



UNIVERSIDADE ESTADUAL DE SANTA CRUZ
PRO-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM COMPUTACIONAL
EM CIÊNCIA E TECNOLOGIA

DÁRCIO ROCHA DA SILVA

TÉCNICAS DE MINERAÇÃO DE DADOS APLICADAS A ANÁLISE DE
SÉRIES TEMPORAIS DE DADOS SOCIOECONÔMICOS DA REGIÃO SUL
DA BAHIA

ILHÉUS-BA
2019

DÁRCIO ROCHA DA SILVA

**TÉCNICAS DE MINERAÇÃO DE DADOS
APLICADAS A ANÁLISE DE SÉRIES TEMPORAIS
DE DADOS SOCIOECONÔMICOS DA REGIÃO
SUL DA BAHIA**

Dissertação apresentada ao Programa de Pós-Graduação em Modelagem Computacional em Ciência e Tecnologia da Universidade Estadual de Santa Cruz, como parte das exigências para obtenção do título de Mestre em Modelagem Computacional em Ciência e Tecnologia.

Orientador: Prof. Dr. Dany Sanchez Dominguez

Coorientador: Prof. Dr. Paulo Eduardo Ambrósio

ILHÉUS-BA
2019

S586

Silva, Dárcio Rocha da.

Técnicas de mineração de dados aplicados a análise de séries temporais de dados socioeconômicos da região sul da Bahia / Dárcio Rocha da Silva. – Ilhéus : UESC, 2019.

83f. : il.

Orientador : Dany Sanchez Dominguez.

Co-orientador : Paulo Eduardo Ambrósio.

Dissertação (Mestrado) – Universidade Estadual de Santa Cruz. Programa de Pós-graduação em Modelagem Computacional em Ciência e Tecnologia.

Inclui referências.

1. Redes neurais (computação). 2. Sistema fuzzy. 3. Métodos estatísticos. I. Dominguez, Dany Sanchez. II. Ambrósio, Paulo Eduardo. III. Título.

CDD – 629.895

DÁRCIO ROCHA DA SILVA

**TÉCNICAS DE MINERAÇÃO DE DADOS
APLICADAS A ANÁLISE DE SÉRIES TEMPORAIS
DE DADOS SOCIOECONÔMICOS DA REGIÃO
SUL DA BAHIA**

Ilhéus-BA, 25/02/2019

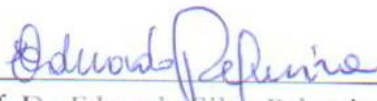
Comissão Examinadora



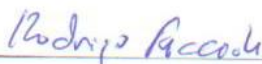
Prof. Dr. Dany Sanchez Dominguez
UESC
(Orientador)



Prof. Dr. Paulo Eduardo Ambrósio
UESC
(Coorientador)



Prof. Dr. Eduardo Silva Palmeira
UESC



Prof. Dr. Rodrigo Antônio Faccioli
UFRB

Dedico este trabalho aos meus pais, Manoel e Edleusa, aos meus irmãos, Taline, Caliane e Deris, e à minha sobrinha, Manuely da Silva.

Agradecimentos

- Ao professor Dany Sanchez Dominguez, por todos os ensinamentos compartilhados e dedicação na orientação, sem os quais seria inviável desenvolver esse trabalho.
- Ao professor Paulo Ambrósio, pela coorientação e os ensinamentos compartilhados.
- Aos membros da banca de defesa, pela disposição, críticas e sugestões.
- Ao professor César Alberto Bravo Pariente, pela amizade, confiança e incentivo.
- Ao Programa de Pós-Graduação em Modelagem Computacional, seu corpo docente pelos ensinamentos e incentivos.
- Ao NBCGIB, pela infraestrutura e recursos disponibilizados para o desenvolvimento da pesquisa.
- Aos colegas de curso, pela amizade e pelos excelentes momentos de convívio pessoal e profissional, em especial a Uerbert Moreira, Romilson Souza, Islane Dutra, Fabilone Silva e Allan Gonçalves.
- Aos amigos de longas datas, Poliane Santos, Elizângela Cardoso, Tâmara Carolina, Barbara Hellen, Rhebeka Oliveira, Erick Moreira, Marccone Lopes, Victor Batista, Messias Vituriano, Cleviton Borges, Gleisson Carrilo, Júnior Bergens e Marcos Silva.
- Aos meus pais e meus irmãos, pelo estímulo, confiança, incentivo nas decisões e pelo amor incondicional.
- A todos que contribuíram direta ou indiretamente para esta realização.

*“Você não pode ensinar nada a ninguém, mas
pode ajudar a pessoas a descobrirem por si mes-
mas.”*

Galileu Galilei

TÉCNICAS DE MINERAÇÃO DE DADOS APLICADAS A ANÁLISE DE SÉRIES TEMPORAIS DE DADOS SOCIOECONÔMICOS DA REGIÃO SUL DA BAHIA

Resumo

O estudo de diversas técnicas de análise e previsão de séries temporais de dados têm sido cada vez mais frequente entre a comunidade acadêmica e empresarial, constituindo um forte aliado no suporte a tomada de decisão. Com isso, o objetivo deste trabalho foi comparar os métodos de estatísticos, de inteligência artificial e abordagens híbridas para o problema de previsão de séries temporais. Foram utilizados os métodos de suavização, ARIMA, redes neurais artificiais (RNA), RNA com ARIMA e sistema *neurofuzzy* na análise e previsão de series temporais de dados socioeconômicos. Para isto, primeiro foi obtido um modelo de ajustamento para cada um dos métodos seguido pela realização das previsões a curto e médio prazo, onde a qualidade da previsão oferecida por cada método foi avaliado mediante medidas de acurácia. Ao final, os resultados mostraram que todos os métodos utilizados foram bastantes competitivos entre si quanto ao ajustamento dos modelos, entretanto, em relação a previsão, o destaque é para as abordagens híbridas as quais apresentaram desempenhos superiores em termo de acurácia. Sendo que a abordagem híbrida ARIMA com RNA apresentou melhores resultados ao realizar previsões a curto prazo e o sistema *neurofuzzy* para previsões a médio prazo.

Palavras-chave: Séries Temporais, Dados Socioeconômicos, Redes Neurais Artificiais, Sistema Neurofuzzy, Métodos Híbridos.

DATA MINING TECHNIQUES APPLIED TO THE ANALYSIS OF TEMPORARY SERIES OF SOCIOECONOMIC DATA IN THE SOUTHERN REGION OF BAHIA

Abstract

The study of several techniques for analysis and time series forecast has been increasingly among the academic and business community, constituting a strong ally in the decision making support. Thus, the objective of this paper was to compare statistical methods, artificial intelligence and hybrid approaches to the problem of forecasting time series. We used smoothing methods, ARIMA, artificial neural networks (ANN), ANN with ARIMA and neurofuzzy system in the analysis and forecast of time series of socioeconomic data. For this, first an adjustment model was obtained for each of technical followed by the realization of the short- and medium-term forecasts, and compared by means of measures of accuracy. In the end, the results showed that all the methods used presented competitive results among themselves regarding the adjustment of the models. However, in itself, when dealing with forecasting, the highlight is for the hybrid approaches which presented superior results in terms of accuracy. Since the ARIMA hybrid approach with ANN presented better results when performing short-term forecasts and the neurofuzzy system for medium-term forecasts.

Keywords: Time Series, Socioeconomic Data, Artificial Neural Networks, Hybrid Methods, Neurofuzzy System.

Lista de figuras

Figura 1 – Neurônio artificial	24
Figura 2 – Função de ativação degrau	26
Figura 3 – Função de ativação degrau bipolar	27
Figura 4 – Função de ativação logística	27
Figura 5 – Função de ativação linear	28
Figura 6 – Exemplo de rede <i>feedforward</i> de camada múltiplas	29
Figura 7 – Rede perceptron multicamadas	31
Figura 8 – Esquema geral de um controlador fuzzy	36
Figura 9 – Modelo ANFIS com duas entradas e uma saída.	39
Figura 10 – Modelo fuzzy Sugeno com duas entradas e duas regras (Jang (1993)).	39
Figura 11 – Representação esquemática do algoritmo Box Jenkins	49
Figura 12 – Arquitetura da RNA para a série temporal do custo da cesta básica da cidade de Ilhéus, jan/2005 a dez/2016	52
Figura 13 – Etapas da abordagem ARIMA com RNA	54
Figura 14 – Sistema neurofuzzy com três entradas para a série temporal do custo da cesta básica da cidade de Ilhéus jan/2005 a dez/2016	55
Figura 15 – Custo da cesta básica da cidade de Ilhéus, jan/2005 a dez/2016 . . .	59
Figura 16 – Amplitude \times média do custo da cesta básica da cidade de Ilhéus, jan/2005 a dez/2016	60
Figura 17 – Autocorrelação e autocorrelação parcial amostrais do custo da cesta básica da cidade de Ilhéus, jan/2005 a dez/2016	61
Figura 18 – (a) Série temporal do custo da cesta básica da cidade de Ilhéus, jan/2005 a dez/2016 (b) Diferença do custo da cesta básica.	62
Figura 19 – Modelo livre de sazonalidade do custo da cesta básica da cidade de Ilhéus, jan/2005 a dez/2016	63
Figura 20 – Funções de autocorrelação e autocorrelação parcial do processo estacionário do custo da cesta básica da cidade de Ilhéus, jan/2005 a dez/2016	63
Figura 21 – Funções de autocorrelação dos resíduos do modelo ARIMA(2,1,3) para o custo da cesta básica da cidade de Ilhéus, jan/2005 a dez/2016	64
Figura 22 – Análise de resíduos do modelo ARIMA(2,1,3) para o custo da cesta básica da cidade de Ilhéus, jan/2005 a dez/2016. (a) Histograma dos resíduos, (b) Probabilidade normal	65
Figura 23 – Ajustamento do modelo ARIMA(2,1,3) para o custo da cesta básica da cidade de Ilhéus, jan/2005 a dez/2016	66

Figura 24 – Ajustamento do modelo SMMS para o custo da cesta básica da cidade de Ilhéus, jan/2005 a dez/2016	67
Figura 25 – Ajustamento do modelo SES para o custo da cesta básica da cidade de Ilhéus, jan/2005 a dez/2016	68
Figura 26 – Ajustamento do modelo SH para o custo da cesta básica da cidade de Ilhéus, jan/2005 a dez/2016	70
Figura 27 – Ajustamento do modelo SHW aditivo para o custo da cesta básica da cidade de Ilhéus, jan/2005 a dez/2016	70
Figura 28 – Ajustamento do modelo SHW multiplicativo para o custo da cesta básica da cidade de Ilhéus, jan/2005 a dez/2016	71
Figura 29 – Previsão dos métodos estatísticos para o custo da cesta básica da cidade de Ilhéus, jan/2017 a dez/2017	72
Figura 30 – Ajustamento do modelo RNA para o custo da cesta básica da cidade de Ilhéus, jan/2005 a dez/2016	73
Figura 31 – Ajustamento do modelo híbrido ARIMA com RNA para o custo da cesta básica da cidade de Ilhéus, jan/2005 a dez/2016	74
Figura 32 – Ajustamento do modelo híbrido sistema neurofuzzy para o custo da cesta básica da cidade de Ilhéus, jan/2005 a dez/2016	76
Figura 33 – Previsões dos métodos de aprendizagem de máquina para o Preço total da cesta em Ilhéus (Período de 01/2017 a 12/2017)	77
Figura 34 – Erro quadrático médio das previsões do custo da cesta básica da cidade de Ilhéus, jan/2017 a dez/2017	78
Figura 35 – Erros quadráticos médios agrupados nos períodos de quatro, seis e doze meses de previsões para o custo da cesta básica da cidade de Ilhéus, jan/2017 a dez/2017	78

Lista de tabelas

Tabela 1 – Provisões mínimas de alimentos estipuladas pelo Decreto Lei n 399 .	45
Tabela 2 – Resumo de treinamento do sistema neurofuzzy para o preço total da cesta em Ilhéus (Período de 01/2005 a 12/2016)	57
Tabela 3 – Resultado do teste Cox-Stuart	60
Tabela 4 – Teste ADF para primeira diferença da série temporal do custo da cesta básica da cidade de Ilhéus, jan/2005 a dez/2016.	61
Tabela 5 – Acurácia dos ajustamentos dos modelos estatísticos para o preço total da cesta em Ilhéus (Período de 01/2005 a 12/2016)	67
Tabela 6 – Previsões dos métodos estatísticos para o preço total da cesta em Ilhéus (Período de 01/2017 a 12/2017)	69
Tabela 7 – Medidas de acurácias das previsões dos métodos estatísticos para o preço total da cesta em Ilhéus (Período de 01/2017 a 04/2017)	69
Tabela 8 – Medidas de acurácias das previsões dos métodos estatísticos para o preço total da cesta em Ilhéus (Período de 01/2017 a 06/2017)	69
Tabela 9 – Medidas de acurácia das previsões dos métodos estatísticos para o preço total da cesta em Ilhéus (Período de 01/2017 a 12/2017)	69
Tabela 10 – Acurácia dos ajustamentos dos modelos para o preço total da cesta em Ilhéus (Período de 01/2005 a 12/2016)	72
Tabela 11 – Previsões obtidas pelos métodos RNA e as abordagens híbridas para o preço total da cesta em Ilhéus (Período de 01/2017 a 12/2017)	75
Tabela 12 – Medidas de acurácia das previsões para o preço total da cesta em Ilhéus (Período de 01/2017 a 04/2017)	75
Tabela 13 – Medidas de acurácias das previsões para o preço total da cesta em Ilhéus (Período de 01/2017 a 06/2017)	75
Tabela 14 – Medidas de acurácia das previsões para o preço total da cesta em Ilhéus (Período de 01/2017 a 12/2017)	75

Lista de abreviaturas e siglas

UESC	Universidade Estadual de Santa Cruz
PPGMC	Programa de Pós-Graduação em Modelagem Computacional em Ciência e Tecnologia
ACCB	Acompanhamento do Custo da Cesta Básica
RNA	Rede Neural Artificial
ARIMA	Autoregressive Integrated Moving Average
ANFIS	Adaptive-Network-Based Fuzzy Inference System
MSE	Mean Squared Error
MAPE	Mean Absolute Percentage Error
SMAPE	Symmetric Mean Absolute Percentage Error
ADF	Augmented Dickey-Fuller
SMMS	Suavização Exponencial por Médias Móveis Simples
SES	Suavização Exponencial Simples
SH	Suavização Exponencial de Holt
SHWM	Suavização Exponencial de Holt-Winters Multiplicativo
SHWA	Suavização Exponencial de Holt-Winters Aditivo
DIEESE	Departamento Intersindical de Estatística e Estudos Socioeconômicos
MLP	Multi-Layer Perceptron
ICMS	Imposto sobre Circulação de Mercadorias e Serviços
MQO	Mínimo Quadrados Ordinários
GD	Gradiente Descendente

Lista de símbolos

\in	Pertence
Z_t	Série temporal
T_t	Tendência
S_t	Sazonalidade
a_t	Comportamento aleatório ou ruído branco
$\Delta^n Z_t$	Diferença da série
Y_t	Série sem tendência
\hat{S}_t	Estimação sazonal
Z_t^{SA}	Série livre de sazonalidade
$I(f_i)$	Intensidade da frequência
f_i	Frequência
H_0	Hipótese nula
H_1	Hipótese alternativa
σ^2	Variância
$\phi(B)$	Operador autorregressivo
B	Operador de defasagem
p	Coeficiente outorregressivo
q	Número de defasagem de média móvel
θ	Parâmetro de média móvel
$\theta(B)$	Operador de média móvel
d	Ordem de integração
M_t	Estimação do nível
$\hat{Z}_t(h)$	Previsão no instante t e horizonte h
r	Número de observações para a média móvel

S	Erro mínimo
\bar{Z}_t	Suavização exponencial
α	Constante de suavização exponencial simples
A	Constante de suavização exponencial de Holt
C	Constante de suavização exponencial Holt-Winters
h	Número de passos de previsão
s	Período de previsão
F_t	Fator sazonal
D	Constante de suavização exponencial simples-Winters sazonal
w	Peso sináptico
u	Potencial de ativação
β	Constante real
x	Sinal de entrada
E	Erro quadrático médio
d_i^x	Saída desejada
O_i^x	Saída produzida
γ	Taxa de aprendizado
$G(B)$	Centro de gravidade
$\mu_{A_i}(x)$	Função de pertinência

Sumário

1 – Introdução	1
1.1 Problema	3
1.2 Hipótese	3
1.3 Objetivos	4
1.4 Motivação	4
2 – Revisão de Literatura	6
2.1 Séries temporais	6
2.1.1 Objetivos da análise de séries temporais	7
2.1.1.1 Modelos estocásticos	9
2.1.2 Tendência	9
2.1.3 Sazonalidade	11
2.2 Modelos estacionários	13
2.2.1 Modelo autorregressivo (AR)	14
2.2.2 Modelos de média móvels (MA)	14
2.2.3 Modelo autorregressivo integrado de média móvel (ARIMA)	15
2.2.4 Função de autocorrelação fac e função de autocorrelação parcial facp	17
2.3 Modelos de suavização exponencial	18
2.3.1 Suavização por médias móveis simples (SMMS)	19
2.3.2 Suavização exponencial simples (SES)	19
2.3.3 Suavização exponencial de Holt (SH)	20
2.3.4 Suavização exponencial de Holt-Winters (SHW)	21
2.4 Redes neurais artificiais	23
2.4.1 Neurônios artificiais	23
2.4.2 Arquitetura de uma RNA	27
2.4.3 Processos de treinamento e aspectos de aprendizado	29
2.4.4 Redes perceptron multicamadas	30
2.4.4.1 Processo de treinamento da MLP	31
2.5 Noções de lógica fuzzy	33
2.5.1 t-normas e t-conormas	34
2.5.2 Raciocínio aproximado e variáveis linguísticas	35
2.5.3 Controladores fuzzy	36
2.5.4 Rede neurofuzzy	37
2.5.4.1 Arquitetura ANFIS	38
2.6 Estado da arte	40

3 – Metodologia	43
3.1 Dados	44
3.2 Análise da série temporal	46
3.3 Descrição computacional e ajustamento dos modelos	47
3.3.1 Modelo ARIMA	48
3.3.2 Métodos de suavização	49
3.3.3 RNA	50
3.3.4 Modelos híbridos	53
3.3.4.1 ARIMA com RNA	53
3.3.4.2 Sistema neurofuzzy	54
3.4 Previsão	56
4 – Resultados e discussão	59
4.1 Métodos estatísticos	59
4.1.1 Análise da série e modelagem ARIMA	60
4.1.2 Ajustamento dos modelos	64
4.1.3 Previsão	66
4.2 Aprendizagem de máquina	71
4.2.1 Ajustamentos dos modelos	71
4.2.2 Previsão	73
4.3 Comparação entre os modelos	76
5 – Conclusões e trabalhos futuros	79
Referências	80

1 Introdução

Nos últimos anos, o termo governança eletrônica, também conhecido como *e-gov* ou *egoverno*, tem se mostrado bastante promissor como ferramenta para auxiliar na reforma administrativa e como um dos meios da prestação de serviços mais rápidos e com mais eficiência para a população [Souza et al. 2013]. Assim, o governo eletrônico faz uso de novas tecnologias de informação e comunicação na aplicação de funções de governo, que podem visar a diferentes setores, cidadão ou mercado. Administração Pública, como entidade fortemente centrada em atividades que envolvem a coleta, processamento e disponibilização de grandes volumes de informação, bem como o fornecimento de serviços à generalidade dos cidadãos, beneficiou-se de maneira significativa das Tecnologias da Informação.

