo uma métrica escolhida, e a camada de treinamento retorna a métrica de desempenho para cada parametrização utilizada. Então, de forma iterativa, encontra-se parâmetros ótimos dentro do universo de busca. Por fim, o algoritmo retorna os dados de desempenho de cada modelo para análise comparativa.

4.2 Bases de Dados

A base de dados utilizada fora montada compondo-se de algumas tabelas, e cada um de seus atributos será descrito a seguir. Como uma série temporal, a indexação dos elementos é feita sempre pela data, sendo assim, esta é a chave única entre as instâncias de cada tabela. Tendo sido coletados dados em bases diárias, esses identificadores de cada instância são um dia,

¹ O termo é utilizado em inglês por ser o usual da literatura, visto a discordância em sua tradução literal. Seu significado é, de forma simplificada, a escolha de determinados atributos das instâncias do banco para a análise preditiva; e o consequente descarte dos demais, visando melhoria de performance e acurácia, com a utilização de atributos realmente correlacionados à variável a ser predita (KOEHRSEN, 2018).

Figura 4.1 – Fluxograma de análise comparativa entre modelo estatístico clássico e 3 modelos de Inteligência Artificial para previsão de séries temporais.



Fonte: O Autor, 2019.

que não se repete. A partir da união de todas as tabelas pela data gera-se o banco completo com as variáveis coletadas. Os dados presentes na base são:

- Tabela de preços e variações das bolsas de Nova York e BM&F (INVESTING.COM, 2019), (NOTÍCIAS AGRÍCOLAS, 2019):
 - Data;
 - Preço do Contrato Futuro de Café Arábica na NYBOT no fechamento do dia;
 - Variação do fechamento do dia na NYBOT em relação à média mensal;

- Preço do Contrato Futuro de Café Arábica na BM&F no fechamento do dia;
 - Variação do fechamento do dia na BM&F em relação à média mensal.
- Tabela de preços e variações diárias do mercado físico brasileiro (CEPEA ESALQ, 2019):
 - Data;
 - Fechamento do preço do Café Arábica no mercado físico em R\$;
 - Variação percentual do fechamento em R\$ em relação ao dia anterior;
 - Fechamento do preço do Café Arábica no mercado físico em US\$;
 - Variação percentual do fechamento em US\$ em relação ao dia anterior;
 - cotação do dólar em R\$.
- Tabela de indicadores de posição de mercado, retirada do relatório semanal² dos negociantes de commodities (U.S COMMODITY FUTURES TRADING COMMISSION, 2019):
 - Data;
 - Contratos sobre posse de bancos de investimento em posição comprado;
 - Contratos sobre posse de bancos de investimento em posição vendido;
 - Diferença entre número de contratos sobre posse de bancos de investimento em posição comprado e vendido;
 - Contratos sobre posse de produtores da commodity em posição comprado;
 - Contratos sobre posse de produtores da commodity em posição vendido;
 - Contratos sobre posse de "outros" em posição comprado;
 - Contratos sobre posse de "outros" em posição vendido;
 - Diferença entre número de contratos sobre posse de "outros" em posição comprado e vendido;
 - Contratos negociados em posição comprado;
 - Contratos negociados em posição vendido;

² Estes dados são expostos *online* semanalmente, sendo assim, foi inferido um segurador de ordem zero para imputação dos valores ausentes do decorrer da semana.

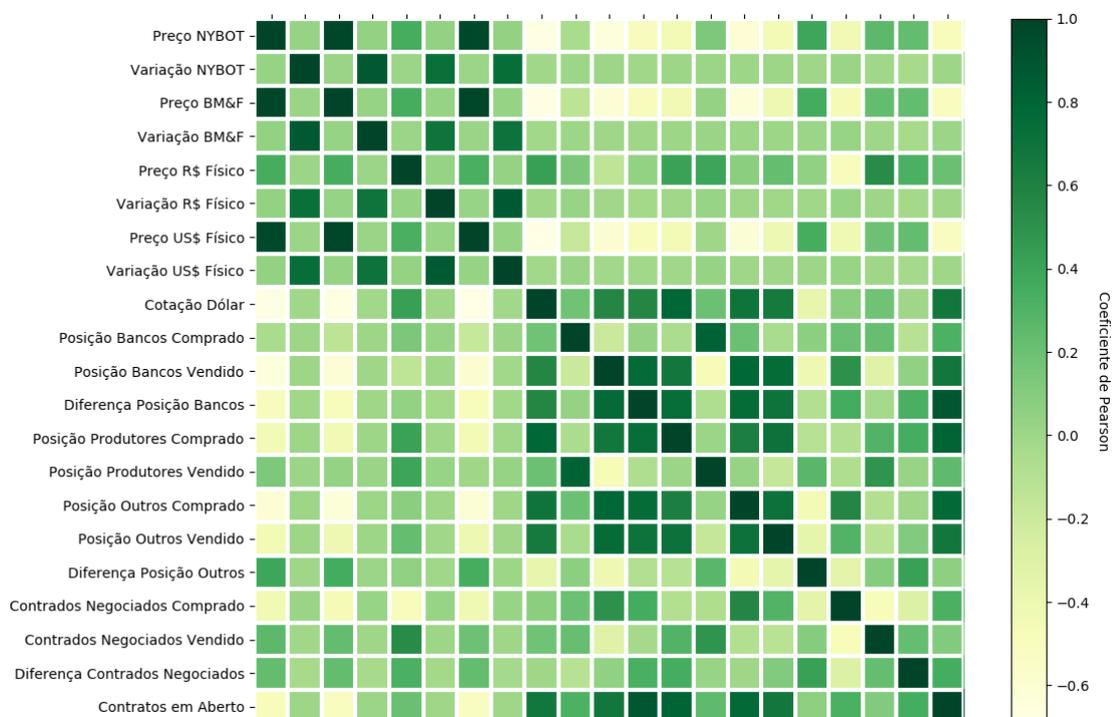
- Diferença entre número de contratos negociados em posição comprado e vendido;
- Número de contratos em aberto.