Em uma etapa mais avançada o governo eletrônico envolve o uso de técnicas computacionais e mineração de dados no suporte à análise de dados e tomada de decisão em relação as políticas públicas e programas governamentais. São vários exemplos de boa utilização dessas tecnologias na interação Estado-Cidadão, tendo como principal função a melhora em termos de eficácia dos serviços prestados. Um bom exemplo, é o Projeto de Extensão Acompanhamento do Custo da Cesta Básica (ACCB) [UESC et al.], desenvolvido pelo Departamento de Ciências Econômicas da Universidade Estadual de Santa Cruz.

Neste contexto, o projeto ACCB é um projeto de extensão que faz o acompanhamento do custo da cesta básica na microrregião Ilhéus-Itabuna. Iniciou-se em março de 1999, e desde então vem acompanhando mensalmente a evolução dos preços de doze produtos de alimentação adotando a metodologia do DIEESE (Departamento Intersindical de Estatística e Estudos Socioeconômicos) [LLC 1955]. A metodologia DIEESE consiste em realizar o acompanhamento em dezoito capitais do Brasil da evolução dos preços de treze produtos de alimentação, necessários para satisfazer às necessidades normais de alimentação, assim como o gasto mensal que cada trabalhador precisa aplicar para comprá-los.

Seguindo a mesma metodologia do DIEESE, o projeto ACCB realiza o acompanhamento do custo da cesta básica e a quantidade de horas que o trabalhador necessita para poder adquirir os produtos. Os dados coletados vem sendo armazenados em uma base de dados desde o seu início. Com isso, têm-se uma sequência de observações referentes ao custo da cesta básica. Essas sequências de observações são denominadas séries temporais, as quais são uma fonte de informação relevante para dar suporte ao processo decisório nas políticas públicas regionais e alicerçar práticas de governo eletrônico na região.

O uso do estudo de séries temporais como auxílio na tomada de decisão é uma prática que vem sendo bastante adotada em diversas áreas. O trabalho de [Sato 2013] propõe avaliações do gerenciamento de doenças infecciosas e não infecciosas através da análise de séries temporais utilizando técnicas estatísticas, o mesmo conclui que o uso desta análise oferece grande contribuição para os gestores os auxiliando na tomada de decisão das intervenções de saúde em dada população.

Em [Sousa et al. 2011], se propõe um modelo para previsão de vendas na indústria retalhista. Para isso, os autores propõem o uso de redes neurais artificiais do tipo *feedforward* para realizar as previsões, podendo assim ajudar a minimizar a incerteza existente a curto e médio prazo no planejamento e distribuição na indústria. Já [Sousa et al. 2012], apresentam um estudo comparando modelos estatísticos clássicos de previsão com um método de inteligência artificial através do uso de rede neural de *Elman* na previsão do preço do etanol no estado de Goiás. Mostrando que os modelos são bastantes competitivos em termos de predição, destacando-se, no entanto, os modelos estatísticos.

Além do uso de métodos com abordagem individualizada, para realizar à análise e previsão de séries temporais, tem-se a combinação de métodos os quais tendem a fornecer previsões mais precisas. Em [Silva 2012], se apresenta uma metodologia híbrida usando métodos estatísticos com métodos de aprendizagem de máquina fornecendo previsões mais acuradas para a previsão de cargas elétricas a curto prazo, auxiliando no planejamento e operação de sistemas elétricos. Já [Faustino 2011], apresenta a construção de um modelo híbrido baseado em regras fundamentadas em conceitos da lógica *fuzzy* e redes neurais artificiais, mostrando que o mesmo apresentou melhores resultados quando comparado a modelos tradicionais de predição apresentados na literatura.

Existem diversas técnicas para realizar o estudo e análise de séries temporais. Muitas delas têm sido utilizadas com sucesso pelo mundo acadêmico e empresarial na gestão de organizações e empresas. Entretanto, ainda não são utilizadas de forma sistemática pelos gestores públicos brasileiros. Nesse contexto, o objeto de estudo deste trabalho consistiu em realizar a análise das séries temporais históricas dos dados do projeto cesta básica usando técnicas computacionais.

Para realizar esta análise, foram feitas o uso de técnicas avançadas de modelagem computacional, especificamente mineração de dados, na análise e previsão de séries temporais de dados gerados no âmbito do projeto ACCB. Para isso, inicialmente as séries temporais foram estudadas utilizando métodos estatísticos de suavização e do tipo *Box-Jenkins*. Em seguida foram estudadas utilizando técnica de redes neurais artificiais, e logo depois através de dois métodos híbridos que combinaram técnicas de redes neurais artificiais com modelo ARIMA e redes neurais artificiais com sistema de inferência *fuzzy*.

Dessa forma, considerando as séries temporais de dados, geradas no âmbito

do projeto ACCB, deve-se determinar qual ou quais técnicas apresentam melhores resultados na análise e previsão, considerando os critérios de acurácia. Para isso, buscou conhecer o comportamento do mecanismo gerador de dados, verificando tendências e sazonalidade, obtenção do modelo de ajustamento e previsão de valores futuros. Este conhecimento acerca dos dados do projeto ACCB, podem ser utilizados pelos órgãos competentes, principalmente na gestão pública, dando suporte à análise de dados e tomada de decisão em relação a políticas públicas e programas governamentais.

A estrutura desta dissertação está assim organizada: o primeiro capítulo apresenta esta introdução juntamente com o problema, hipótese, objetivos e motivação; o capítulo 2 apresenta o referencial teórico necessário para o desenvolvimento da pesquisa; o capítulo 3 apresenta a metodologia abordada, apresentado as etapas seguidas para chegar aos resultados; o capítulo 4 são expostos os resultados obtidos; o capítulo 5 são apresentadas as conclusões obtidas através deste estudo.

1.1 Problema

A mineração de dados é um processo de descoberta de conhecimento adquirido na realização da análise de dados, e na aplicação de algoritmos específicos para a extração de modelos, que sob certas limitações computacionais, produzem um conjunto de padrões de certos dados [Fayyad et al. 1996]. As técnicas de mineração consistem na especificação de métodos que garantam como descobrir os padrões persistentes nos dados em análise. Dentre as principais técnicas utilizadas em mineração de dados, destacamos as técnicas estatísticas e técnicas de aprendizado de máquina. Existindo uma grande variedade de abordagens, um dos principais problemas na mineração de dados envolve decidir qual e ou quais técnicas geram os melhores resultados em uma determinada análise. Dessa forma, considerando as séries temporais de dados, geradas no âmbito do projeto Acompanhamento do Custo da Cesta Básica da microrregião Ilhéus-Itabuna, deve-se determinar qual é a técnica de mineração de dados que apresenta melhores resultados na análise e previsão considerando critérios de acurácia?

1.2 Hipótese

Técnicas estatísticas e de aprendizado de máquina podem ser utilizadas satisfatoriamente na análise de dados de séries temporais históricas do custo da cesta básica das cidades de Ilhéus. Dentre estas técnicas temos as técnicas estatísticas e técnicas de aprendizado de máquina, que serão implementadas e validadas visando avaliar o mecanismo gerador de dados; descrevendo o comportamento; verificando as tendências, ciclos e sazonalidade; e a previsão de valores futuros. A comparação entre as

técnicas permitirá estabelecer qual é a técnica de mineração de dados a ser usada na caracterização e análise desse tipo de série temporal de dados.

1.3 Objetivos

O objetivo geral deste projeto é utilizar técnicas avançadas de modelagem computacional, especificamente mineração de dados, na análise e previsão de séries temporais de dados, geradas no âmbito do projeto Acompanhamento do Custo da Cesta Básica da microrregião Ilhéus-Itabuna. Os resultados desse projeto têm um forte impacto social e contribuem a fortalecer a governança eletrônica, podendo subsidiar a gestão pública regional oferecendo suporte à tomada de decisão, ao lançamento de programas e políticas públicas e governamentais.

Os objetivos específicos deste projeto são:

- (i) Implementar e avaliar o uso de métodos estatísticos (suavização por médias móveis simples, suavização exponencial simples, suavização exponencial de Holt, suavização exponencial de Holt-Winters e métodos de *Box-Jenkins*) na caracterização e previsão de séries temporais do custo da cesta básica nas cidade de Ilhéus;
- (ii) Propor, desenvolver e avaliar o uso de redes neurais artificiais na caracterização e previsão de séries temporais do projeto ACCB na cidades Ilhéus;
- (iii) Propor, desenvolver e avaliar o uso de métodos híbridos que combinem RNA com métodos estatísticos e sistema de inferência *fuzzy* na análise de séries de dados geradas no projeto ACCB da UESC;
- (iv) Avaliar os métodos propostos na tarefa de previsão usando medidas de acurácia, estabelecendo qual abordagem oferece melhores resultados.

1.4 Motivação

As profundas mudanças provocadas na gestão dos governos com o dramático desenvolvimento das tecnologias da informação, orientou-se o governo eletrônico a uma maior transparência na gestão dos gastos públicos e a melhorar a qualidade dos serviços ao cidadão [Chahin 2004, Ferrer e Santos 2004]. No Brasil, principalmente o governo federal tem lançados múltiplas iniciativas de *e-gov*, com destaque para o Portal de Transparência [Federal 2004] e o Sistema Público de Escrituração Digital [Federal 2007]. Na segunda etapa, o governo eletrônico inclui o uso de técnicas computacionais avançadas como ferramentas para tomada de decisão e gestão de políticas públicas.

Estas técnicas, dentre as outras, incluem inteligência artificial, reconhecimento de padrões, modelagem computacional e computação bioinspirada. Muitas delas têm sido utilizadas com sucesso pelo mundo acadêmico e empresarial na gestão de organizações e empresas. Entretanto, ainda não são utilizadas de forma sistemática pelos gestores públicos brasileiros. Nesse contexto, o objeto de estudo desta proposta é a análise das séries temporais históricas dos dados do projeto cesta básica usando técnicas computacionais. Com isso, busca conhecer o comportamento do mecanismo gerador de dados, verificando tendências, ciclos e sazonalidade; prever valores futuros e procurar propriedades relevantes. Esses conhecimentos acerca dos dados analisados, podem ser utilizados pelos órgãos competentes, principalmente pela gestão pública, dando suporte à análise de dados e tomada de decisão em relação a políticas públicas e programas governamentais. Isso é de grande relevância, principalmente para a região Sul da Bahia, que juntamente com todas as outras sub-regiões do nordeste, possui baixos indicadores socioeconômicos comparados com o restante do país.

2 Revisão de Literatura

Neste capítulo apresentaremos os conhecimentos teóricos necessários para o desenvolvimento desta pesquisa. Discorreremos sobre o conceito de séries temporais, apresentando os objetivos de sua análise e sua decomposição clássica. Depois, apresentamos os principais métodos estatísticos e de aprendizagem de máquina, abordado nesta pesquisa.

2.1 Séries temporais

Uma série temporal é definida como qualquer conjunto de observações de uma variável ordenadas no tempo. Assim, as séries temporais podem ser classificadas como discretas ou contínuas. Matematicamente, uma série temporal é definida como,

$$Z_t = \{Z_t \in \mathbb{R} | t = 1, 2, 3 \dots n\}, \quad (1)$$

onde t é um índice temporal e n é o número de observações.

Uma série temporal é dita discreta quando o conjunto de observações no tempo for finito ou infinito enumerável. Caso contrário, a série for infinita e não enumerável é classificada como contínua. Além disso, uma série temporal pode ser classificada de acordo com a sua formação, isto é, podemos classificar uma série temporal como determinística quando os valores futuros são estabelecidos por uma relação funcional matemática do tipo $y = f(t)$ ou estocástica, quando os valores futuros são apresentados em termos probabilísticos, onde a série é descrita por meio de uma relação matemática funcional que envolve não só o tempo, mas também uma variável aleatória do tipo $y = f(t, a_t)$, em que a_t é a variável aleatória residual [Morettin e Toloí 2006]. A seguir, serão apresentados alguns exemplos de séries temporais:

- i valores diários de poluição na cidade de Ilhéus;
- ii valores mensais de temperaturas na cidade de Itabuna;
- iii índices diário da Bolsa de Valores de São Paulo;
- iv número médio anual de manchas solares;
- v registro de marés no porto de Ilhéus.

Nos itens i, ii, iii e iv são exemplos de séries temporais discretas, enquanto que o item v é um exemplo de uma série contínua. Uma série temporal discreta pode ser obtida

através de uma série temporal contínua. Para isso, basta tomar intervalos de tempos iguais da série temporal contínua. Assim, para realizar a análise da série temporal do item v , será necessário amostrá-la em intervalos de tempos iguais, convertendo-a em uma série contínua.

2.1.1 Objetivos da análise de séries temporais

O objetivo de se aplicar técnicas de previsão de séries temporais dada uma série Z_t , é identificar padrões presentes no conjunto de pontos da série, criando um modelo que seja capaz de construir os próximos pontos temporais. Assim, dada uma série temporal $Z(t_1), \dots, Z(t_n)$, observada nos instantes t_1, \dots, t_n , podemos estar interessados em:

$$Z_t = \{Z_t \in \mathbb{R} | t = 1, 2, 3 \dots N\}, \quad (2)$$

- investigar o mecanismo gerador da série temporal; por exemplo, analisando uma série de alturas de ondas, querendo determinar como estas ondas foram geradas;
- fazer previsões de valores futuros da série; estas podem ser a curto prazo, como para séries de vendas, produção ou estoque, ou ao longo prazo, como para séries populacionais, de produtividade etc;
- descrever apenas o comportamento da série; neste caso, a construção do gráfico, a verificação de histograma e diagramas de dispersão etc., podem ser ferramentas úteis para a análise.
- Procurar propriedades relevantes nos dados, como periodicidades, sazonalidade, tendência, componentes aleatórias e níveis.

A análise de séries temporais tem como objetivo construir modelos para a série, com propósitos determinados, como descrever o comportamento da mesma, investigar o mecanismo gerador da série, fazer previsão de valores futuros e procurar periodicidades relevantes nos dados. Um modelo clássico assume que uma série temporal $\{Z_1, \dots, Z_n\}$ pode ser escrita como

$$Z_t = T_t + S_t + a_t, \quad t = 1, 2, \dots, n \quad (3)$$

sendo que Z_t representa a série temporal, T_t a tendência, S_t a sazonalidade e a_t a componente aleatória.

[Morettin e Toloi 2006] definem a tendência como o aumento ou diminuição gradual das observações ao longo do período. Já a sazonalidade pode ser entendida

como flutuações ocorridas em períodos menores ou iguais a doze meses e a componente aleatória mostra oscilações aleatórias irregulares.

O Modelo (3) é dito aditivo devido a componente sazonal ser independente das outras componentes. Se as amplitudes sazonais variam de acordo com a tendência, o modelo é dito multiplicativo e definido como

$$Z_t = T_t S_t a_t, \quad t = 1, 2, \dots, n. \quad (4)$$

Nesse caso, a variância não é constante, sendo necessária uma transformação de dados. Tomando-se o logaritmo de (4), tem-se

$$Z_t^* = \log T_t + \log S_t + \log a_t \quad t = 1, 2, \dots, n, \quad (5)$$

em que Z_t^* é o $\log Z_t$.

Para ajustar os modelos de *Box-Jenkins* a uma série temporal Z_t , é necessário que ela seja estacionária, ou seja, se desenvolva aleatoriamente no tempo em torno de uma média constante. A grande maioria das séries temporais apresentam componentes como tendência e ou sazonalidade. Para que a série se desenvolva aleatoriamente no tempo em torno de uma média constante, um procedimento comum é tomar diferenças sucessivas da série original. A primeira diferença é dada por

$$\Delta Z_t = Z_t - Z_{t-1} \quad t = 1, 2, \dots, n. \quad (6)$$

A segunda diferença aparece na forma

$$\Delta^2 Z_t = Z_t - 2Z_{t-1} + Z_{t-2} \quad t = 1, 2, \dots, n. \quad (7)$$

Generalizando, a n-ésima diferença é dada por

$$\Delta^n Z_t = \Delta \{ \Delta^{n-1} Z_t \} \quad t = 1, 2, \dots, n. \quad (8)$$

Normalmente, são necessárias uma ou duas diferenças para que a série se torne estacionária [Morettin e Toloi 2006].

Para [Morettin e Toloi 2006], a palavra previsão sugere que se quer ver uma coisa antes de ela existir. Com tudo, a previsão não constitui um fim em si, mas um meio de oferecer informações para uma consequente tomada de decisão, visando determinados objetivos, em que os procedimentos de previsão utilizados na prática variam muito, podendo ser simples e interativos ou mais complexos.

2.1.1.1 Modelos estocásticos

Definimos um processo estocástico como um modelo matemático utilizado para o estudo de fenômenos aleatórios que têm como resultados funções. Estas funções recebem como parâmetros um conjunto arbitrário T a qual é definida como função trajetória. Com isso, um processo estocástico pode ser definido como sendo uma coleção de variáveis aleatórias ordenadas no tempo $\{x_t, t, \in T\}$, onde T é um conjunto ordenado de índices.

De maneira mais geral, o processo estocástico é uma variável que se comporta durante o tempo, pelo menos parte deste comportamento é considerado aleatório. Isto é, o comportamento é definido pela probabilidade da evolução x_t da variável x durante o tempo t . Uma série temporal pode ser considerada como a realização de um processo estocástico e que as variáveis estão ordenadas em intervalos regulares de tempo $z_t = \{z_1, z_2, z_3, \dots, z_t\}$. Se definimos as funções de distribuição conjunta das variáveis aleatórias $\{z_1, z_2, z_3, \dots, z_t\}$ para qualquer valor t , fica caracterizado o processo estocástico [Morettin e Toloi 2006].

Um processo estocástico é classificado em dois processos, são eles: processos em tempo discreto e processo em tempo contínuo. O processo estocástico em tempo discreto é aquele que o valor da variável varia em determinado instante de tempo. Já o processo estocástico em tempo contínuo, a variável possui valores em qualquer instante de tempo. Nos casos em que tem o processo em tempo discreto, tem-se variáveis discretas e nos casos em que tem o processo em tempo contínuo, tem-se variáveis contínuas.

2.1.2 Tendência

Na ausência da componente sazonal, o Modelo (3) se reduz a

$$Z_t = T_t + a_t, t = 1, 2, \dots, n. \quad (9)$$

Em que a_t é o ruído branco. [Morettin e Toloi 2006] apresentam vários métodos para estimar a tendência, dentre eles, o ajuste polinomial e suavização (filtro).

Uma vez estimada a tendência \hat{T}_t , a série pode ser escrita livre da mesma através da expressão

$$Y_t = Z_t - \hat{T}_t, t = 1, 2, \dots, n. \quad (10)$$

Já um outro procedimento mais prático para eliminar a tendência é tomar a diferença. Conforme definido nas equações (6),(7) e (8)

Usando a análise gráfica visual é possível avaliar se uma série temporal possui tendência. Entretanto, é necessário confirmar esta avaliação prévia. Para isto, são usados dois procedimentos: a análise do gráfico da autocorrelação e ou o teste para tendência. Entre os testes para comprovar a existência da tendência apresentados por [Morettin e Toloi 2006] têm-se: teste do sinal (Cox-Stuart) e o teste F para análise de variância.

O teste Cox-Stuart é usado para verificar a existência da componente tendencial em uma série temporal, o qual é realizado através dos seguintes passos:

1. Agrupar as observações em pares $(Z_1, Z_{c+1}), (Z_2, Z_{c+2}), \dots, (Z_{N-c}, Z_N)$, em que $c = \frac{N}{2}$ se N for um número par e $c = \frac{N+1}{2}$ se N for ímpar;
2. Para cada par (Z_i, Z_{c+i}) , associa-se um sinal. Este sinal é positivo (+) se $Z_i < Z_{c+i}$ e negativo (-) se $Z_i > Z_{c+i}$. Descarta-se a comparação em que $Z_i = Z_{c+i}$;
3. Testa a hipótese de nulidade.

H_0 : Não existe tendência. A estatística do teste a ser utilizado é R , sendo que R é igual ao número de sinais positivos atribuídos no final da comparação entre Z_i e Z_{c+i} ;

4. Calcula-se o valor de r , que é dado através da expressão

$$r = \frac{1}{2}(n + \omega_{\alpha/2}\sqrt{n}) \quad (11)$$

em que n é o número de vezes que não houve empate na comparação entre Z_i e Z_{c+i} e $\omega_{\alpha/2}$ é o valor tabelado ao nível de significância α . Caso $n \leq 20$, o valor tabelado de $\omega_{\alpha/2}$ é dado pela distribuição binomial com $p = 0,5$ e, caso $n > 20$, utiliza-se a distribuição normal;

5. Se $R \leq r$ ou $R \geq n - r$, rejeita H_0 , ou seja, a série possui tendência.

Quando uma série temporal é estacionária, os resultados da estatística tradicional são válidos, porém, se a série apresenta raiz unitária há estimadores que não corresponde ao padrão da série, comprometendo a validade dos resultados [MARGARIDO e ANEFALOS 1999]. Logo, o teste da raiz unitária torna-se importante para verificar se a integração foi suficiente para tornar a série não-estacionária em uma série estacionária.

De acordo com [Brooks 2014], o teste Augmented Dickey-Fuller (ADF) tem como objetivo básico examinar a hipótese nula de que $\pi = 1$, contra a alternativa $\pi < 1$, na expressão

$$\Delta y_t = \pi y_{t-1} + a_t \quad (12)$$

As hipóteses de interesse são: H_0 : a série possui raiz unitária; H_1 : a série é estacionária.

Segundo [Sousa et al. 2012], o teste ADF verifica a existência de uma raiz unitária em y_t , quando o processo gerador da série é expresso por um dos três modelos dados por

$$\Delta y_t = \pi y_{t-1} + a_t \quad (13)$$

$$\Delta y_t = \alpha + \pi y_{t-1} + a_t \quad (14)$$

$$\Delta y_t = \alpha + \beta_t + \pi y_{t-1} + a_t \quad (15)$$

Assim, a Eq. (13) demonstra um processo estocástico com caminho aleatório puro. O modelo (14) adiciona um intercepto α e (15) inclui o intercepto α e uma tendência linear β_t .

O coeficiente estimado π é dividido por seu erro padrão para calcular a estatística τ de ADF, onde τ para o modelo (13), τ_μ para o modelo (14) e τ_τ para o modelo (15). Depois de obtido o valor, o mesmo é comparado com o valor tabelado de DF para verificar se a hipótese nula $\pi = 1$ é rejeitada, isto é, a série é estacionária.

Os testes ADF consistem em estimar as Eq. (13) a (15) através de mínimos quadrados ordinários e comparar as estatísticas t resultantes, aos valores críticos gerados por *Dickey – Fuller*.

2.1.3 Sazonalidade

A sazonalidade é uma componente presente na grande maioria das séries temporais, onde as características ocorrem regularmente dentro do período de um ano, ou seja, características que ocorrem frequentemente em período de no máximo doze meses. [Morettin e Toloi 2006] descrevem os procedimentos mais comuns para se estimar a sazonalidade que são através do método de regressão (sazonalidade determinística) e do método de média móveis para sazonalidade estocástica.

A presença da sazonalidade, a depender do método de previsão abordado, pode produzir previsão com baixa qualidade, uma vez que o método usado não consegue modelar o comportamento sazonal. Com isso, torna-se necessário a estimação do fator

sazonal \hat{S}_t , e após isso, pode-se escrever a série temporal livre dela. Com isso, pode-se estimar o modelo livre de sazonalidade, no caso aditivo, através de

$$Z_t^{SA} = Z_t - \hat{S}_t, t = 1, 2, \dots, n, \quad (16)$$

e se for multiplicativo, através da expressão

$$Z_t^{SA} = \frac{Z_t}{\hat{S}_t}, t = 1, 2, \dots, n. \quad (17)$$

Além disso, a componente sazonal também pode ser constatada através da análise visual do gráfico da série e para se confirmar a existência da mesma, deve-se aplicar um teste para a sazonalidade. Dentre os teste para identificar a sazonalidade, tem-se: teste de Fridman, teste de Kruskal-Wallis, teste F para Análise de Variância descrito em [Morettin e Toloi 2006] e teste de Fisher, encontrado em [Priestley 1981].

Através da análise do gráfico do periodograma também é possível detectar a existência de sazonalidade. Para [Morettin e Toloi 2006], esta análise é fundamental quando o interesse é a procura de periodicidade nos dados. Essa análise é feita no domínio da frequência, portanto, é baseada em modelos não paramétricos.

[Priestley 1981] define o periodograma do processo estacionário (Z_t) como

$$I_p(f_i) = \frac{2}{n} \left[\left(\sum_{t=1}^n Z_t \cos \frac{2\pi i}{n} t \right)^2 + \left(\sum_{t=1}^n Z_t \sin \frac{2\pi i}{n} t \right)^2 \right], \quad (18)$$

onde $0 < f_i < \frac{1}{2}$ e $t = 1, 2, \dots, n$ e $I(f_i)$ representa a intensidade da frequência f_i . Para verificar a periodicidade, observa-se a existência de picos na frequência $f_i = \frac{i}{n}$, o que indica uma periodicidade de período $\frac{1}{f_i}$.

O gráfico do periodograma, o eixo das ordenadas representa a frequência f_i e no eixo das abscissas a intensidade da frequência $I(f_i)$, onde o pico de maior intensidade é o componente periódico. Caso exista mais de um pico, aplica-se o teste de Fisher para verificar se o pico é uma componente periódica genuína.

Este teste de Fisher é utilizado para confirmar a existência da sazonalidade em uma série temporal. [Priestley 1981] define que este teste utiliza os valores do periodograma e detecta as grandes periodicidades.

A aplicação do teste acompanha o seguinte roteiro:

1. Traçar o periodograma por meio da frequência padrão dada pela Eq. (18)

2. Tomar a maior periodicidade encontrada no periodograma, $\max(I_p)$ e depois calcular a estatística usando a expressão

$$g = \frac{\max I_p}{\sum_{p=1}^{N/2} I_p} \quad (19)$$

3. Calcular a estatística do Teste de Fisher, z_α dada por

$$z_\alpha = \left(\frac{\alpha}{n} \right)^{\frac{1}{(n-1)}} + 1 \quad (20)$$

em que $n = N/2$ e α é o nível de significância;

4. Testar a hipótese de nulidade.

H_0 : Não existe periodicidade.

Se $g > z_\alpha$, rejeita-se H_0 , ou seja, existe sazonalidade no período analisado.

Outro modelo tradicional para determinar a sazonalidade é o modelo sazonal determinístico. Este modelo pressupõe um padrão sazonal regular e estável no tempo com picos e vales sempre localizados nos mesmos trimestres.

O processo trimestral m_t é um processo de sazonalidade determinística quando sua representação é expressa da seguinte forma

$$m_t = \sum_{j=1}^4 \Theta_j P_{tj} + Z_t, t = 1, 2, \dots, n, \quad (21)$$

com $P_{tj} (j = 1, \dots, 4)$ representa as usuais *dummies* sazonais, Z_t um processo com média nula e fracamente estacionária e Θ_j representa a taxa média da variação trimestral da série original. Embora Z_t possa conter também alguma sazonalidade, o ideal é ser representado pelos coeficientes das *dummies* sazonais que representam a média do processo para cada trimestre, ou seja, representa o ciclo sazonal presente na série.

2.2 Modelos estacionários

A metodologia de *Box-jenkins* descrevem tanto o comportamento estacionário como o não-estacionário. Isso permite que esta metodologia possua uma grande flexibilidade para tratar séries temporais de comportamento mais geral, em que as previsões com base nesta técnica são feitas a partir dos valores correntes e passados da série [Mortin e Toloi 2006].

Assim, o objetivo fundamental deste método é encontrar uma fórmula apropriada para que os resíduos sejam tão pequenos quanto possível e não apresentem padrões.

Esta metodologia é resultante da combinação dos componentes autoregressivo (AR), integração (I) e de médias móveis (MA). A seguir, será feita uma breve introdução dos modelos autorregressivo AR e de média móvel MA, necessários para a formulação do modelo ARIMA.

2.2.1 Modelo autorregressivo (AR)

O modelo autorregressivo (AR) considera que a variável endógena de um período é explicada por seus valores passados acrescido de um termo de ruído branco (parte aleatória). Para processos estacionários com distribuição normal, este modelo é dado por

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p} + a_t \quad (22)$$

onde Z_t é valor observado da série temporal no instante t , ϕ_i é o i -ésimo parâmetro autorregressivo estimado, p é o ordem do modelo autorregressivo e $i = 1, 2, \dots, p$.

Se os valores observados Z_t pertencem a uma série temporal estacionária, então os parâmetros autorregressivo $\phi_1, \phi_2, \dots, \phi_p$ são constantes, tal que $\phi_p \neq 0$, o termo de erro a_t possui média igual a zero e variância constante σ_a^2 . Se a média for diferente de zero, Z_t deve ser substituído por $Z_t - \mu$ na expressão (22), resultando em

$$Z_t - \mu = \phi_1 (Z_{t-1} - \mu) + \phi_2 (Z_{t-2} - \mu) + \dots + \phi_p (Z_{t-p} - \mu) + a_t \quad (23)$$

Uma forma mais simplificada de escrever a Eq. (23) é através do operador de defasagem B [Morettin e Toloí 2006] dado por

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) Z_t = a_t, \quad (24)$$

onde $BZ_t = Z_{t-1}$ e $B^2 Z_t = B(BZ_t) = BZ_{t-1} = Z_{t-2}$.

Usando o operador de defasagem, a Eq. (22) é reescrita da seguinte forma

$$\phi(B) Z_t = a_t. \quad (25)$$

2.2.2 Modelos de média móvels (MA)

O modelo Média Móvels tem como resultado a série Z_t que é resultado da combinação linear dos choques aleatórios ocorridos no período corrente e passado.

[Morettin e Toloi 2006] definem o modelo genérico envolvendo q valores defasados de a_t e, sua notação é dada por $MA(q)$, através da equação

$$Z_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q}, \quad (26)$$

em que: Z_t é o valor observado da série temporal no instante t ; $\theta_1, \theta_2, \dots, \theta_q$ ($\theta_q \neq 0$) são os parâmetros de média móvel estimados, $i = 1, 2, \dots, q$; $a \sim N(0, \sigma_a^2)$, ou seja, ruído branco no instante t ; q é o número de defasagens da média móvel.

Partindo da Eq. (26) é possível reescrever o processo $MA(q)$ na forma equivalente

$$Z_t = \theta(B)a_t, \quad (27)$$

em que o operador de média móvel é dado por

$$\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q. \quad (28)$$

2.2.3 Modelo autorregressivo integrado de média móvel (ARIMA)

A metodologia *Box-Jenkins*, também conhecida como Modelo Autoregressivo Integrado de Média Móvel, ou ARIMA (*Autoregressive Integrated Moving Average*) foi proposto por [Box et al. 2015]. Estes modelos partem da ideia de que os valores passados da série temporal são fortemente dependentes, isto é, cada um de seus valores podem ser explicados por pontos anteriores da própria série.

Sua representação é dada por $ARIMA(p, d, q)$, onde p é o número de defasagens da série, d é a ordem de integração e q o número da combinação dos processos AR e MA, são determinados com base em funções de autocorrelação parcial e da autocorrelação amostral [Morettin e Toloi 2006].

Para representar o modelo $ARIMA(p, d, q)$, a variável Z_t é substituída por \tilde{Z}_t que representa uma série estacionária, de acordo com

$$\tilde{Z}_t = \phi_1 \tilde{Z}_{t-1} + \dots + \phi_p \tilde{Z}_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}. \quad (29)$$

Portanto, $\tilde{Z}_t = \Delta^d Z_t = (1 - B)^d Z_t$, ou seja, a série é estacionária de ordem d . Já $\theta_1, \dots, \theta_q$ são os parâmetros de média móvel de ordem q e ϕ_1, \dots, ϕ_p são os parâmetros autorregressivo de ordem p

Na forma generalizada, o modelo é definido por

$$\phi(B)(1 - B)^d \tilde{Z}_t = \theta(B)a_t \quad (30)$$

Segundo [Morettin e Toloi 2006], para fazer uso da metodologia *Box-Jenkins* é necessário descobrir qual entre as várias versões do modelo ARIMA, o melhor que descreva o comportamento da série. Este processo é definido em etapas, sendo a primeira a identificação, onde se determina os valores de p , d e q do modelo. As fases posteriores são a estimação dos parâmetros e a validação.