No desmonte dessa arquitetura, é possível traçar uma simplificada matriz de correlações entre as variáveis, para analisar simultaneamente, por meio do coeficiente de Pearson, a relação dentre as variáveis e comparar com a teoria das variáveis relevantes encontrada na literatura supracitada. O coeficiente de Pearson pode ser calculado como

$$\rho = \frac{\sum_{i=0}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=0}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=0}^n (y_i - \bar{y})^2}},$$

sendo x e y dois vetores de tamanho n , o coeficiente indicará um valor de correlação entre os dois vetores (BECKER; CHAMBERS; WILKS, 1988). Variáveis correlatas exercem influência uma na outra, fazendo com que aquelas que mais se correlacionam à variável a ser predita são importantes como entradas exógenas em seus modelos de predição. A matriz de correlação para esta base de dados se encontra na Figura 4.2.

Figura 4.2 – Matriz de correlação dentre as séries temporais que compõem a base de dados.



Fonte: O Autor, 2019.

Esta Figura representa um mapa de calor, onde a cor é relacionada a um valor (descrito na legenda), e pode ser interpretada como uma matriz. Onde a variável em cada linha possui

tal correlação de Pearson com a variável em cada coluna. Inclusive observa-se que a matriz é simétrica em sua diagonal. As altas correlações indicam uma proporcionalidade direta das variáveis ao longo da série, enquanto que correlações negativas indicam proporcionalidades inversas.

Esta análise ascende evidências para os relatos de (MINGA; ALVES; PARRÉ, 2016) e (RIBEIRO; SOUSA; ROGERS, 2006a), que sugerem o uso somente dos históricos de preço para regressão do preço futuro, sendo a própria variável sua melhor base de predição. Visto esse escopo, os valores de atraso das séries de preço do contrato futuro do Café Arábica serão utilizadas como entrada nos experimentos, assim como os preços dos contratos futuros da Bolsa de Nova York, e o preço do café físico em dólares, que se apresentam com coeficiente de Pearson maiores que 0,9. A base de dados completa conforme descrita, fora organizada e publicada juntamente ao repositório do software de experimentação conforme consta na Seção 4.3.

Por fim, determinada a variável a ser usada como entrada para os modelos, a divisão da série foi feita de acordo com a seguinte proporção: dada a série de preços do café de 2010 à 2018, os primeiros 85% primeiros dados sequencialmente foram separados para treinamento das RNAs, os últimos 15% foram colocados em *holdout* para testes. Importante ressaltar que os dados de *holdout* não são vistos pelo modelo durante nenhuma fase de treinamento. Dentre os dados de treinamento, 90% foram aqueles de treinamento de fato, e 10% foram utilizados para validação ao longo de todas as épocas de treino das redes. Esta divisão se deu distribuída aleatoriamente entre os dados.

4.3 Considerações de Software

Para a implementação dos experimentos com os modelos descritos, fora estudadas as tecnologias presentes hodiernamente no cenário de inteligência computacional, e optada pela implementação na linguagem de programação *Python*. Esta é uma moderna linguagem de uso geral, com grande aplicabilidade no escopo científico devido à seu alto nível. Sendo voltada à legibilidade de código, esta linguagem possui uma rápida curva de aprendizagem, e uma ampla gama de bibliotecas e funcionalidades para diversas áreas da ciência, mentidas por sua comunidade que é suportada por grandes apoiadores ao redor do mundo.

Devido à seu funcionamento como uma linguagem de *script*³, e sua interface com as bibliotecas que serão usadas, por já implementarem de forma otimizada ferramentas necessárias nos experimentos propostos, o *Python* se adequou corretamente ao trabalho. Diferentes bibliotecas foram utilizadas para o cumprimento do propósito, sendo elas:

- **ScyPy:** Ecossistema de bibliotecas de código aberto com funcionalidades para implementações matemáticas, científicas e de engenharia. Destas, foram utilizadas nos experimentos as bibliotecas: **Pandas**, para importação e estruturação da base de dados proveniente de um arquivo de texto; **NumPy**, complementar ao anterior, é uma biblioteca de estruturação e funcionalidades para o tratamento de estruturas de dados em vetores, foram utilizados seus vetores para armazenamento e tratamento das séries temporais; e **Matplotlib**, com funcionalidades de plotagem de gráficos para visualização de dados (JONES et al., 2001).
- **StatsModels:** Módulo em *Python* composto de classes e funções que implementam a estimação de diversos modelos estatísticos, além de funcionalidades para a condução de experimentações estatísticas. Utilizado neste trabalho para implementação do preditor ARIMA (SEABOLD; PERKTOLD, 2010).
- **SciKit Learn:** Biblioteca de aprendizado de máquina preparada para ambientes de experimentação e produção, e amplamente utilizada para diferentes modelos, atuando com funcionalidades desde o pré processamento, *feature engeneering*, e atuação dos modelos. Neste trabalho, as RNAs utilizadas foram implementadas sobre esta base (PEDREGOSA et al., 2011).
- **TensorFlow:** Interface de desenvolvimento sustentada pelo *Google* que implementa funcionalidades para a construção e execução de algoritmos de aprendizado de máquina. O modelo ANFIS a ser experimentado neste trabalho consome funções desta biblioteca (ABADI et al., 2015).

É importante ressaltar, no desmonte dessa arquitetura, que todas as bibliotecas utilizadas são de código aberto e mantidas com contribuição de suas comunidades. Assim como os

³ Programas escritos para rodar comandos de sistema em tempo de execução. Estas possibilitam a escrita de programas para rodar um determinado conjunto de tarefas sem a necessidade de escrita e compilação de um sistema, ou da colocação de um sistema em produção para execução dos comandos. Encontram ampla aplicabilidade em experimentações com necessidade de auxílio computacional (BARRON, 2000).

códigos e bases de dados para as experimentações propostas neste trabalho também o são, e podem ser encontrados no repositório aberto <https://github.com/mateussantos21/coffe_price_prediction>.