Para [Morettin e Toloi 2006], a identificação é a parte mais crítica do método *Box-Jenkins*. As funções de autocorrelação e autocorrelação parcial estimadas servirão como base para a escolha dos parâmetro p e q . Este procedimento é eficiente, porém, é importante salientar que pode ser identificado mais de um modelo para a mesma série temporal.

Assim, o objetivo desta fase é determinar os valores de p , d e q do modelo ARIMA(p, d, q). O procedimento de identificação consistem em:

1. Verificar se a série é estacionária, por meio da análise gráfica da série original, função de autocorrelação e do teste Cox-Stuart;
2. Diferenciar a série até que se torne estacionária por meio do operador diferença. O número d de diferenças necessárias é alcançado quando a *fac* amostral da série decai rapidamente para zero;
3. Identificar os valores de p e q por meio da análise das funções de autocorrelação e autocorrelação parcial.

Segundo [Lütkepohl e Xu 2012], em alguns casos é necessário também uma transformação logarítmica para obter-se homogeneidade de variância quanto de uma distribuição dos dados mais próximos da distribuição normal. Esta transformação, segundo [Morettin e Toloi 2006], propociona previsões mais acurada dos modelos ARIMA quando comparados às previsões realizadas com dados não transformados. Esta verificação é realizada de forma visual através da representação gráfica da média versus amplitude ou desvio padrão versus amplitude do conjunto de dados. Divide-se a série original em subconjuntos para o cálculo de tais estatísticas para a construção do gráfico. Segundo [Morettin e Toloi 2006], se a amplitude for independente da média, não há necessidade de aplicar a transformação, mas se a amplitude for propocional a média, neste caso é necessário aplicar a transformação logarítmica.

Construido o modelo, o passo seguinte consiste na verificação da adequação do modelo ajustado. O principal procedimento utilizado nessa fase é análise dos resíduos. Se os resíduos apresentarem comportamento de um processo de ruído branco, deverão ser uma sequência de variáveis aleatórias não correlacionadas, provenientes de uma distribuição fixa, com média constante e variância constante. Assim, as hipóteses so-

bre a componente residual são verificados através de representações gráficas e testes estatísticos adequados.

Assim, deve ser feita a análise gráfica das funções de autocorrelação e autocorrelação parcial amostrais, onde deve revelar uma média próximo de zero e não apresentar movimentos periódicos, sem picos fora do intervalo de confiança. A normalidade da componente residual é verificada através do histograma dos resíduos e da representação em papel de probabilidades.

Caso os resíduos tenham todas as características de ruído branco, o modelo em causa é usado para obter a previsão.

2.2.4 Função de autocorrelação fac e função de autocorrelação parcial $facp$

Os valores fac caem entre -1 e +1, calculados da série temporal nas diferentes defasagens para medir a significância das correlações entre a observação presente e as observações passadas e determinar o quanto a se voltar no tempo, isto é, de quantas defasagens no tempo eles estão correlacionados.

Já os valores $facp$ são os coeficientes de uma regressão linear da série temporal usando valores defasados como variáveis independentes. Os valores de $facp$ também caem entre -1 e 1, se a série temporal for estacionária.

Em [Morettin e Toloi 2006] apresenta o processo para a obtenção das funções de autocorrelação e autocorrelação parcial. Multiplicando ambos os membros da Eq. (29) por \tilde{Z}_{t-k} e tomando-se a esperança, obtém a função de autocovariância (fac) de lag k , dado por

$$\begin{aligned}\gamma_k &= E[\tilde{Z}_t \tilde{Z}_{t-k}] \\ &= E[(\phi_1 \tilde{Z}_{t-1} + \dots + \phi_p \tilde{Z}_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}) \tilde{Z}_{t-k}] \\ &= \phi_1 \gamma_{k-1} + \dots + \phi_p \gamma_{k-p} + \gamma_{za}(k) - \theta_1 \gamma_{za}(k-1) - \dots - \theta_q \gamma_{za}(k-q).\end{aligned}\quad (31)$$

onde γ_{za} é a covariância cruzada entre \tilde{Z}_t e a_t , definida por $\gamma_{za} = E[\tilde{Z}_{t-ka_t}]$.

Como \tilde{Z}_{t-k} depende apenas dos choques a_t ocorridos até o instante $t - k$, obtém-se

$$\gamma_{za} \begin{cases} = 0 & \text{se } k > 0, \\ \neq & \text{se } k \leq 0. \end{cases}\quad (32)$$

E para $k > q$, tem-se

$$\gamma_k = \phi_1 \gamma_{k-1} + \dots + \phi_p \gamma_{k-p}\quad (33)$$

dividindo (33) por $\gamma_0 = \text{Var}(\tilde{Z}_t)$, obtém a função de autocorrelação:

$$\rho_k = \phi_1 \rho_{k-1} + \dots + \phi_p \rho_{k-p}. \quad (34)$$

De (33), pode-se deduzir que as autocorrelações de lags $k = 1, 2, \dots, q$ são afetadas diretamente pelos parâmetros de médias móveis e para $k > q$ se comportam como uma mistura de polinômios, exponenciais e senóides amortecidas.

Em [Box et al. 2015] pode-se verificar que, para $q < p$, a *fac* consiste em uma mistura de exponenciais e senóides amortecidas, porém, se $q \geq p$, os primeiros $p - q - 1$ valores de $\rho_0, \rho_1, \dots, \rho_{q-p}$ não seguirão este padrão. [Box et al. 2015] propõem a utilização da função de autocorrelação parcial (*facp*) como para identificação do modelo.

O *facp* é utilizado para medir o grau de associação entre as observações Z_t e Z_{t-k} , quando os efeitos das defasagens até $k - 1$ são fixadas. A obtenção dos coeficientes de autocorrelação parcial são obtidos através das equações de Yule-Walker, expressas na forma matricial por

$$\begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_k \end{bmatrix} \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \rho_2 & \cdots & \rho_{k-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & 1 \end{bmatrix}^{-1} = \begin{bmatrix} \phi_{k1} \\ \phi_{k2} \\ \vdots \\ \phi_{kk} \end{bmatrix}, \quad (35)$$

em que ϕ_{kj} é o j -ésimo coeficiente de um modelo $\text{AR}(k)$ e ϕ_{kk} é o último coeficiente.

2.3 Modelos de suavização exponencial

Os métodos de suavização baseiam-se na idéia que as observações contêm conhecimento sobre o padrão que descreve a série temporal. Logo, estes métodos distinguem o padrão persistente de qualquer ruído que possa estar contido nas observações e, então usa este padrão para prever valores futuros da série.

As técnicas deste tipo assumem que os valores dos extremos da série apresentam aleatoriedade e, por meio da suavização desses extremos, pode-se identificar o padrão básico presente na série e inferir novos valores. Segundo [Morettin e Toloi 2006], esses métodos são de grande popularidade devido à simplicidade, à eficiência computacional e à razoável precisão.

2.3.1 Suavização por médias móveis simples (SMMS)

A técnica por média móvel simples consiste em calcular a média aritmética das r observações mais recentes, isto é, de acordo com a expressão

$$M_t = \frac{Z_t + Z_{t-1} + \dots + Z_{t-r+1}}{r}. \quad (36)$$

Portanto, M_t é uma estimativa do nível que não leva em conta as observações mais antigas. A previsão de todos os valores futuros é dada pela última média móvel calculada conforme a expressão

$$\hat{Z}_t(h) = \hat{Z}_{t-1}(h+1) + \frac{Z_t - Z_{t-r}}{r}, \forall h > 0. \quad (37)$$

As propriedades deste método dependem do número de observações utilizadas para calcular a média, ou seja, o valor de r . Se r for um valor muito grande, isso fará com que a previsão acompanhe lentamente as mudanças presentes na série; um valor pequeno implica numa reação mais rápida [Morettin e Toloi 2006].

Assim, deve-se escolher o valor de r que minimize o erro de previsão. De acordo com [Morettin e Toloi 2006] uma maneira objetiva de escolher o valor de r é selecionar o valor que forneça a "melhor previsão" a um passo das observações já obtidas, isto é, encontrar o valor de r que minimize

$$S = \sum_{t=l+1}^N (Z_t - \hat{Z}_{t-1}(1))^2, \quad (38)$$

em que l é escolhido de tal modo que o valor inicial utilizado em (36) não influencie na previsão.

2.3.2 Suavização exponencial simples (SES)

O método de suavização exponencial simples considera que a série pode ser descrita pela expressão

$$\bar{Z}_t = \alpha \sum_{k=0}^{t-1} (1-\alpha)^k Z_{t-k} + (1-\alpha)^t \bar{Z}_0, t = 1, \dots, n, \quad (39)$$

onde \bar{Z}_t é denominado valor exponencialmente suavizado e α é a constante de suavização, $0 \leq \alpha \leq 1$. Este método pondera a média, atribuindo pesos maiores às observações mais recentes.

A previsão dos valores futuros é obtida pelo último valor exponencialmente suavizado conforme as expressões

$$\hat{Z}_t(h) = \bar{Z}_t, h = 1, \dots, n. \quad (40)$$

$$\hat{Z}_t(h) = \alpha Z_t + (1 - \alpha) \hat{Z}_{t-1}(h + 1), \quad (41)$$

o que pode ser interpretado como uma equação de atualização da previsão, quando adquirida uma nova observação. Assim, cada nova previsão pode ser obtida da anterior, adicionando um múltiplo do erro de previsão, garantido que as novas previsões acompanhem as mudanças de nível da série, relevadas pelo erro de previsão.

A constante de suavização α quanto menor for, mais estáveis serão as previsões finais, uma vez que o valor baixo implica que pesos maiores serão dados às observações passadas e, qualquer flutuação aleatória no presente, exercerá um peso menor no cálculo de previsão [Morettin e Toloi 2006]. Assim, para determiná-la é necessário adotar alguns critérios, tais como tipo de autocorrelação entre os dados e o custo de previsão. [Morettin e Toloi 2006] afirmam que um procedimento mais objetivo é selecionar o valor que fornece a melhor previsão das observações já obtidas, como foi definido no método MMS, Subseção 2.3.1, Eq. (38).

2.3.3 Suavização exponencial de Holt (SH)

O método SH é similar ao método SES; a diferença é que em vez de suavizar só o nível, ele utiliza uma nova constante de suavização para modelar a tendência da série.

Os valores do nível e da tendência da série, no instante t , serão inferidos

$$\bar{Z}_t = AZ_t(1 - A)(\bar{Z}_{t-1} + \hat{T}_{t-1}), 0 < A < 1 \text{ e } t = 2, \dots, n, \quad (42)$$

$$\hat{T}_t = C(\bar{Z}_t - \bar{Z}_{t-1}) + (1 - C)\hat{T}_{t-1}, 0 < C < 1 \text{ e } t = 2, \dots, n, \quad (43)$$

onde A e C são as constantes de suavização e \hat{T}_t representa o nível da série. A previsão para o valor Z_{t+h} , com origem em t é realizada adicionando-se ao valor básico \bar{Z}_t a tendência multiplicada pelo número de passos à frente que se deseja prever (h), isto é

$$\hat{Z}_t(h) = \bar{Z}_t + h\hat{T}_t, h = 1, \dots, n. \quad (44)$$

Assim, uma nova previsão para o valor Z_{t+h} é obtida como

$$\hat{Z}_{t+1}(h-1) = \bar{Z}_{t+1} + (h-1)\hat{T}_{t+1}. \quad (45)$$

Para utilizar as equações do método SEH (42-45), é preciso fazer hipóteses sobre os valores iniciais. O procedimento mais simples é atribuir $\hat{T}_2 = Z_2 - Z_1$ e $\bar{Z}_2 = Z_2$. Para determinar as constantes de suavização, segue o mesmo procedimento do método SES, só que em vez de escolher o valor de α que torna a soma dos erros quadráticos de previsão mínima, deve-se escolher o vetor (A, C) de forma que satisfaça esta condição.

2.3.4 Suavização exponencial de Holt-Winters (SHW)

A técnica de suavização exponencial de Hot-Winters é baseada em três equações com constantes de suavização diferentes, que são associadas a cada uma das componentes do padrão da série: nível, tendência e sazonalidade. Assim, esta técnica possui dois tipos de procedimentos, sendo o primeiro para série sazonal multiplicativa e o segundo para série sazonal aditiva.

Seja Z uma série temporal com período s , a variante mais usal do método, suavização de Holt-Winter Multiplicativo (SHWM), considera o fator sazonal F_t como sendo multiplicativo e a tendência aditiva conforme a expressão

$$Z_t = \mu_t F_t + T_t + a_t, t = 1, \dots, n. \quad (46)$$

As três equações de suavização para cada uma das componentes da série aparecem na forma

$$\hat{F}_t = D \left(\frac{Z_t}{\bar{Z}_t} \right) + (1-D)\hat{F}_{t-s}, 0 < D < 1, \quad (47)$$

$$\bar{Z}_t = A \left(\frac{Z_t}{\hat{F}_{t-s}} \right) + (1-A)(\bar{Z}_{t-1} + \hat{T}_{t-1}), 0 < A < 1, \quad (48)$$

$$\hat{T}_t = C(\bar{Z}_t - \bar{Z}_{t-1}) + (1-C)\hat{T}_{t-1}, 0 < C < 1, \quad (49)$$

onde $t = s+1, \dots, n$. Estas expressões representam estimativas do fator sazonal, do nível e da tendência, respectivamente. A , C e D são as constantes de suavização, que novamente são calculadas minimizando os erros de previsão a um passo a frente.

As previsões do procedimento para séries sazonal multiplicativa são obtidas como

$$\begin{aligned}\hat{Z}_t(h) &= (\bar{Z}_t + h\hat{T}_t)\hat{F}_{t+h-s}, h = 1, 2, \dots, s, \\ \hat{Z}_t(h) &= (\bar{Z}_t + h\hat{T}_t)\hat{F}_{t+h-2s}, h = s + 1, \dots, 2s,\end{aligned}\quad (50)$$

Para atualizar as previsões quando inserido uma nova observação Z_{t+1} , deve-se utilizar as seguintes equações

$$\hat{F}_{t+1} = D\left(\frac{Z_{t+1}}{\bar{Z}_{t+1}}\right) + (1 - D)\hat{F}_{t+1-s}, \quad (51)$$

$$\bar{Z}_{t+1} = A\left(\frac{Z_{t+1}}{\hat{F}_{t+1-s}}\right) + (1 - A)(\bar{Z}_t + \hat{T}_t), \quad (52)$$

$$\hat{T}_{t+1} = C(\bar{Z}_{t+1} - \bar{Z}_t) + (1 - C)\hat{T}_t. \quad (53)$$

Para série sazonal aditiva, a variante do método, suavização exponencial de Holt-Winter Aditivo (SHWA), considera o fator sazonal aditivo. O procedimento anterior pode ser modificado para tratar com situação onde o fator sazonal é aditivo,

$$Z_t = \mu_t + F_t + T_t + a_t. \quad (54)$$

As estimativas do fator sazonal, nível e tendência são obtidas por

$$\hat{F}_t = D(Z_t - \bar{Z}_t) + (1 - D)\hat{F}_{t-s}, 0 < D < 1, \quad (55)$$

$$\bar{Z}_t = A(Z_t - \hat{F}_{t-s}) + (1 - A)(\bar{Z}_{t-1} + \hat{T}_{t-1}), 0 < A < 1, \quad (56)$$

$$\hat{T}_t = C(\bar{Z}_t - \bar{Z}_{t-1}) + (1 - C)\hat{T}_{t-1}, 0 < C < 1, \quad (57)$$

respectivamente; A, C, D são as constantes de suavização. Para realizar previsão, a Eq. (50) é modificada para

$$\begin{aligned}\hat{Z}_t(h) &= (\bar{Z}_t + h\hat{T}_t) + \hat{F}_{t+h-s}, h = 1, 2, \dots, s, \\ \hat{Z}_t(h) &= (\bar{Z}_t + h\hat{T}_t) + \hat{F}_{t+h-2s}, h = s + 1, \dots, 2s,\end{aligned}\quad (58)$$

onde \bar{Z}_t, \hat{T}_t e \hat{F}_t são dados por (56), (57) e (55), respectivamente.

2.4 Redes neurais artificiais

As Redes Neurais Artificiais computacionais são inspiradas no modelo do sistema nervoso biológico. Sua estrutura é definida como um conjunto de unidades de processamento, o qual é caracterizado por neurônios artificiais, que são interligados entre si, criando um grande número de interconexões permitindo que a mesma possua a capacidade de aquisição e manutenção do conhecimento com base nas informações apresentada a ela [Silva et al. 2010].