4.3.1 Parametrização

A otimização de hiper parâmetros dos modelos neurais por busca em rede⁴ consiste na busca exaustiva, com a utilização de força bruta, do melhor desempenho dentro de um universo de possíveis parâmetros a serem atribuídos ao modelo no momento do treinamento. Ou seja, dado um vetor com n valores para um determinado parâmetro, o algoritmo treinará a rede n vezes, utilizando um valor por vez, e retornará aquele que chegar ao final do treinamento com o melhor desempenho mensurado (PEDREGOSA et al., 2011).

A escolha de utilização deste algoritmo para pesquisa operacional no universo de combinações dos parâmetros se deu pela sua simplicidade de implementação e entendimento, visto uma abordagem introdutória, onde o foco é o comportamento das RNAs de forma mais geral, sem a análise de sensibilidade das mesmas a cada parâmetro que as compõem.

4.4 Medição de Desempenho

Os resultados de experimentação do presente trabalho objetivam comparação dos métodos inteligência computacional em detrimento da técnica econométrica tradicional utilizada. Além de traçar um perfil comparativo entre as três técnicas mais avançadas entre si para a base de dados proposta. Isso posto, os resultados não serão comparados àqueles encontrados na literatura para as mesmas técnicas.

Três métricas de desempenho com funções distintas guiarão as análises de qualidade, sendo elas o erro quadrático médio (MSE⁵), a raiz do erro quadrático médio (RMSE⁶), e o erro percentual absoluto médio (MAPE⁷). Sendo y o vetor de valores esperados da variável, \hat{y} o vetor de valores preditos pelo modelo e n o número de amostras dos vetores; as métricas são definidas por:

⁴ Termo traduzido livremente do inglês *grid search*, como é encontrado na literatura sobre as bibliotecas utilizadas para construção dos modelos.

⁵ Do inglês *Mean Squared Error*.

⁶ Do inglês *Root Mean Squared Error*.

⁷ Do inglês *Mean Absolute Percentage Error*.

$$MSE(y, \hat{y}) = \frac{1}{n} \sum_{i=0}^{n-1} (y_i - \hat{y}_i)^2;$$

$$RMSE(y, \hat{y}) = \frac{1}{n} \sum_{i=0}^{n-1} \sqrt{(y_i - \hat{y}_i)^2};$$

$$MAPE(y, \hat{y}) = \frac{1}{n} \sum_{i=0}^{n-1} \left| \frac{y_i - \hat{y}_i}{y_i} \right|;$$

onde o primeiro fornecerá um indicador onde diferenças menores afetarão de forma menos significativa que as diferenças maiores, dando vantagem para métodos com maior precisão. O segundo, como raiz quadrada do primeiro, retorna as medidas para suas unidades originais, fornecendo uma análise de peso dos erros de forma mais linear. Por fim, o MAPE fornecerá uma interpretação do ponto de vista percentual, contribuindo com uma análise econômica de quanto o modelo difere da realidade (FILHO, 2018).

Além dessa análise, tabelas que permitem a ilustração da reação de alguma métrica do modelo em função da variação de algum parâmetro de treinamento ou construção, juntamente à curvas de aprendizado, serão utilizadas com a finalidade de contribuir com a proposta de soluções para melhoria dos resultados obtidos em trabalhos futuros (PEDREGOSA et al., 2011).

5 RESULTADOS

5.1 Análise Comparativa

Sobre a ótica do estado da arte levantado neste documento, este capítulo descreve os resultados obtidos com a experimentação realizada para testar a capacidade de aproximação de funções das RNAs descritas frente à metodologia econométrica com maior relevância em citações acadêmicas. A base de experimentação foi dada pela série temporal montada conforme a seção 4.2.

É válido ressaltar que a comparação do desempenho da MLP, RBF e ANFIS, frente ao ARIMA, não foi feita por base em resultados deste último encontrados na literatura, e sim do próprio método reproduzido em software juntamente aos outros modelos. O *script* utilizado para treinar os modelos, gerar as análises de resultado, assim como as imagens utilizadas neste capítulo seguem o padrão descrito na seção 4.3.

O experimento consistiu em duas partes de forma macro, com suas subdivisões, sendo elas uma rodada de cada um dos métodos com a parametrização padrão encontrada na literatura básica da utilização dessas técnicas na mesma ferramenta, encontradas na referência da seção 4.3. Para estes parâmetros serão analisados o desempenho calculado em MSE, RMSE e MAPE, e plotado gráfico dos dados utilizados para teste em comparação às previsões tidas como saída dos modelos.

Em um segundo momento, as mesmas análises foram feitas, entretanto com os modelos parametrizados por meio de uma busca em rede conforme descrito na subseção 4.3.1. Visando que, com essas análises postas, fosse possível analisar a viabilidade de utilização de inteligência computacional mais rebuscada para modelagem da função proposta, e os efeitos de melhoria de desempenho causados por uma busca operacional sobre o universo dos hiper parâmetros de cada modelo face ao custo computacional de sua aplicação.

Todos os treinamentos e testes foram executados tendo por entrada um total de 10 atrasos da variável alvo de regressão, convencionado por corresponder à duas semanas úteis de dados de variação de preços no mercado de commodities, que não indexa valores em finais de semana e feriados, juntamente com as variáveis exógenas dadas como relevantes por correlação de Pearson.

5.1.1 Parametrização Padrão

Por padrão das bibliotecas utilizadas, os parâmetros propostos para o modelo ARIMA foram $(p, q, d) = [2, 1, 1]$. Para a rede neural MLP fora configurada a taxa de aprendizado $\eta = 0,001$, e um número de neurônios nas camadas ocultas, respectivamente para a primeira e segunda camadas ocultas, igual à $(100, 1)$, enquanto este método foi treinado por 500 iterações, otimizando seus pesos em cada uma delas para se aproximar à função. Já a RBF fora treinada com um número padrão de neurônios na camada oculta de 100, e um $\alpha = 0,00001$, sendo este o valor adicionado à diagonal da matriz *kernel* a cada iteração, e o processo de treinamento otimizado ao longo de 100 iterações. Por fim, o modelo ANFIS foi treinado neste primeiro momento com taxa de aprendizado $\eta = 0.01$, um número de regras se-então *fuzzy* igual à 100 (correspondente ao padrão de n^2 , sendo n o numero de entradas do modelo), por um número de épocas igual à 500.