Segundo [Silva et al. 2010], as características mais relevantes quando se trata de aplicações de redes neurais são:

- Adaptação por experiência: Há a adaptação dos parâmetros internos da rede, fazendo os ajustes dos mesmos sucessivas vezes, relacionados ao comportamento do processo, possibilitando a aquisição do conhecimento por experimentação;
- Capacidade de aprendizado: Através de algum procedimento de treinamento, a rede extrai o relacionamento existente entre as diversas variáveis que compõem a aplicação;
- Habilidade de generalização: Uma vez treinada a rede, ela é capaz de generalizar o conhecimento, isto é, estimar soluções que antes eram desconhecidas;
- Organização dos dados: Uma vez treinada a rede com um conjunto de informações, a mesma é capaz de realizar sua organização interna visando possibilitar o agrupamento de padrões que apresentam algum tipo de característica em comum;
- Tolerância a falhas: como a rede possui muitas interconexões entre os neurônios, a rede acaba tornando-se um sistema tolerante a falhas;
- Armazenamento distribuído: O conhecimento a respeito do comportamento de determinado processo de uma arquitetura neural é realizado de forma distribuída entre as diversas sinapses de seus neurônios;
- Facilidade de prototipagem: A maioria das arquiteturas de redes neurais podem ser facilmente implementadas, pois, após o processo de treinamento, os resultados são normalmente obtidos por algumas operações matemáticas elementares.

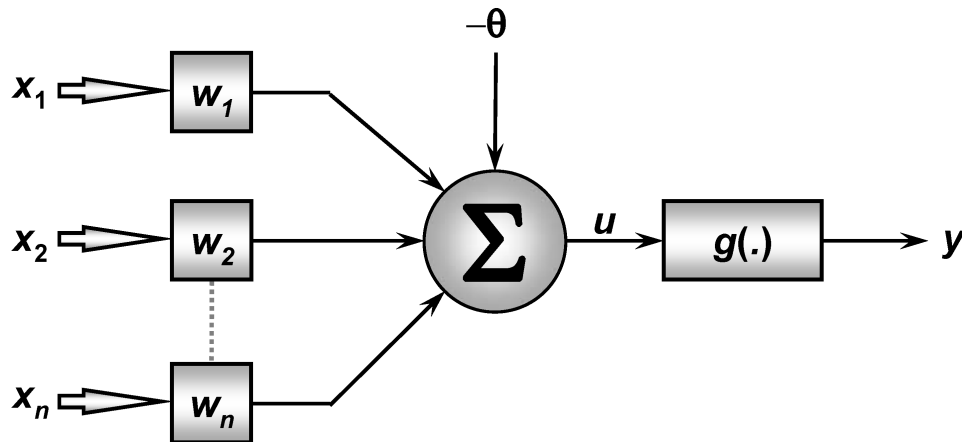
2.4.1 Neurônios artificiais

Como mencionado anteriormente, as redes neurais foram desenvolvidas a partir do modelo conhecido do sistema nervoso biológico. Assim, damos o nome de neurônios artificiais aos modelos simplificados dos neurônios biológicos, os quais representam os elementos computacionais ou unidades processadoras. Estes modelos foram inspirados

a partir da análise de geração e propagação dos impulsos elétricos pela membrana celular dos neurônios [Hodgkin e Huxley 1952].

A Figura 1 ilustra o modelo de neurônio mais simples que engloba estas características proposto por [McCulloch e Pitts 1943], este é o modelo mais utilizado nas diferentes arquiteturas de redes neurais artificiais.

Figura 1 – Neurônio artificial



Fonte: [Silva et al. 2010]

O neurônio artificial proposto por [McCulloch e Pitts 1943], é composto por sete elementos básicos:

- Sinais de entrada $\{x_1, x_2, \dots, x_n\}$

São os sinais de entrada da rede e que representam os valores assumidos pelas variáveis de alguma aplicação específica.

- Pesos sinápticos $\{w_1, w_2, \dots, w_n\}$

São os valores que servirão para ponderar cada uma das variáveis de entrada da rede.

- Combinador linear $\{\Sigma\}$

A função deste combinador é agregar todos os sinais de entrada que foram ponderados pelos respectivos pesos sinápticos a fim de produzir um valor de potencial de ativação

- Limiar de ativação $\{\theta\}$

Define o patamar apropriado para que o resultado produzido pelo combinador, possa gerar um valor de disparo em direção à saída do neurônio.

- Potencial de ativação $\{u\}$

Este é o valor atribuído pela diferença do valor produzido entre o combinador linear e o limiar de ativação. Se $u \geq \theta$, o neurônio produz um potencial excitatório; caso contrário, o potencial será inibitório.

- Função de ativação $\{g\}$

Tem como objetivo limitar a saída do neurônio, ou seja, produzir valores dentro do intervalo de valores razoáveis a serem assumidos por sua própria imagem funcional.

- Sinal de saída $\{y\}$

Consiste no valor final produzido pelo neurônio em relação a um determinado conjunto de entrada.

Na representação anterior, os diversos sinais de entrada advindo do meio externo são espelhados pelo conjunto $\{x_1, x_2, x_3, \dots, x_n\}$, que são similares aos impulsos elétricos externos captados pelos dendritos no neurônio biológico. Já as ponderações exercidas pelas junções sináptica do modelo biológico são representadas pelos pesos $\{w_1, w_2, \dots, w_2\}$.

Novamente, de forma semelhante, a relevância de cada entrada $\{x_i\}$ do neurônio é executada por meio de uma multiplicação pelo peso sináptico $\{w_i\}$, ponderando-se todas as informações externas que chegam ao neurônio. Desse modo, torna-se possível verificar que a saída do corpo celular artificial, representado por u , é a soma ponderada de suas entradas.

As expressões seguintes sintetizam os resultados gerados pelo neurônio artificial proposto por [McCulloch e Pitts 1943]

$$u = \sum_{i=1}^n w_i x_i - \theta, \quad (59)$$

$$y = g(u). \quad (60)$$

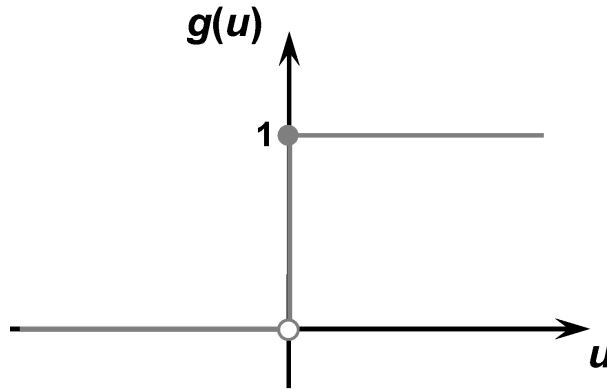
As funções de ativação são divididas em dois grupos principais, isto é, as funções parcialmente diferenciáveis e as funções totalmente diferenciáveis, considerando-se para tanto todo o domínio de definição das mesmas. As funções de ativação parcialmente diferenciáveis são aquelas que apresentam pontos cujas derivadas de primeira ordem são inexistentes. Entre as principais funções desse grupo temos a função degrau e a função degrau bipolar.

A função degrau assume como retorno valores unitários positivos quando o potencial de ativação do neurônio for maior ou igual a zero; caso contrário, retornará valores nulos conforme

$$g(n) = \begin{cases} 1, & \text{se } u \geq 0, \\ 0, & \text{se } u < 0. \end{cases} \quad (61)$$

A representação gráfica desta função é ilustrada na Figura 2.

Figura 2 – Função de ativação degrau



Fonte: [Silva et al. 2010]

A função degrau bipolar ou função sinal retornará valores unitários positivos quando o potencial de ativação do neurônio for maior que zero, valor nulo quando o potencial for também nulo e retornará valores unitários negativos quando o potencial for menor que zero. Sua expressão aparece como

$$g(n) = \begin{cases} 1, & \text{se } u > 0, \\ 0, & \text{se } u = 0, \\ -1, & \text{se } u < 0. \end{cases} \quad (62)$$

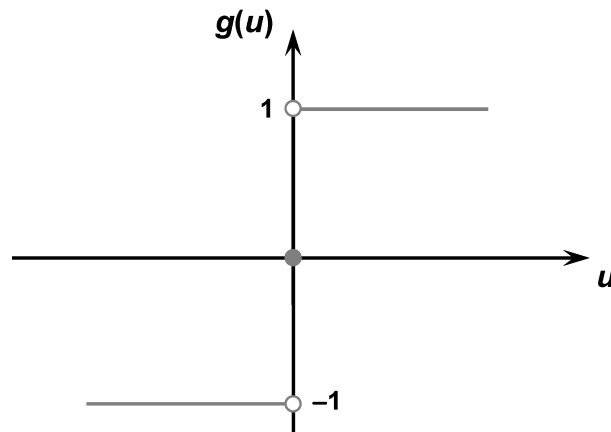
A Figura 3 ilustra a representação gráfica desta função.

Por outro lado, as funções de ativação totalmente diferenciáveis são aquelas cujas derivadas de primeira ordem existem e são conhecidas em todos os pontos de seu domínio de definição. Entre as principais funções deste tipo temos a função logística e a função linear.

A função logística assume sempre valores reais entre zero e um, tendo-se sua expressão matemática dada por

$$g(u) = \frac{1}{1 + e^{-\beta u}}, \quad (63)$$

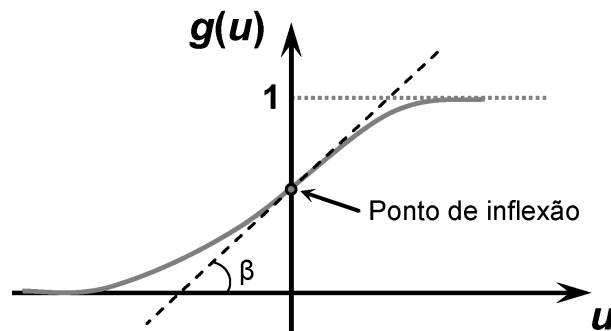
Figura 3 – Função de ativação degrau bipolar



Fonte: [Silva et al. 2010]

onde β é uma constante real associada ao nível de inclinação da função logística com seu ponto de inflexão. A Figura 4 ilustra a representação gráfica desta função.

Figura 4 – Função de ativação logística



Fonte: [Silva et al. 2010]

A função de ativação linear produz resultados de saída idênticos aos valores do potencial de ativação $\{u\}$, sendo expressa por

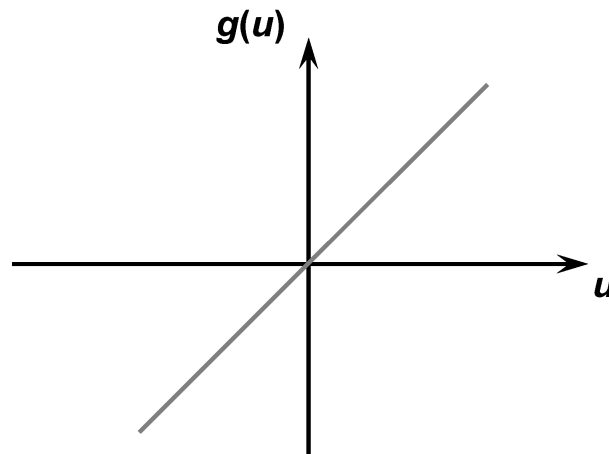
$$g(u) = u. \quad (64)$$

A Figura 5 ilustra a representação gráfica desta função.

2.4.2 Arquitetura de uma RNA

A forma como os neurônios são arranjados, ou dispostos, uns em relação aos outros define a arquitetura de uma rede neural artificial. Esses arranjos são organizados através do direcionamento das conexões sinápticas dos neurônios. Considerando uma determinada arquitetura, a topologia de uma rede neural artificial é definida como sendo as diferentes formas de composição na estrutura que poderá assumir.

Figura 5 – Função de ativação linear



Fonte: [Silva et al. 2010]

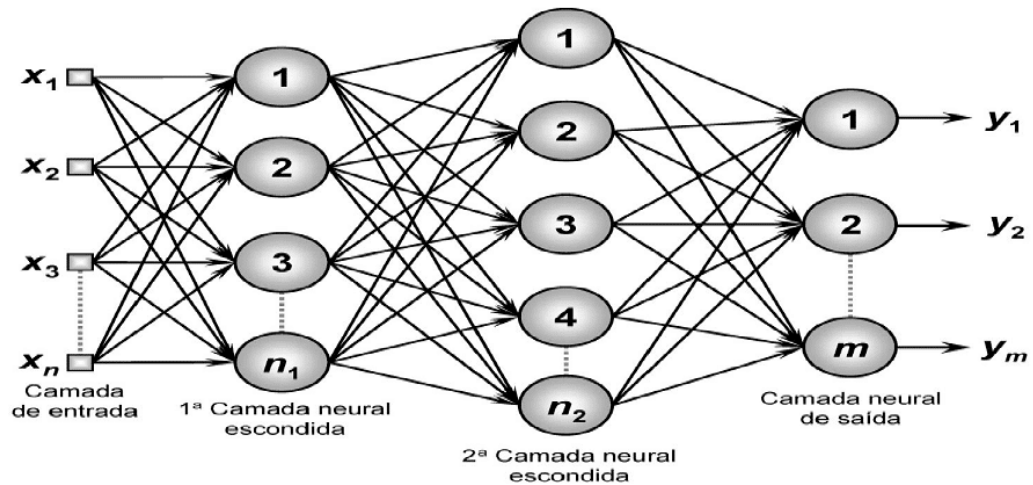
Segundo [Silva et al. 2010], uma rede neural artificial pode ser dividida em três partes, denominadas de camadas, as quais são nomeadas da seguinte forma:

- Camada de entrada: A camada de entrada é responsável pelo recebimento de informações, sinais, características ou medições advindas do meio externo, em que estas entradas são geralmente normalizadas em relação às faixas de variação dinâmicas produzidas pelas funções de ativação.
- Camadas escondidas, intermediárias, ocultas ou invisíveis: São as camadas compostas por neurônios que possuem a responsabilidade de extrair as características associadas ao processo.
- Camada de saída: Esta camada é responsável pela produção e apresentação dos resultados finais da rede, os quais são advindos dos processamentos efetuados pelos neurônios das camadas ocultas.

Considerando a disposição dos neurônios, as formas de interligação entre eles e a constituição de suas camadas, as principais arquiteturas de redes neurais artificiais podem ser divididas em: redes *feedforward* de camada simples, redes *feedforward* de camadas múltiplas, redes recorrentes e redes reticuladas. Para nosso propósito será apenas explicados as redes *feedforward* de camadas múltiplas. Para mais informações consultar [Silva et al. 2010].

As redes *feedforward* de camadas múltiplas são constituídas pela presença de uma ou mais camadas escondidas de neurônios, como pode ser visto na Figura 6.

Na Figura 6 temos uma rede *feedforward* de camadas múltiplas, formada por uma camada de entrada com n sinais, duas camadas escondidas constituídas respectivamente

Figura 6 – Exemplo de rede *feedforward* de camada múltiplas

Fonte: [Silva et al. 2010]

por n_1 e n_2 neurônios e, por último, uma camada neural de saída composta de m neurônios representando os valores de saídas da aplicação.

2.4.3 Processos de treinamento e aspectos de aprendizado

Uma das principais características das redes neurais artificiais está na capacidade de aprender a partir dos padrões extraídos das amostras de entrada. Após a rede ter aprendido o relacionamento entre as entradas e saídas, esta é capaz de generalizar soluções.

O processo de treinamento de uma rede neural consiste na aplicação de passos ordenados que são necessários para realizar os ajustes dos pesos sinápticos e limiares de seus neurônios, em que o principal objetivo é obter a generalização de soluções a serem produzidas pelas saídas da rede. Este conjunto de passos ordenados, visando o treinamento da rede, é denominado de algoritmo de aprendizagem.

Freqüentemente, o conjunto total das amostras é dividido em dois subconjuntos, os quais são denominados de subconjunto de treinamento e subconjunto de teste. O subconjunto de treinamento é composto aleatoriamente com cerca de 60% a 90% das amostras disponíveis. Já o subconjunto de teste está entre 10% a 40% do conjunto total de amostras. Ao longo do treinamento da rede neural artificial, cada apresentação completa das amostras pertencentes ao subconjunto de treinamento, será denominada de época de treinamento.

Segundo [Silva et al. 2010], o processo de treinamento pode ocorrer dos seguintes modos: treinamento supervisionado, treinamento não-supervisionado, treinamento com reforço, aprendizagem usando lote de padrões (off-line) e, aprendizagem usando

padrão-por-padrão (on-line). Neste trabalho será utilizado o aprendizado supervisionado, portanto, faremos apenas a descrição do mesmo. A descrição dos outros processos de aprendizagem podem ser consultados em [Silva et al. 2010].

O processo de treinamento supervisionado consiste em ter disponível, considerando cada amostra dos sinais de entrada, as respectivas saídas desejadas, ou seja, cada amostra de treinamento é composta pelos sinais de entrada e suas correspondentes saídas.

Os pesos sináptico e limiares são continuamente ajustados mediante a aplicação de ações comparativas, realizada pelo algoritmo de aprendizagem, que supervisiona a defasagem entre as respostas produzidas pela rede com relação as saídas desejadas, esta diferença é usada no procedimento de ajuste.

A rede é considerada treinada quando a defasagem estiver dentro de valores aceitáveis, levando em consideração os propósitos de generalização de soluções do sistema. A primeira estratégia de treinamento supervisionado foi proposta em 1949 por Donald Hebb, cuja inspiração adveio de observações neurofisiológicas [Hebb et al. 1949].

2.4.4 Redes perceptron multicamadas

As redes neurais do tipo *multi-layer perceptron* (MLP) são caracterizadas pela presença de pelo menos uma camada intermediária de neurônios. Isto é, as rede desse tipo possuem no mínimo duas camadas de neurônios, os quais estão distribuídos entre as camadas intermediária e a camada de saída.