Os resultados encontrado nessa remessa do experimento foi inserido na Tabela 5.1. Sendo divididos por medida de desempenho por modelo treinado, e na última linha acrescentado o tempo total de treinamento (em segundos) para aquisição destes resultados sobre estes parâmetros¹.

Tabela 5.1 – Resultados da experimentação com parâmetros padrões.

	ARIMA	MLP	RBF	ANFIS
MSE	6.89	7.50	186.69	25.97
RMSE	2.63	2.74	13.66	5.10
MAPE	1.36	1.43	5.14	2.64
Tempo de execução	19.62	1.93	88.35	10.06

Fonte: O Autor, 2019.

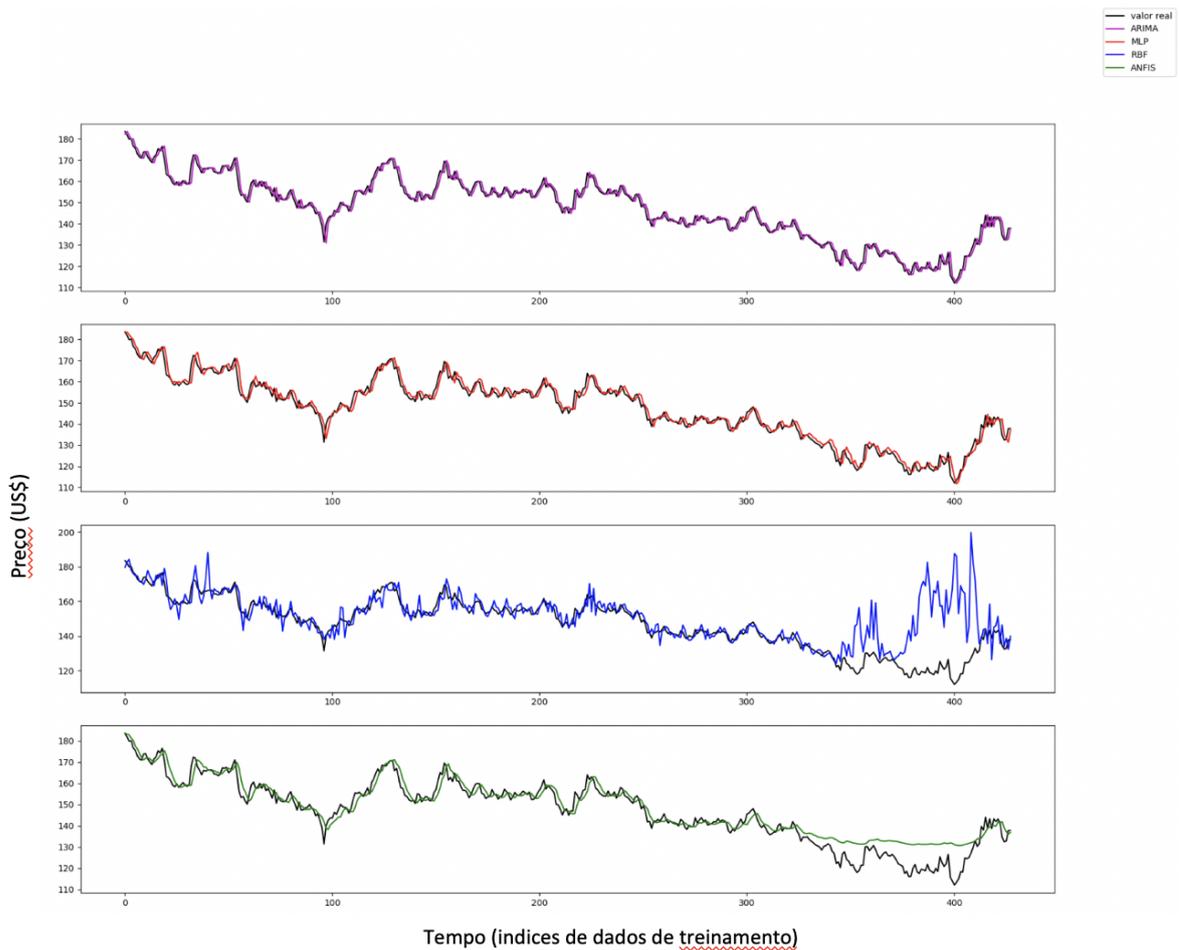
Já na Figura 5.1, observa-se a representação visual da predição feita pelos modelos face aos valores reais dos dados de teste. As medidas de desempenho das curvas podem ser acompanhadas pela Tabela 5.1.

5.1.2 Parametrização Otimizada

Já com a busca em rede no universo dos hiper parâmetros, foram encontrados os parâmetros ótimos dentro dos respectivos universos de busca, descritos a seguir.

¹ Todos os experimentos foram executados na máquina do autor, sem alteração de configurações, rodando sem concorrência com outros processos.

Figura 5.1 – Gráficos de resultados de predição em dados separados para teste com parametrização padrão das bibliotecas utilizadas.



Fonte: O Autor, 2019.

Para o ARIMA, a busca encontrou $(p, q, d) = [1, 0, 0]$, buscando em um alcance de 0 a 5, iterado a cada unidade inteira, para cada parâmetro. Resultado que indica o padrão estocástico dos dados, onde o modelo tem maior facilidade em aproximar a função sobre a auto regressão de apenas um valor, sem o recurso integrador. Para a MLP, a taxa de aprendizado $\eta = 0,00085$, número de neurônios nas camadas ocultas igual à $(600, 400)$, enquanto este método foi treinado por 500 iterações, com um vetor de busca para η de 0.00001 a 0.1 iterado a cada 0.00001, e para número de neurônios, de 100 a 800 iterado a cada 100 unidades inteiras. A RBF foi treinada com um número de neurônios na camada oculta de 100, um $\alpha = 0,0001$, buscado em um vetor de 0.00001 a 0.01 iterado em 0.00001, e o processo de treinamento otimizado ao longo de 500 iterações. E o modelo ANFIS foi treinado com taxa de aprendizado $\eta = 0.04$, alcançada em uma busca de 0.0001 a 0.1 iterada em 0.0001, um número de regras se-então *fuzzy* igual à 175,

encontrada no universo de 100 a 300 iterado a cada 5 unidades inteiras, por um número de épocas igual à 500.

A Tabela 5.2 contém os resultados dessa segunda remessa de experimentos sendo dividida pelo mesmo padrão da seção anterior para futura comparação.

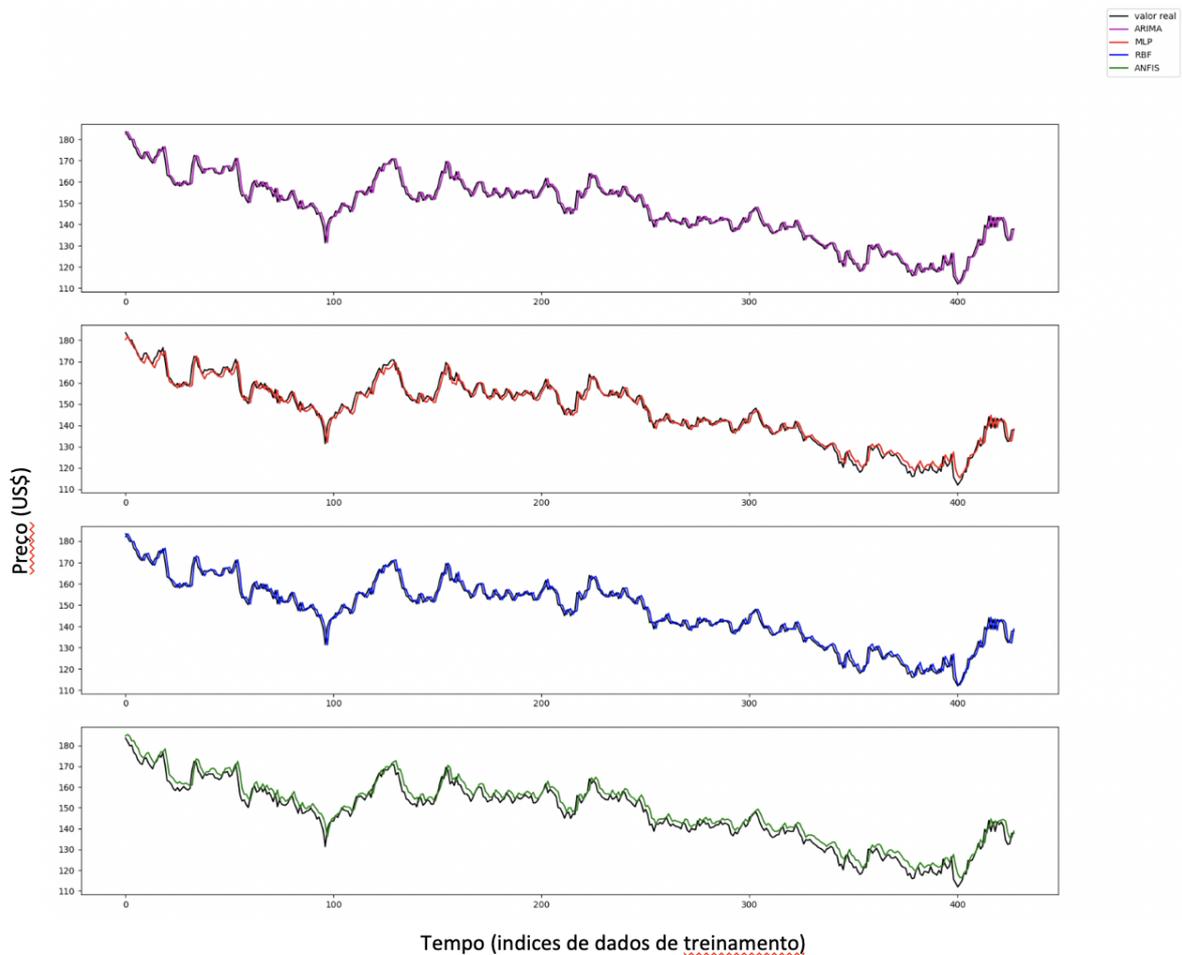
Tabela 5.2 – Resultados da experimentação com parâmetros padrões.

	ARIMA	MLP	RBF	ANFIS
MSE	6.88	7.18	6.98	12.45
RMSE	2.62	2.68	2.64	3.53
MAPE	1.36	1.42	1.38	2.01
Tempo de execução	16.63	110.30	130.15	55.59

Fonte: O Autor, 2019.

E a representação gráfica dos resultados dessa remessa são postos na Figura 5.2.