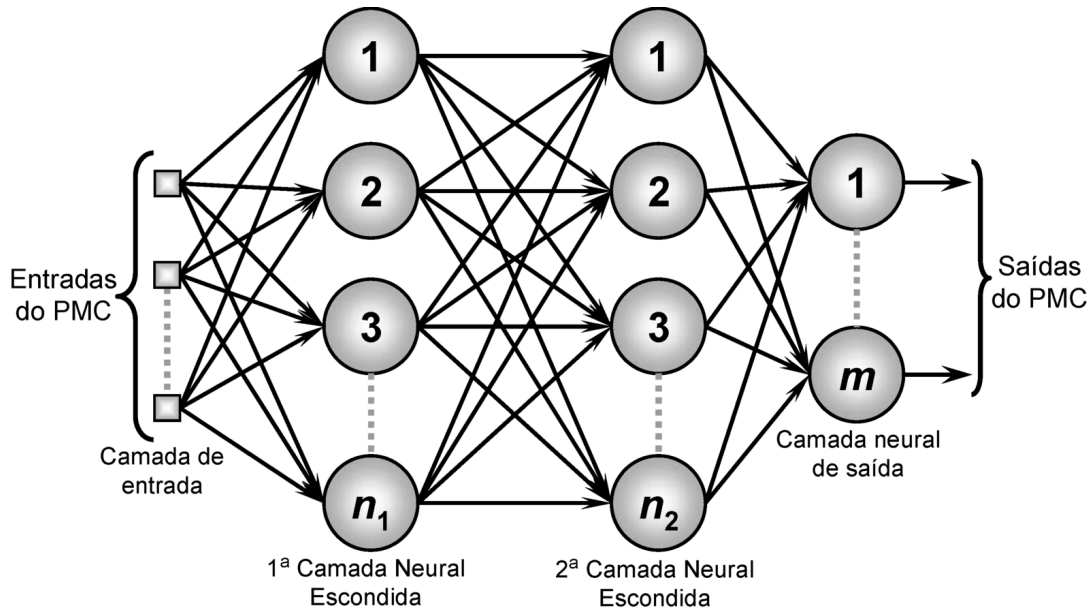
As MLP pertencem à arquitetura *feedforward* de múltiplas camadas, cujo treinamento é efetuado de forma supervisionada. A Figura 7 ilustra a arquitetura de uma MLP, onde o fluxo de informação na estrutura da rede se inicia na camada de entrada, percorrendo as camadas intermediária, sendo então finalizada na camada de saída.

Assim, analisando a Figura 7, observamos que cada uma das entradas da rede será propagado em direção à camada neural de saída do MLP. Com isso, as saídas dos neurônios da primeira camada neural escondida serão as próprias entradas daqueles neurônios pertencentes à segunda camada neural escondida.

Logo, a propagação dos sinais de entrada em uma rede MLP é sempre realizada num único sentido, ou seja, da camada de entrada em direção à camada neural de saída.

O ajuste dos pesos e do limiar de cada neurônio da rede MLP é efetuado através do processo de treinamento supervisionado. O algoritmo de aprendizado utilizado no processo de treinamento da MLP é denominado *backpropagation*.

Figura 7 – Rede perceptron multicamadas



Fonte: [Silva et al. 2010]

2.4.4.1 Processo de treinamento da MLP

O processo de treinamento utilizando o algoritmo *backpropagation* (regra Delta generalizado) é comumente realizado mediante as aplicações sucessivas de duas fases bem específicas. A primeira fase é denominada de “propagação adiante” (*forward*), na qual os sinais $\{x_1, x_2, \dots, x_n\}$ de uma amostra do conjunto de treinamento são inseridos na entrada da rede e são propagados camada a camada até a produção das respectivas saídas, visando somente obter as respostas da rede.

As respostas produzidas pelas saídas da rede são comparadas com as respectivas respostas desejadas, pois, trata-se de um processo de aprendizado supervisionado. Comparando as respostas da rede com as saídas desejadas, calcula-se o erro ou desvio, os quais serão subsequentemente utilizados para ajustar os pesos e limiares de todos os seus neurônios.

Com os valores dos erros, aplica-se em seguida, a segunda fase do método *backpropagation*, denominada de “propagação reversa” (*backward*). Os ajustes dos pesos sinápticos e limiares de todos os neurônios da rede são executados no decorrer desta fase. Assim, o desempenho da rede é medido por uma função erro que considera, para cada uma das x entrada, o quadrado da diferença entre o valor esperado e a respectiva saída calculada definido por

$$E = \sum_{i,x} E_i^x = \frac{1}{2} \sum_{i,x} (d_i^x - O_i^x)^2, \quad (65)$$

onde E_i^x é o erro no i -ésimo elemento neural, para x -ésima entrada, d_i^x é a saída desejada no i -ésimo elemento neural para a x -ésima entrada e O_i^x é a saída produzida por

$$O_i^x = g(u_i^x) = g\left(\sum_j w_{ij} v_j^x - \theta_i\right), \quad (66)$$

onde v_j^x é a j -ésima componente do conjunto de entrada $\{x_1, x_2, \dots, x_n\}$.

O algoritmo atua nos pesos sinápticos minimizando a função erro, por meio da técnica do gradiente descendente. Os valores dos pesos são modificados proporcionalmente ao oposto da derivada do erro, dado por

$$\Delta w_{ij} = \gamma \frac{\partial E_i^x}{\partial w_{ij}} \quad (67)$$

onde γ é a taxa de aprendizado que controla quão grande deve ser o "passo" a ser dado.

Definido-se $d_i^{x,k}$ como sendo a saída esperada na i -ésima unidade da k -ésima camada, quando a x -ésima entrada é apresentada a rede e

$$O_i^{x,k} = g(u_i^{x,k}), \quad (68)$$

como sendo a saída real da unidade, onde $u_i^{x,k} = \sum_j w_{ij}^k O_j^{x,k-1} - \theta_i$ e $O_i^{x,0} = v_i^x$, a função erro, na camada k , pode ser escrito como:

$$E^k = \frac{1}{2} \sum_{i,x} (d_i^{x,k} - O_i^{x,k})^2. \quad (69)$$

A correção nos passos na camada de saída k é dada através da aplicação da regra da cadeia dada por

$$\Delta w_{ij}^k = -\gamma \frac{\partial E^k}{\partial w_{ij}^k} = -\gamma \sum_x \frac{\partial E^k}{\partial O_i^{x,k}} \frac{\partial O_i^{x,k}}{\partial u_i^{x,k}} \frac{\partial u_i^{x,k}}{\partial w_{ij}^k} = \gamma \sum_x \delta_i^{x,k} O_j^{x,k-1}, \quad (70)$$

onde

$$\delta_i^{x,k} = g'(u_i^{x,k})(d_i^{x,k} - O_i^{x,k}). \quad (71)$$

De forma similar, pode ser mostrado que nas camadas intermediária a correção é dada por

$$\Delta w_{ij}^k = -\gamma \frac{\partial E^k}{\partial w_{ij}^k} = \gamma \sum_x \delta_i^{x,k} O_j^{x,k-1}, \quad (72)$$

definido-se

$$\delta_i^{x,k} = g'(u_i^{x,k}) \sum_j w_{ij}^{k+1} \delta_j^{x,k+1}. \quad (73)$$

2.5 Noções de lógica fuzzy

A lógica fuzzy, diferentemente da teoria clássica, toma as premissas e as conclusões como subjetivas, isto é, quanto maior for o grau de incerteza das premissas, maior será a incerteza das conclusões. Entretanto, a lógica fuzzy se destaca pelo seu caráter prático diferenciando, permitindo manipular incertezas podendo obter conclusões a partir de pressupostos incertos, o que não é possível na lógica clássica.

Em se tratando da lógica clássica, a sentença só pode assumir dois possíveis valores, isto é, a sentença é verdadeira ou falsa. Desse modo, apresentaremos a seguir os conectivos básicos da lógica clássica e a partir deles a extensão dessa teoria.

1. O conectivo **e**, representado por \wedge é definido através da seguinte função:

$$\wedge : \{0, 1\} \times \{0, 1\} \longrightarrow \{0, 1\}$$

$$(p, q) \longmapsto \wedge(p, q) = p \wedge q = \min\{p, q\}$$

2. O conectivo **ou**, representado por \vee é definido através da seguinte função:

$$\vee : \{0, 1\} \times \{0, 1\} \longrightarrow \{0, 1\}$$

$$(p, q) \longmapsto \vee(p, q) = p \vee q = \max\{p, q\}$$

3. O conectivo **negação**, representado por \neg é definido através da seguinte função:

$$\neg : \{0, 1\} \longrightarrow \{0, 1\}$$

$$p \longmapsto \neg p$$

4. O conectivo condicional, representado por \Rightarrow é definido através da seguinte função:

$$\Rightarrow : \{0, 1\} \times \{0, 1\} \longrightarrow \{0, 1\}$$

$$(p, q) \longmapsto \Rightarrow(p, q) = (p \Rightarrow q)$$

Tratando-se da lógica clássica, esses conectivos só assumem valores 0 ou 1. Olhando para os conjuntos fuzzy, estes conceitos são definidos através da t – norma e t – conorma. Estes operadores têm origem nos estudos de Espaços Métricos Estatísticos [Menger 1942]. Suas denominações advêm da generalização da propriedade triangular para tais espaços [Klement et al. 2013].

2.5.1 t-normas e t-conormas

Uma t-norma é uma operação binária utilizada para representar o operador e , ou a intersecção [Bedregal e Takahashi 2005]. Definimos a t-norma como uma função $\Delta: [0, 1] \times [0, 1] \longrightarrow [0, 1]$ que satisfaz:

1. Elemento neutro: $\Delta(1, x) = 1\Delta x = x$
2. Comutativa: $\Delta(x, y) = x\Delta y = y\Delta x = \Delta(y, x)$
3. Associativa: $x\Delta(y\Delta z) = (x\Delta y)\Delta z$
4. Monotonicidade: se $x \leq u$ e $y \leq v$ então $x\Delta y \leq u\Delta v$

Já a t-conorma é utilizada em lógica fuzzy para representar o operador clássico de união entre conjuntos e a operação de disjunção na lógica clássica [Bedregal e Takahashi 2005]. Definimos a t-conorma como uma função $\nabla: [0, 1] \times [0, 1] \longrightarrow [0, 1]$ que satisfaz:

1. Elemento neutro: $\nabla(0, x) = 0\nabla x = x$
2. Comutativa: $\nabla(x, y) = x\nabla y = y\nabla x = \nabla(y, x)$
3. Associativa: $x\nabla(y\nabla z) = (x\nabla y)\nabla z$
4. Monotonicidade: se $x \leq u$ e $y \leq v$ então $x\nabla y \leq u\nabla v$

A complementação interpreta o operador lógico de negação. Assim, uma função $\eta: [0, 1] \longrightarrow [0, 1]$ é uma negação fuzzy se satisfaz:

1. Fronteira: $\eta(0) = 1$ e $\eta(1) = 0$
2. Involução: $\eta(\eta(x)) = x$
3. Monotonicidade: η é decrescente

Já o operador de implicação é uma forma de modelar regras de inferência do tipo:

se <premissa> então <conclusão>

temos que o operador $\Rightarrow: [0, 1] \times [0, 1] \longrightarrow [0, 1]$ é uma implicação fuzzy se satisfaz:

1. Reproduz a tabela da implicação clássica.
2. É decrescente na primeira variável, isto é, para cada $x \in [0, 1]$ tem-se $(a \Rightarrow x) \leq (b \Rightarrow x)$ se $a \geq b$
3. É crescente na segunda variável, ou seja, para cada $x \in [0, 1]$ tem-se $(x \Rightarrow a) \geq (x \Rightarrow b)$ se $a \geq b$

2.5.2 Raciocínio aproximado e variáveis linguísticas

O raciocínio aproximado consiste no processo de obter conclusões a partir de premissas incertas. Se a incerteza for considerada fuzzy, é frequente o uso do termo raciocínio fuzzy.

Na vida diária, é comum tirarmos conclusões a partir destas premissas, como por exemplo:

"Se a banana está amarela então a banana está madura"

ou seja,

"Caso se tenha uma banana amarela, é automaticamente classificada como madura".

De forma mais geral, podemos ter algo como:

"Se X é Q então Y é W."

Neste caso, conhecendo X e Q, conclui-se que Y é W. Esta é uma generalização do conhecimento método dedutivo de "modus poneus" [Arnold 2013], isto é, utiliza a subjetividade dos predicados envolvidos no *modus poneus* clássico.

[Arnold 2013] define uma variável linguística X no universo U como sendo uma variável cujos valores assumidos são subconjuntos fuzzy de U. As variáveis linguísticas são definidas em termos de uma base de variáveis, sendo esta uma variável no sentido clássico exemplificada por qualquer variável física (temperatura, pressão, velocidade, ...), assim como, pode também ser exemplificada por qualquer variável numérica (idade, salário, probabilidade, ...).

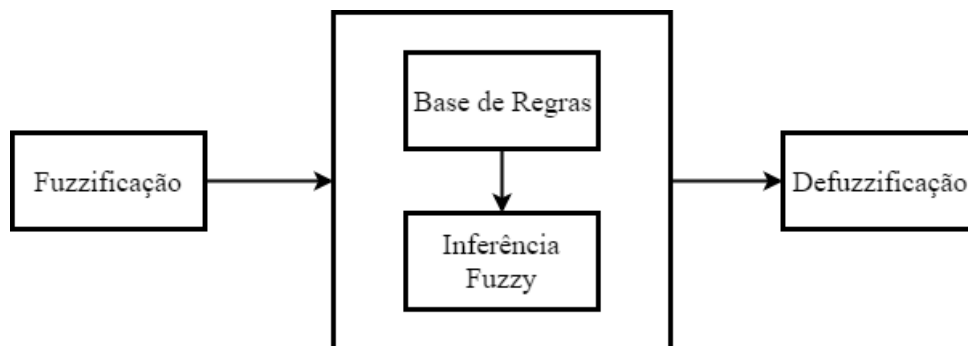
De maneira geral, uma variável linguística é um substantivo, enquanto que seus valores são adjetivos, representados por conjuntos *fuzzy*. Por exemplo, “dor” é uma variável linguística que pode assumir os atributos “forte” ou “fraca” que são modelados por conjuntos fuzzy. Consideraremos os casos em que as variáveis assumirão valores de números *fuzzy* com o universo sendo os números reais e X sendo uma variável linguística real, em que as sentenças que possuem variáveis linguísticas juntamente com seus atributos são chamadas de proposições *fuzzy*.

2.5.3 Controladores fuzzy

Os controladores *fuzzy* propõem um cálculo matemático para traduzir e processar o conhecimento subjetivo obtido de um especialista e ou uma base de dados, em que este conhecimento é processado com um determinado nível de incerteza [TEIXEIRA et al. 2007], as informações são convertidas para variáveis linguísticas. Tais informações na maioria das vezes são difíceis de serem expressas em termos precisos, tornando o uso de termos linguísticos necessários. Estes termos linguísticos são constituídos por um conjunto de regras que são decodificadas por um controlador.

Seja um sistema *fuzzy* qualquer, cada entrada *fuzzy* faz-se corresponder uma saída *fuzzy*. Os controladores *fuzzy* também possuem essa característica, mas se a entrada for crisp, espera-se também que a saída seja crisp. Um controlador *fuzzy* é constituído de 4 etapas conforme apresentado na Figura 8 composto pelo processo de fuzzificação, base de regras, inferência *fuzzy* e defuzzificação.

Figura 8 – Esquema geral de um controlador fuzzy



Fonte: [TEIXEIRA et al. 2007]

- **Fuzzificação:** Consiste em modelarmos matematicamente a informação das variáveis de entrada por meio de conjuntos *fuzzy*. Para cada variável de entrada deve ser atribuída um termo linguístico que representa o estado desta variável e, a cada termo linguístico, deve ser associado um conjunto *fuzzy* por uma função de pertinência. Nesta etapa se faz necessário a presença de um especialista no processo que está sendo estudado, para que essa modelagem possa ser feita.

- **Base de Regras:** É composto por proposições *fuzzy* e consiste em catalogarmos as variáveis e suas classificações linguísticas. Esta etapa é considerada o núcleo do sistema *fuzzy*.
- **Inferência Fuzzy:** É onde se definem quais são os conectivos lógicos usados para estabelecer a relação fuzzy que modela a base de regras. É desde módulo que depende o sucesso do sistema *fuzzy* já que ele fornecerá a saída *fuzzy* a ser adotada pelo controlado a partir de cada entrada *fuzzy*.
- **Defuzzificação:** Como no geral um sistema *fuzzy* mapeia cada entrada *fuzzy* a uma saída *fuzzy*, espera-se que a cada entrada crisp faça corresponder a uma saída crisp, ou até mesmo uma entrada *fuzzy* corresponder a uma saída crisp. É nessa etapa que o conjunto *fuzzy* de saída é transformado em um valor crisp, ou seja, um número real. Para realizar a defuzzificação, o método mais comum é o centro de gravidade (centróide). Assim, quando o domínio é discreto, o centro de gravidade $G(B)$ é dado por

$$G(B) = \frac{\sum_{i=0}^n u_i \varphi_B(u_i)}{\sum_{i=0}^n \varphi_B(u_i)}, \quad (74)$$

em que $\varphi_B(u_i)$ corresponde aos valores que indicam o grau de compatibilidade do valor u_i com o conceito modelado pelo conjunto *fuzzy* B . Já quando o domínio é contínuo, $G(B)$ é dado por

$$G(B) = \frac{\int_{\mathbb{R}} u \varphi_B(u) du}{\int_{\mathbb{R}} \varphi_B(u) du}. \quad (75)$$

Para mais informações sobre outros métodos de defuzzificação, consultar [[Pedrycz e Gomide 1998](#)].