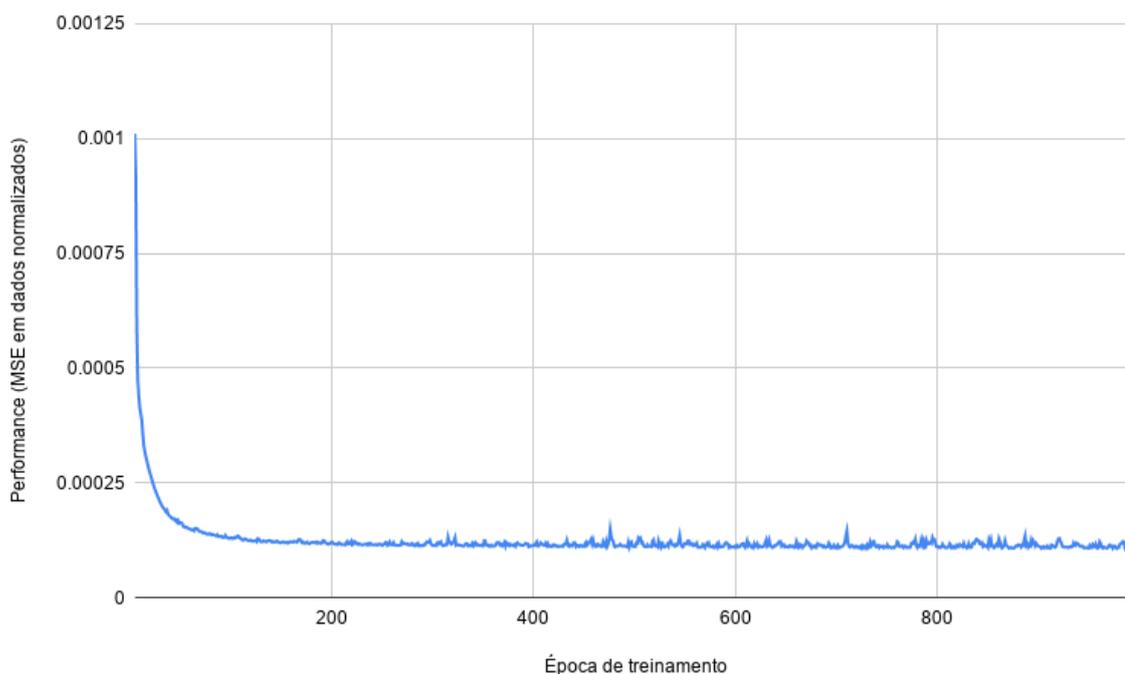
Figura 5.2 – Gráficos de resultados de predição em dados separados para teste com parametrização otimizada com busca em rede.



Fonte: O Autor, 2019.

Aqui, observa-se que o número de épocas de treinamento para as RNAs não foi inserido na busca em rede para parametrização. O motivo pode ser visto representado graficamente pela Figura 5.3, onde observa-se a medida de desempenho final ao longo do número de épocas de treinamento. Aqui, entende-se que a aprendizagem da rede chega em uma saturação por base nos demais hiper parâmetros, e devido a tal, dispensa-se a necessidade de buscar em um universo de épocas diferentes desempenhos melhores que aqueles na estabilização da curva.

Figura 5.3 – Curvas de desempenho por número de épocas de treinamento para MLP.



Fonte: O Autor, 2019.

5.2 Sumário e Discussões

Em primeira instância, observa-se nos dois cenários experimentados a superioridade da técnica econométrica clássica sobre os modelos de inteligência computacional propostos. E, além disso, vale salientar, para início da análise, a forma como a parametrização desta técnica se comportou. Sua parametrização padrão é dada por um fator auto regressivo de ordem 2, integrador e médias móveis com ordem 1; sobre os quais o desempenho RMSE se apresenta igual à 2,63. A pesquisa operacional sobre a ordem dos parâmetros apontou uma melhoria de 0,01 para esta métrica neste modelo, melhoria de baixa significância no contexto da métrica, entretanto conclusiva ao se analisar os parâmetros encontrados de auto regressão em ordem 1,

integrador e médias móveis em ordem 0. Este comportamento indica a forma como a adição de dados passados para regressão acabam prejudicando o modelo em suas previsões. Pode-se afirmar que a melhor forma de regredir a variável alvo é por defini-la no próximo instante de tempo como igual à observada no instante anterior (auto regressão de ordem 1).

Esse íterim remonta contexto para entendimento da incapacidade das RNAs de aproximarem a função de preço da commodity ao longo do tempo. Visto que os próprios valores passados prejudicam as métricas, entende-se um modelo de estoicidade elevada em relação a seu determinismo, fazendo com que mesmo aproximadores não lineares sejam incapazes de obter boa precisão. Entretanto, conclui-se melhorias interessantes na aplicação da busca na região dos hiper parâmetros.

Na rede MLP, o aumento do número de neurônios por camada oculta demonstra melhoria significativa em sua capacidade de atribuir pesos a entrada que aproxima melhor sua saída, por outro lado, percebe-se um aumento desproporcional do tempo de execução necessário para concluir seu treinamento. Para a RBF, a primeira observação demonstra uma incapacidade de agir com valores preditivos dado um determinado comportamento, visto que as previsões destoam de forma mais visível para certa área da série. Este comportamento é amenizado com o aumento da relevância de seu α , aumentando sua capacidade de regressão pelos valores próximos àquele a ser predito. Por fim, na rede ANFIS observa-se melhoria de sua capacidade de aproximação da função com o aumento do número de regras, conforme esperado, por motivo similar à MLP em sua forma de atribuição de pesos. Contudo, nos três casos em estudo, o incremento de esforço computacional no treinamento não se faz proporcional à melhoria de desempenho.

6 CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS

Sob a perspectiva do impacto hodierno causado pelas diversas técnicas de inteligência computacional presentes nos modelos de negócio mais modernos, este trabalho permitiu a observação da aplicabilidade das tecnologias em uma área de singular importância para o contexto econômico brasileiro. O aprofundamento nas análises de novas metodologias para estudos econométricos nas movimentações de commodities abre precedente para um universo de possibilidades que agregam inovação ao agronegócio brasileiro.

Com esta produção foi possível observar a grande complexidade de aplicação de modelos de inteligência computacional em uma ciência com grande quantidade de fatores inexatos, como a economia. Produzindo séries de dados não-estacionárias e com alto fator estocástico, o que dificulta a adequação de modelos aproximadores universais para estas funções. Mesmo com a utilização de algoritmos de busca no universo dos meta-parâmetros de redes neurais, houve dificuldade em treinar modelos que se aproximassem do desempenho de uma metodologia econométrica clássica.

Contudo, em contexto geral, foi possível observar de forma holística o comportamento dos dados coletados frente aos modelos, e abrir perspectivas para futuros melhoramentos e descobertas. São três os principais caminhos observados para possíveis melhorias dos métodos. Primeiramente, há várias propriedades não-exploradas na base de dados. A instância de cada dia possui outros valores que podem ser de grande valia para modelagem e estimação de preços futuros. Cabe um trabalho aprofundado de *feature engineering* para explorar tal mutualidade entre variáveis. Em segundo lugar, dentro do contexto das mesmas redes neurais exploradas neste trabalho, podem ser aplicadas técnicas de otimização baseadas em regularização, e em estratégias populacionais visando encontrar melhores soluções globais no espaço dos parâmetros. Cita-se como exemplo a técnica de corrida iterada (*iRace*¹), conforme (LOPEZ-IBANEZ et al., 2016).

Por fim, trabalhos futuros possuem um oceano de técnicas de inteligência computacional contemporâneas, que não foram exploradas neste trabalho por seu contexto matemático fugir do escopo. Estas técnicas podem ser aplicadas em diferentes fases do processo de projeto de modelos para contribuir com o desempenho das predições. Nesse ínterim, conclui-se que as possibilidades e desafios são diversas, e a complexidade do problema impõe uma abertura para aplicabilidade nas mais diversas áreas da economia tanto quanto.

¹ Do inglês *iterated race*.

REFERÊNCIAS

- ABADI, M. et al. **TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems**. 2015. Software available from tensorflow.org. Disponível em: <<https://www.tensorflow.org/>>. Acesso em: 20 jul. 2019.
- ABIC. **Produção e consumo de café em nível mundial batem recordes no ano-safra 2017/18**. 2018. Disponível em: <<http://abic.com.br/producao-e-consumo-de-cafe-em-nivel-mundial-batem-records-no-ano-safra-201718/>>. Acesso em: 21 abr. 2019.
- AIHARA, K.; TAKABE, T.; TOYODA, M. Chaotic neural networks. **Physics letters A**, Elsevier, v. 144, n. 6-7, p. 333–340, 1990.
- ATIVO. In: WIKIPÉDIA: a enciclopédia livre. Wikimedia, 2010. Disponível em: <<https://pt.wikipedia.org/wiki/Ativo>>. Acesso em: 12 abr. 2019.
- BACCI, L. A. Combinação de métodos de séries temporais para previsão de demanda de café no brasil. 2007.
- BARRON, D. **The World of Scripting Languages**. Wiley, 2000. (North-Holland Mathematical Library). ISBN 9780471998860. Disponível em: <<https://books.google.com.br/books?id=ZpFQAAAAMAAJ>>.
- BECKER, R. A.; CHAMBERS, J. M.; WILKS, A. R. The new s language. **Wadsworth Brooks/Cole**, 1988.
- BOX, G. E. P.; JENKINS, G. **Time Series Analysis, Forecasting and Control**. San Francisco, CA, USA: Holden-Day, Inc., 1990. ISBN 0816211043.
- BROOMHEAD, D. S.; LOWE, D. **Radial basis functions, multi-variable functional interpolation and adaptive networks**. [S.l.], 1988.
- CASTRO, F. D.; CASTRO, M. D. Redes neurais artificiais. **Porto Alegre, RS: Pontifícia Universidade Católica do Rio Grande do Sul**, 2001.
- CEPEA ESALQ. **Indicador do Café Arábica**. 2019. Disponível em: <<https://www.cepea.esalq.usp.br/br/indicador/cafe.aspx>>. Acesso em: 20 jul. 2019.
- COELHO, L. d. S.; SANTOS, A. A. A. P.; JR., N. C. A. d. C. Podemos prever a taxa de cambio brasileira? *scielo*, v. 15, p. 635 – 647, 12 2008. ISSN 0104-530X. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0104-530X2008000300016&nrm=iso>.
- CONAB. **Regiões Cafeeiras**. 2019. Disponível em: <<http://www.rossignolli.com.br/regioes-cafeeiras.php>>.
- DUDA, R. O.; HART, P. E.; STORK, D. G. **Pattern Classification (2Nd Edition)**. New York, NY, USA: Wiley-Interscience, 2000. ISBN 0471056693.
- FILHO, M. **As Métricas Mais Populares para Avaliar Modelos de Machine Learning**. 2018. Disponível em: <<http://mariofilho.com/as-metricas-mais-populares-para-avaliar-modelos-de-machine-learning/>>. Acesso em: 20 jul. 2019.

GUTIERREZ, C. E. C.; ALMEIDA, F. M. de M. Modelagem e previsão do preço do café brasileiro. **Revista de Economia**, v. 39, n. 2, 2013.

HAYKIN, S. **Neural networks: a comprehensive foundation**. [S.l.]: Prentice Hall PTR, 1994.

INC, T. S. **SANN Overview - Network Types**. 2017. Disponível em: <<http://documentation.statsoft.com/STATISTICAHelp.aspx?path=SANN/Overview/SANNOverviewsNetworkTypes>>.

INVESTING.COM. **Arabica Coffee 4/5 Futures - May 19 (ICFc1)**. 2019. Disponível em: <<https://www.investing.com/commodities/arabica-coffee-4-5>>. Acesso em: 20 jul. 2019.

JANG, J.; SUN, C.; MIZUTANI, E. **Neuro-fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence**. Prentice Hall, 1997. (MATLAB curriculum series). ISBN 9780132610667. Disponível em: <<https://books.google.com.br/books?id=vN5QAAAAMAAJ>>.

JANG, J. shing R. Anfis: Adaptive-network-based fuzzy inference system. **IEEE Transactions on Systems, Man, and Cybernetics**, v. 23, p. 665–685, 1993.

JONES, E. et al. **SciPy: Open source scientific tools for Python**. 2001. Disponível em: <<http://www.scipy.org/>>. Acesso em: 20 jul. 2019.

KIA, A. Using mlp and rbf neural networks to improve the prediction of exchange rate time series with arima. **International Journal of Information and Electronics Engineering**, 01 2012.

KOEHRSEN, W. **Feature Engineering: What Powers Machine Learning**. 2018. Disponível em: <<https://towardsdatascience.com/feature-engineering-what-powers-machine-learning-93ab191bcc2d>>.