2.5.4 Rede neurofuzzy

Uma rede *neurofuzzy* é um tipo de sistema *fuzzy* onde é aplicado algum algoritmo de treinamento de aprendizado de máquina, sobre o mesmo. A união de sistemas *fuzzy* com algum algoritmo de treinamento de aprendizado de máquina resulta em um sistema híbrido denominado de rede *neurofuzzy* [[Oliveira et al. 2007](#)]. Essa abordagem é considerada uma ferramenta poderosa no desenvolvimento de sistemas *fuzzy*, sendo sua utilidade evidenciada quando surge a necessidade de lidar com sistemas fortemente não-lineares, de comportamento variável no tempo, como no caso de séries temporais.

A ideia básica é simples: os modelos *neurofuzzy* nada mais são que representações de sistemas *fuzzy* (formulados por meio de regras *fuzzy* do tipo se-então) na forma de redes (ou grafos direcionados) passíveis de treinamento por técnicas similares as usadas

em redes neurais artificiais. Tal estrutura viabiliza a adaptação do sistema *fuzzy* sob treinamento ao conjunto de treinamento, de modo que, ao final, obtenhamos resultados interpretáveis, ao contrário dos casos típicos das redes neurais artificiais, de difícil interpretação por humanos.

O que torna possível essa integração é o fato de as redes neurais e os sistemas *fuzzy* serem aproximadores universais. Logo, a rede *neurofuzzy* mantém essa característica fundamental no desenvolvimento das aplicações. Outra característica importante é que, tanto as redes neurais, como os sistemas *fuzzy*, manipulam dados incompletos, imprecisos, complexos e não-lineares, além do alto grau de abstração de informações irrelevante para o sistema [Campos 2008].

Assim, as redes *neurofuzzy* herdam da rede neural a capacidade de aprendizagem, generalização, classificação, robustez, adaptações e agrupamento de dados comuns em *clusters*. As redes *neurofuzzy* permitem que mantenha uma certa relação de suas entradas e suas respectivas saídas, porque o comportamento deste modelo pode ser entendido através das observações das variáveis linguísticas, das funções de pertinência e das próprias regras *fuzzy*, as quais explicam facilmente o funcionamento do sistema, devido a simplicidade e proximidade com a linguagem humana.

2.5.4.1 Arquitetura ANFIS

A arquitetura ANFIS (*adaptive-network-based fuzzy inference system*) foi proposta por [Jang 1993], e é um dos sistemas *neurofuzzy* mais conhecido e empregado devido as suas qualidades de bom classificador e previsor [Campos 2008]. Este sistema processa as variáveis linguísticas com a capacidade de aprendizado de uma rede neural, fazendo com que o mesmo se ajuste e otimize uma solução qualquer para o ambiente no qual foi inserido.

Esta arquitetura tem por única limitação a configuração da rede que é limitada a rede do tipo *feedforward* [Jang et al. 1997]. Entretanto, as redes adaptativa podem ser empregadas em uma ampla variedade de aplicações, como na tomada de decisão, no processamento de sinais e em controle. A seguir será descrita a arquitetura do sistema de inferência ANFIS.

Para simplificar o entendimento, será considerado um sistema de inferência *fuzzy* com duas entradas x e y e uma saída z . Para o modelo *fuzzy* Sugeno de primeira ordem, um conjunto usual de regras “se-então” é o seguinte:

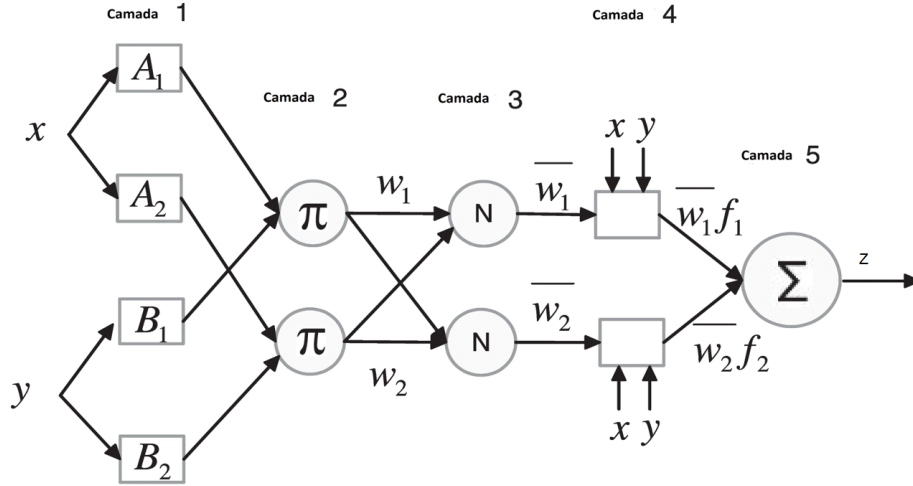
Regra 1: Se x é A_1 e y é B_1 então $f_1(x, y) = p_1x + q_1y + r$,

Regra 2: Se x é A_2 e y é B_2 então $f_2(x, y) = p_2x + q_2y + r$.

A Figura 14 ilustra a mecânica do raciocínio para o modelo *fuzzy* Sugeno de primeira ordem com duas entradas e duas regras. A Figura 10 ilustra a arquitetura

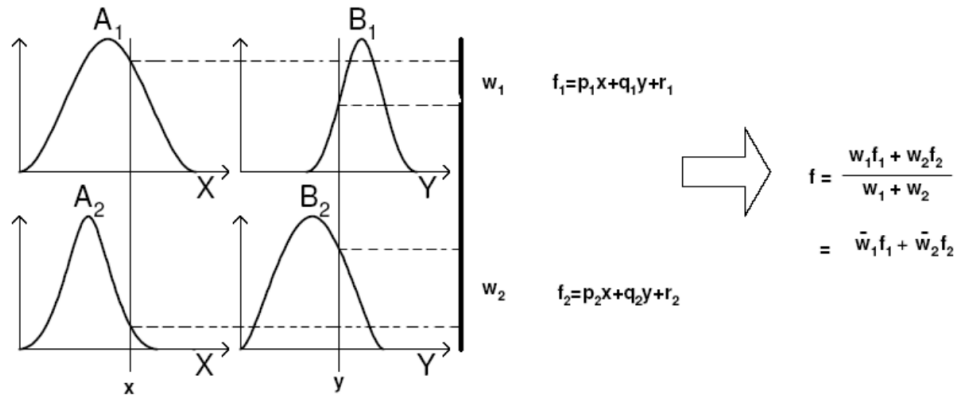
ANFIS equivalente, onde os nós da mesma camada têm funções similares. A seguir, será descrita cada camada do modelo.

Figura 9 – Modelo ANFIS com duas entradas e uma saída.



Fonte: [Jang 1993]

Figura 10 – Modelo fuzzy Sugeno com duas entradas e duas regras (Jang (1993)).



Fonte: [Jang 1993]

- Camada 1: Esta camada é responsável por transformar cada variável de entrada em variáveis do universo *fuzzy*. Cada neurônio desta camada representa uma função de pertinência das diversas variáveis *fuzzy* dada por

$$\begin{aligned} \mu_{A_i}(x), \text{ para } i = 1, 2 \text{ ou} \\ \mu_{B_{i-2}}(y), \text{ para } i = 3, 4, \end{aligned} \quad (76)$$

sendo x e y as entradas para os nós dos dois neurônios da camada inicial.

- Camada 2: Esta camada realiza a aplicação da t-norma às saídas das camadas anteriores, logo, é nesta camada que se encontram as regras de conhecimento do

especialista *fuzzy*. Isso equivale à aplicação das operações se-então, cuja saída é o produto de todos os sinais da entrada da camada anterior na forma

$$w_i = \mu_{A_i}(x) \cdot \mu_{B_i}(y), \quad i = 1, 2. \quad (77)$$

- Camada 3: É responsável pelo cálculo do grau de ativação normalizado das variáveis, dos valores gerados pelas regras se-então, evitando assim uma distorção nos parâmetros de consequência do sistema. A saída desse nó é dada pela razão entre o *i*-ésimo nível de disparo da regra e a soma de todas os níveis de disparo, isto é

$$\bar{w}_i = \frac{w_i}{w_1 + w_2} \quad i = 1, 2. \quad (78)$$

- Camada 4: Cada nó desta camada é adaptativo por conter uma função adaptativa para realizar o cálculo com os parâmetros de consequência do sistema dada por

$$\bar{w}_i f_i = \bar{w}_i (p_i x + q_i y + r_i), \quad (79)$$

onde \bar{w}_i é o nível de disparo normalizado da camada 3 e p_i , q_i e r_i é o conjunto de parâmetros deste nó.

- Camada 5: O nó desta última camada da arquitetura, calcula a saída crisp do sistema e juntamente com os nós das camadas 3 e 4, promovem a defuzzificação ou soma total de todos os sinais de entrada, ou seja, transforma as variáveis *fuzzy* de consequência obtidas na camada 4 em um único valor dos conjuntos clássicos, usando a expressão

$$\sum_i \bar{w}_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i}. \quad (80)$$

2.6 Estado da arte

Vários esforços foram e estão sendo aplicados no objetivo de aprimorar métodos de previsão e, consequentemente fazer previsão acurada de séries temporais. Grande parte desses esforços são focados na aplicação de métodos estatísticos. Mas, nos últimos anos, vários pesquisadores apontaram abordagens como redes neurais artificiais e lógica *fuzzy* como uma alternativa bastante satisfatória na abordagem de previsão de séries temporais.

Os métodos mais comum no uso de análise de séries temporais são os métodos estatísticos. Entre estes métodos, os mais aplicados são os métodos de suavização exponencial e os modelos ARIMA. [BARBOSA et al. 2015] propõe aplicar a metodologia

Box-Jenkins na obtenção de um modelo que permitisse prever os valores futuros de temperatura média mensal da cidade de Bauru (SP). O estudo mostrou que a metodologia *Box-Jenkins* conseguiu modelar adequadamente os dados climáticos, realizando previsões bem próximas dos valores reais.

Outro estudo feito por [Azevedo et al. 2017], verifica se a metodologia ARIMA prevê com exatidão as arrecadações de ICMS. Os resultados mostraram que a metodologia de previsão ARIMA pode ser utilizada para aumentar a previsibilidade da arrecadação de ICMS. Já [Scheffer et al. 2014] construíram um modelo para previsão do ICMS (Imposto sobre Circulação de Mercadorias e Serviços) no Estado da Paraíba. De acordo com os autores, a metodologia de séries temporais pode melhorar a previsão, sugerindo um modelo de suavização exponencial de Holt-Winters aditivo.

O uso de métodos heurísticos e modelos econométricos lineares foram investigados por [Miranda et al. 2013], que utilizou uma rede neural *multi-layer perceptron*, treinada com o algoritmo de retropropagação de erro como método eurístico. Quanto aos modelos lineares, utilizaram-se os Autoregressivos, o de Médias Móveis e os Autoregressivos de Média Móveis. Foram realizados testes entre os parâmetros modificáveis destes modelos e experimentos a fim de determinar o melhor ajuste dos parâmetros de suavização nos modelos de média móvel, ajustes da linha de tendência nos modelos de projeção, a melhor ordem p , q no modelo ARIMA, ou o menor erro na modificação da taxa de aprendizado, topologia, erro desejado e ciclo da rede neural. Por fim, os resultados mostram que os dois modelos tiveram um desempenho satisfatório, mas os critérios de avaliação adotados pelos autores, demonstraram que a rede neural possui um poder maior na predição.

[Medeiros et al. 2016] aplicaram métodos de RNA e lógica *fuzzy* na previsão de demanda de sistemas de energia na subestação de Itaporanga (PB). Os autores desenvolveram dois sistemas de previsão, sendo o primeiro por uma RNA e o segundo por sistema de inferência *fuzzy*. Os dois sistemas alcançaram resultados significativos com valores de MAPE (*Mean Absolute Percentage Error*) abaixo de 4%. O sistema de inferência *fuzzy* apresentou melhor resultado do que a rede neural multicamada, entretanto, os autores afirmam que ambos foram bastante equiparados. Assim, os autores concluíram que a lógica *fuzzy* pode ser bastante eficiente para realizar previsão.

Há uma grande necessidade em realizar previsões acuradas. O aprimoramento tecnológico, tem estimulado a aplicação e comparação de diferentes técnicas de modelagem, assim como métodos de combinação. Historicamente, pesquisadores consideram que combinar diferentes técnicas de previsão melhora a sua acurácia. [Martins e Werner 2014] procuram verificar se há diferença entre a acurácia das previsões individuais e a de suas combinações por meio da combinação da metodologia *Box-Jenkins* e a RNA. Como resultado, foi observado por meio das medidas de acurácia que as previsões

obtidas pela combinação apresentaram desempenho superior em relação às demais previsões produzidas pelos métodos separadamente.

Outro estudo que discute a combinação de *Box-Jenkins* e RNA foi feito [ABREU et al. 2012]. Neste trabalho, é proposta uma modelagem híbrida tendo como objetivo investigar a precisão dos resultados obtidos na previsão a curto prazo. O modelo proposto utiliza os modelos ARIMA para fornecer a série temporal simulada do modelo real, que foi utilizada como entradas da RNA. Neste estudo, são apresentados resultados para uma série temporal de demanda de energia elétrica. O estudo também mostrou que utilizando esta abordagem, quando maior for a série temporal, melhor será a previsão. [Sousa et al. 2012] realizaram também um estudo comparativo entre os métodos de inteligência artificial e estatística para o problema de previsão. Foram comparados a metodologia *Box-Jenkins* com uma rede neural de *Elman* na previsão um passo a frente dos preços do etanol no estado de Goiás e comparados através medidas de erros específicas. Os resultados mostram que o modelo estatístico de *Box-Jenkins* se mostrou mais adequado que a rede *Elman* em termos de parcimônia entre desempenho e complexidade.

3 Metodologia

Tendo como objetivo geral desta dissertação a análise e previsão de séries temporais de dados obtidas do projeto de acompanhamento do custo da cesta básica da microrregião Ilhéus-Itabuna, foram modelados vários métodos estatísticos e de aprendizado de máquina para esta finalidade. Existem na literatura vários métodos estatísticos e de aprendizado de máquina que são empregados na análise e previsão de séries temporais, onde cada um se comporta de maneira diferente ao lidar com determinadas séries temporais.

Como estamos lidando com variáveis que variam ao longo do tempo, no caso o preço da cesta básica, escolhemos métodos estatísticos que conseguem modelar a série assumindo uma certa aleatoriedade ao longo do tempo. Fazendo uma busca na literatura, encontramos os métodos de suavização, que por sua vez, assumem que os valores extremos da série se comportam de forma aleatória e, assim por meio da suavização desses extremos identifica o padrão básico distinguindo qualquer ruído que possa estar contido nas observações [Morettin e Toloí 2006], tornando-os uma abordagem viável para o nosso problema.

Além dos métodos de suavização, escolhemos a abordagem *Box-Jenkins* que consiste em fazer o ajuste de um modelo autorregressivo integrado de médias móveis, a um conjunto de dados. A escolha deste modelo se deu exatamente pela propriedade de considerar uma certa aleatoriedade presente na série, além da capacidade de conseguir modelar de forma eficiente séries temporais não estacionárias.

Os métodos estatísticos como mencionados nos parágrafos anteriores, conseguem modelar de forma satisfatória séries temporais considerando-as a presença de variáveis aleatórias. Estes modelos conseguem captar o comportamento linear da série, entretanto, não conseguem captar o comportamento não-linear. Visando contornar essa deficiência dos métodos estatísticos, encontramos na literatura abordagens usando o aprendizado de máquina que foram aplicados de forma satisfatória na análise e previsão de séries temporais como apresentado na Seção 2.6.

Em particular, a previsão de séries temporais consiste em obter valores futuros a partir do conhecimento de seus valores passados. Dentre as técnicas de aprendizado de máquina, estão aquelas baseadas em diferentes arquiteturas de RNA, do tipo *feedforward* como o tipo MLP. Este tipo de rede se destaca pela capacidade de lidar melhor com a não linearidade dos dados, sendo indicados para analisar sistemas dinâmicos como séries temporais não estacionárias [Müller et al. 2012]. Por estes motivos escolhemos como modelo de aprendizado de máquina a RNA do tipo MLP para realizar a modelagem da

série objeto de estudo.

Tanto as abordagens estatísticas como as de aprendizado de máquina abstraem características da série, as quais são usadas para realizar as previsões. Ao combinar abordagens diferentes, temos como resultado um modelo final com a absorção das características individuais de cada um, o que possibilita previsões mais acuradas. Baseado nessa ideia, realizamos duas abordagens híbridas, onde a primeira abordagem consistiu na combinação do modelo ARIMA com RNA e a segunda abordagem consistiu na combinação de uma RNA com sistema de inferência *fuzzy*.

A estrutura deste capítulo está assim organizada: a primeira seção apresenta uma descrição dos dados a serem analisados, a segunda seção mostra como foram realizadas as análise da série, a terceira seção fornece a descrição computacional das técnicas e como foram obtidos os ajustamentos dos modelos e a quarta e última seção como foram obtidas as previsões e as métricas de avaliação.