LI, Y. et al. Forecasting mineral commodity prices with arima-markov chain. In: **2012 4th International Conference on Intelligent Human-Machine Systems and Cybernetics**. [S.l.: s.n.], 2012. v. 1, p. 49–52.

LOPEZ-IBANEZ, M. et al. The irace package: Iterated racing for automatic algorithm configuration. **Operations Research Perspectives**, v. 3, p. 43 – 58, 2016. ISSN 2214-7160. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S2214716015300270>>.

MATHUR, N.; GLESK, I.; BUIS, A. Comparison of adaptive neuro-fuzzy inference system (anfis) and gaussian processes for machine learning (gpml) algorithms for the prediction of skin temperature in lower limb prostheses. **Medical Engineering Physics**, v. 38, n. 10, p. 1083 – 1089, 2016. ISSN 1350-4533. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1350453316301515>>.

MERCADO, S. e. **CAFÉ: Faturamento das lavouras do Brasil atinge R\$ 25 bi em 2018**. 2019. Disponível em: <<http://www2.safras.com.br/cafe-faturamento-das-lavouras-do-brasil-atinge-r-25-bi-em-2018/>>. Acesso em: 21 abr. 2019.

MESQUITA, J. et al. Mercado de café: variáveis que influenciam o preço pago ao produtor. **Ciência e Agrotecnologia, Lavras**, v. 24, n. 2, p. 379–386, 2000.

- MINGA, E.; ALVES, A. F.; PARRÉ, J. L. Especulação afeta o preço das commodities agrícolas? **Revista Espacios**, v. 37, n. 20, p. 13, 2016. Disponível em: <<https://www.revistaespacios.com/a16v37n20/16372013.html>>. Acesso em: 20 jun. 2019.
- MINSKY, M.; PAPERT, S. A. **Perceptrons: An introduction to computational geometry**. [S.l.]: MIT press, 1969.
- MITRA, S.; HAYASHI, Y. Neuro-fuzzy rule generation: survey in soft computing framework. **IEEE Transactions on Neural Networks**, v. 11, n. 3, p. 748–768, May 2000. ISSN 1045-9227.
- NAUCK, D.; KLAWONN, F.; KRUSE, R. **Foundations of neuro-fuzzy systems**. [S.l.]: John Wiley & Sons, Inc., 1997.
- NOTÍCIAS AGRÍCOLAS. **Cotações de Café**. 2019. Disponível em: <<https://www.noticiasagricolas.com.br/cotacoes/cafe>>. Acesso em: 20 jul. 2019.
- PEDREGOSA, F. et al. Scikit-learn: Machine learning in python. **Journal of Machine Learning Research**, v. 12, p. 2825–2830, 10 2011. Disponível em: <<http://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf>>. Acesso em: 20 jul. 2019.
- PINDYCK, R.; RUBINFELD, D. **Microeconomics**. Prentice Hall, 2001. (Economics Series). ISBN 9780130165831. Disponível em: <<https://books.google.com.br/books?id=h6e7AAAAIAAJ>>.
- RIBEIRO, K.; SOUSA, A.; ROGERS, P. Coffee prices in brazil: predictive variables in cash and future markets. **REGE Revista de Gestão**, v. 13, n. 1, p. 11–30, 2006. Disponível em: <<http://www.periodicos.usp.br/rege/article/view/36547>>. Acesso em: 20 jun. 2019.
- RIBEIRO, K. C. de S.; SOUSA, A. F. de; ROGERS, P. Preços do café no brasil: variáveis preditivas no mercado à vista e futuro. **REGE Revista de Gestão**, v. 13, p. 11–30, 2006.
- SAMPAIO, L. M.; OLIVEIRA, M. J.; IGNACIO, A. A. Lógica nebulosa: Aplicações e tendências. **SPOLM2007**, 11 2007.
- SANDMANN, H. R. **Predição não-linear de séries temporais usando sistemas de arquitetura neuro-fuzzy**. Dissertação (Mestrado) — Escola Politécnica da Universidade de São Paulo, 2006.
- SEABOLD, S.; PERKTOLD, J. Statsmodels: Econometric and statistical modeling with python. In: **9th Python in Science Conference**. [S.l.: s.n.], 2010.
- SILBERSCHATZ, A.; SUNDARSHAN, S.; KORTH, H. F. **Sistema de banco de dados**. [S.l.]: Elsevier Brasil, 2016.
- SOARES, D. G.; TEIVE, R. Estudo comparativo entre as redes neurais artificiais mlp e rbf para previsão de cheias em curto prazo. **Revista de Informática Teórica e Aplicada**, v. 22, 08 2015.
- SOMPOLINSKY, H. The theory of neural networks: The hebb rule and beyond. In: SPRINGER. **Heidelberg colloquium on glassy dynamics**. [S.l.], 1987. p. 485–527.
- TAKAGI, T.; SUGENO, M. Derivation of fuzzy control rules from human operator's control actions. **IFAC Proceedings Volumes**, Elsevier, v. 16, n. 13, p. 55–60, 1983.

U.S COMMODITY FUTURES TRADING COMMISSION. **Commitment of Traders**. 2019. Disponível em: <<https://www.cftc.gov/MarketReports/CommitmentsofTraders/index.htm>>. Acesso em: 20 jul. 2019.

WALLER, M. A.; FAWCETT, S. E. Data science, predictive analytics, and big data: a revolution that will transform supply chain design and management. **Journal of Business Logistics**, Wiley Online Library, v. 34, n. 2, p. 77–84, 2013.

WERBOS, P. J. Generalization of backpropagation with application to a recurrent gas market model. **Neural networks**, Elsevier, v. 1, n. 4, p. 339–356, 1988.

ZADEH, L. A. Fuzzy sets, fuzzy logic, and fuzzy systems. In: KLIR, G. J.; YUAN, B. (Ed.). River Edge, NJ, USA: World Scientific Publishing Co., Inc., 1996. cap. Fuzzy Sets, p. 19–34. ISBN 981-02-2422-2. Disponível em: <<http://dl.acm.org/citation.cfm?id=234347.234356>>.