3.1 Dados

O preço total do custo da cesta básica de alimentos da cidade de Ilhéus, foi coletado pelo projeto Acompanhamento do Custo da Cesta Básica (ACCB) na microrregião Ilhéus-Itabuna de iniciativa do departamento de Ciência Econômicas da Universidade Estadual de Santa Cruz (UESC). O ACCB é um projeto de extensão que desde 1999 vem realizando o levantamento mensal do custo da cesta básica de alimentos, assim como o preço individual de cada produto que compõe a cesta básica.

O objetivo deste projeto de extensão é acompanhar a evolução do preço da cesta básica de alimentos na microrregião Ilhéus-Itabuna, tabular os dados e os submeter a diversas instâncias governamentais. Para isto, a metodologia adotada é baseada na metodologia DIEESE [LLC 1955]. O DIEESE realiza a pesquisa da Cesta Básica de Alimentos (Ração Essencial Mínima) em dezoito capitais do Brasil acompanhando mensalmente a evolução de preços de treze produtos de alimentação, assim como o gasto mensal que o trabalhador teria para comprá-los.

A metodologia utilizada pelo DIEESE para a pesquisa da cesta básica de alimentos é estabelecida com base no Decreto Lei nº 399, que regulamenta o salário mínimo no Brasil. O Decreto Lei nº 399 regulamenta a Lei nº 185 de 14 de Janeiro de 1936 [Cassar 2011], que estabelece o salário mínimo como sendo a remuneração devida ao trabalhador adulto, sem distinção de sexo, por dia normal de serviço, capaz de satisfazer, em determinada época e região do país, às suas necessidades normais de alimentação, habitação, vestuário, higiene e transporte [Brasil 1938].

O estudo censitário realizado em cada localidade, juntamente com as informa-

ções salariais obtidas juntos às empresas das várias regiões, as Comissões do Salário Mínimo [Mattoso 1988] estabeleceram os valores mínimos regionais a serem pagos aos trabalhadores. O Decreto Lei nº 399 apresenta também uma lista de alimentos, com suas respectivas quantidades. Esta cesta, chamada de Cesta Básica de Alimentos, seria o suficiente para o sustento e bem estar de um trabalhador com idade adulta. Os bens e as quantidades estipuladas são diferenciados por região, como mostra a Tabela 1.

Tabela 1 – Provisões mínimas de alimentos estipuladas pelo Decreto Lei n 399

Alimentos	Região 1	Região 2	Região 3	Nacional
Carne	6,0 kg	4,5 kg	6,6 kg	6,0 kg
Leite	7,5 l	6,0 l	7,5 l	15,0 l
Feijão	4,5 kg	4,5 kg	4,5 kg	4,5 kg
Arroz	3,0 kg	3,6 kg	3,0 kg	3,0 kg
Farinha	1,5 kg	3,0 kg	1,5 kg	1,5 kg
Batata	6,0 kg		6,0 kg	6,0 kg
Legumes (Tomate)	9,0 kg	12,0 kg	9,0 kg	9,0 kg
Pão francês	6,0 kg	6,0 kg	6,0 kg	6,0 kg
Café em pó	600 gr	300 gr	600 gr	600 gr
Frutas (Banana)	90 unid	90 unid	90 unid	90 unid
Açúcar	3,0 kg	3,0 kg	3,0 kg	3,0 kg
Banha/Óleo	750 gr	750 gr	900 gr	1,5 kg
Manteiga	750 gr	750 gr	750 gr	900 gr

A região 1 é composta pelos Estados de São Paulo, Minas Gerais, Espírito Santo, Rio de Janeiro, Goiás e Distrito Federal. A região 2 é composta pelos Estados de Pernambuco, Bahia, Ceará, Rio Grande do Norte, Alagoas, Sergipe, Amazonas, Pará, Piauí, Tocantins, Acre, Paraíba, Rondônia, Amapá, Roraima e Maranhão. Já a região 3 é composta pelos Estados do Paraná, Santa Catarina, Rio Grande do Sul, Mato Grosso e Mato Grosso do Sul.

Em síntese, o Decreto Lei nº 399 determina que a parcela do salário mínimo correspondentes aos gastos com alimentação não pode ter valor inferior ao custo da Cesta Básica de Alimentos. No decorrer do tempo, a partir da criação dos Escritórios Regionais do DIEESE, foi implementado o acompanhamento da cesta em várias capitais do Brasil. Hoje, esta coleta é feita em dezoito capitais, permitindo assim acompanhar, a variação dos preços de cada produto, o custo de cada um deles e quantas horas um indivíduo que ganha um salário mínimo precisa trabalhar para poder adquiri-los mensalmente.

Assim, o ACCB realiza a pesquisa referente a Cesta Básica de Alimentos definida no Decreto Lei nº 399, tanto em relação aos produtos a serem pesquisados, quanto suas respectivas quantidades. O levantamento dos preços leva em conta a região 2 definida na Tabela 1, da qual o Estado da Bahia faz parte. A coleta é feita nos estabelecimentos de maior fluxo de compra das cidades de Ilhéus e Itabuna.

Para calcular o preço total da cesta básica de alimentos, somam-se os vários resultados coletados nos estabelecimentos comerciais e obtém-se, assim, o preço médio para cada produto. Este preço multiplicado pelas quantidades determina o gasto mensal do trabalhador com cada produto, cuja soma resulta no custo mensal da Cesta Básica de Alimentos para um indivíduo adulto.

Em síntese, o projeto ACCB realiza o levantamento, analisa os dados coletados e os submetem a diversas instâncias governamentais. Desde o seu início, os levantamentos são armazenados em uma base de dados, tendo assim uma série temporal. Entretanto, de março de 1999 a dezembro de 2004 a base de dados está fragmentada, pois alguns meses não foram devidamente registrados. Sendo assim, os dados coletados do ano 1999 até 2004 foram desconsiderados justamente por apresentar esta fragmentação, o que poderia influenciar no resultado final de análise da série, resultando em previsões pouco acuradas, por exemplo.

Já a partir de janeiro de 2005 a base de dados apresenta consistência, ou seja, depois de 2005 a base dados não apresenta fragmentação, contendo todos os meses de coleta referente aos produtos pesquisados e o custo total da cesta básica de alimentos para as duas cidades consideradas (Ilhéus e Itabuna). A base de dados possui no total 14 séries temporais, em que, 12 séries correspondem ao preço individual de cada produto e 2 séries referentes ao custo total da cesta básica de alimentos nas cidades de Ilhéus e Itabuna.

Como o objetivo deste projeto é fazer a análise de séries do tipo socioeconômico, foi escolhida a série, dentre as disponíveis na base de dados, para a qual a análise pudesse gerar um maior impacto socioeconômico. Levando isso em conta, a análise do custo total da cesta básica geraria um impacto maior, uma vez que a mesma é composta pela ração mínima necessária para a sobrevivência de um adulto durante um mês, com a composição de itens que não poderia faltar para suprir suas necessidades básicas.

Na base de dados há duas séries temporais referente ao custo total da cesta básica de alimentos, a primeira consiste na pesquisa nos estabelecimentos na cidade de Ilhéus e a segunda referente a pesquisa nos estabelecimentos na cidade de Itabuna. Dentre essas duas, foi escolhida para análise a série temporal da cidade de Ilhéus, isto porque, ao realizar a análise gráfica, a mesma apresentou um comportamento mais complexo do que a da cidade de Itabuna, embora a variação nos pontos da série seja pequena.

3.2 Análise da série temporal

A primeira parte deste projeto foi realizar a identificação das componentes presentes na série. Primeiro, verificamos a necessidade de transformar os dados da série usando uma transformação não-linear logarítmica com o objetivo de estabilizar a vari-

ância e tornar o efeito sazonal aditivo. Esta verificação foi realizada analisando o gráfico que traz no eixo das abscissas a média de subconjuntos de observações da série original e no eixo das ordenadas a amplitude de cada um desses subconjuntos.

A Segunda parte deste projeto foi realizar a decomposição da série temporal em suas componentes não-observáveis, são elas: sazonalidade e tendência. Para realizar a identificação da componente tendencial, agrupamos os pontos da série temporal em pares da forma $(Z_1, Z_{1+c}), (Z_2, Z_{2+c}), \dots, (Z_{N-c}, Z_N)$, onde Z são os pontos da série, $c = N/2$ e N é o total de pontos presente na série. Após isso, foi aplicado o teste do sinal ou teste Cox-Stuart conforme a Subseção 2.1.2. Além do teste Cox-Stuart, também realizamos o diagnóstico da tendência através das análise gráfica das funções de autocorrelação e autocorrelação parcial da série.

A maioria das séries macroeconômicas apresentam algum tipo de sazonalidade e, antes de aplicar determinados métodos, devemos tornar a série livre de sazonalidade. Para isto, primeiro realizamos o teste da presença de uma sazonalidade trimestral determinística nos dados usando o pacote *forecast* disponível no ambiente estatístico R [Team et al. 2013] e, depois realizamos o estudo do periodograma para verificar a possível presença sazonal.

Após isso, construímos um modelo livre de sazonalidade, onde a eliminação da sazonalidade ocorreu através do método *dummies*, removendo a parte sazonal determinística. Após à construção do modelo livre de sazonalidade, realizamos a comparação da série original com a série livre de sazonalidade através da análise gráfica, reforçando a análise do periodograma.

3.3 Descrição computacional e ajustamento dos modelos

Segundo [Samohyl et al. 2001], as técnicas de análise de séries temporais são importantes, pois permitem a identificação de padrões do comportamento histórico não-aleatórios, possibilitando não só a previsão, como também a organização de ações preventivas em relação às suas consequências. Com isso, antes de realizar a previsão, é necessário analisar como as técnicas se comportam perante a série temporal, isto é, através dos padrões obtidos pelas técnicas advindos da série temporal analisada, estimar a mesma. A esta estimação, damos o nome de ajustamento do modelo.

Os ajustamento dos modelos é importante na análise de séries temporais, pois os mesmos permitem realizar análises gráficas de como os modelos ajustados se comportam diante da série temporal, permitindo inferir se os padrões adquiridos pelas técnicas descrevem de modo satisfatório o comportamento da série.

Após a identificação das principais componentes envolvidas na série, o passo

seguinte consiste em realizar a implementação computacional e os ajustamentos dos modelos para cada uma das abordagens propostas nesse trabalho. Assim, no ajustamento dos modelos, primeiro realizamos o ajuste usando a metodologia *Box-Jenkins* ou modelo ARIMA, seguido pela obtenção dos modelos de suavização, redes neurais artificiais e as abordagens híbridas. A seguir, será apresentado a descrição computacional e obtenção do ajustamento de cada um dos modelos.

3.3.1 Modelo ARIMA

O ajustamento do modelo através da metodologia *Box-Jenkins* consistiu em ciclos iterativos envolvendo análises das funções de autocorrelação e autocorrelação parcial, seguido da fase de estimação e teste. Para realizar a validação, realizamos à análise dos resíduos e o mesmo foi considerado admissível caso o comportamento do mesmo indique ser ruído branco. Todo o processo foi implementado no software R, usando a biblioteca *forecast* [Hyndman et al. 2007].

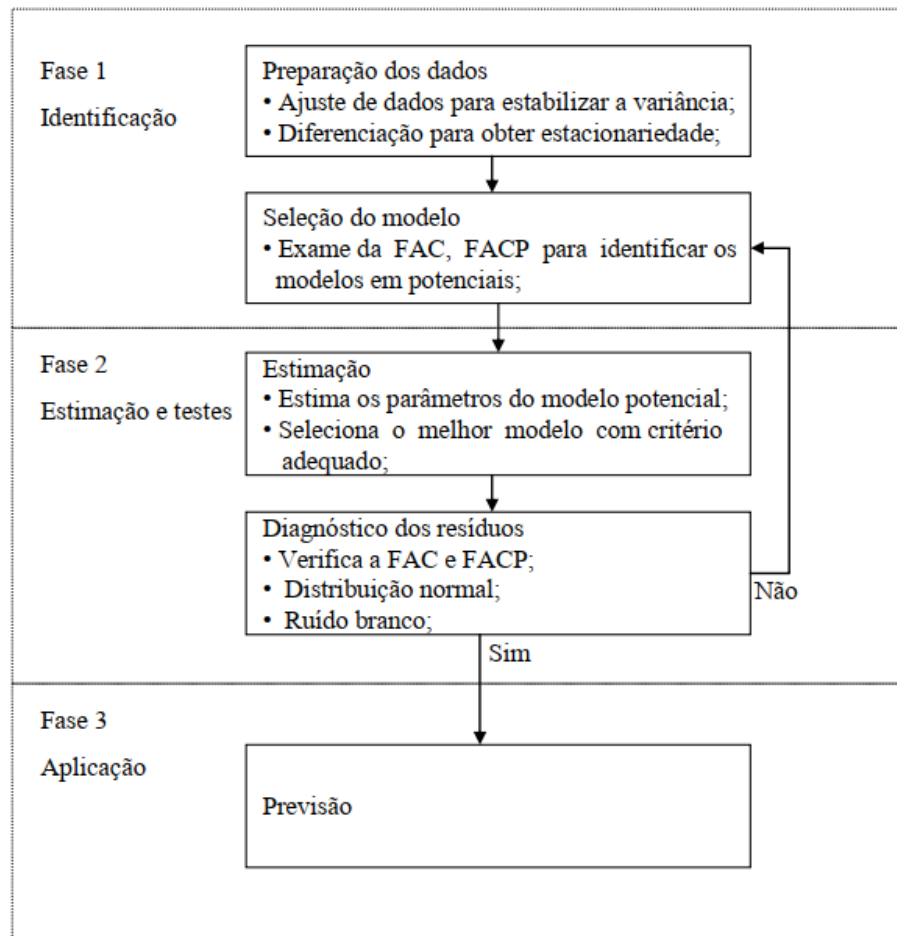
A Figura 11 ilustra todas as etapas da metodologia *Box-Jenkins* seguida neste trabalho. As etapas que envolveram o ajustamento do modelo correspondem a fase 1 e 2 na figura, e serão discutidas a seguir.

Na preparação dos dados, fizemos a identificação das principais componentes presentes na série, as quais foram definidas na Subseção 3.2. Depois disso, tomamos a diferença sucessivas da série para torná-la estacionária de acordo com a Subseção 2.1.1. Após tomadas as diferenças sucessivas, realizamos o teste de estacionariedade através do teste da raiz unitária *ADF* de acordo com a Subseção 2.1.2. Após a preparação dos dados, o passo seguinte consistiu em selecionar o modelo ARIMA(p, q) através da análise gráfica das funções de autocorrelação e autocorrelação parcial estimadas.

Encerrada a fase de identificação, procedemos com a fase de estimação e teste que corresponde ao diagnóstico dos resíduos. A estimação envolve determinar os parâmetros dos modelos, e em seguida realizar o diagnóstico do modelo estimado com a finalidade de verificar se ele representa, ou não, adequadamente, os dados. Para isso, primeiro realizamos o estudo dos resíduos fazendo à análise das *fac* verificando a condição de autocorrelação nos resíduos gerados pelo modelo estimado. Esta verificação é necessária devido a necessidade de investigar a existência de autocorrelação fora do intervalo de confiança o que indica que determinados padrões da série não foram capturados devidamente pelo modelo estimado, sugerindo que o modelo não é adequado.

Feito a análise das *fac*, o passo seguinte foi realizar a verificação de normalidade dos resíduos. Esta verificação ocorreu com a análise do histograma dos resíduos e sua representação em papel de probabilidade.

Figura 11 – Representação esquemática do algoritmo Box Jenkins



Fonte: Adaptado de [Makridakis et al. 1999]

Após a validação do modelo estimado através das etapas descritas nesta seção, obtemos o ajustamento do modelo. A fase final consistiu em utilizar este modelo para realizar as previsões apresentadas no capítulo de resultados.

3.3.2 Métodos de suavização

Para se obter os ajustamentos dos modelos, os métodos de suavização assumem que os valores extremos da série representam a aleatoriedade e, assim, por meio da suavização desses extremos, pode-se identificar o padrão básico presente na série e por meio desse padrão obter o ajustamento do modelo e depois utilizá-los na previsão.

Realizamos a implementação dos seguintes métodos de suavização: Média Móvel Simples, Suavização Exponencial Simples, Suavização Exponencial de Holt, Suavização Exponencial de Holt-Winter Aditivo e Multiplicativo. Para todos estes métodos, primeiro realizamos a implementação computacional das expressões definidas na Seção 2.3 no ambiente estatístico R. Após a implementação computacional, o passo seguinte

foi encontrar os parâmetros para cada método ainda de acordo com a Seção 2.3 e assim obter o ajustamento do modelo.

O primeiro parâmetro que precisou ser encontrado foi o número de observações (r) para calcular a média móvel do método de suavização por média móvel. O ajustamento utilizando este método depende do valor de r , portanto, selecionamos o valor de r que forneceu a melhor previsão a uma passo das observações já obtidas, ou seja, o valor de r que minimize a Eq. (38)

Para o método de suavização exponencial simples foi preciso encontrar o parâmetro que corresponde a constante de suavização conforme a expressão (39). Já o método de suavização exponencial de Holt foi necessário encontrar os parâmetros de nível e tendência conforme as expressões (42) e (43). E, por último, o método de suavização exponencial de Holt-Winters que precisou definir os parâmetros de nível, tendência e sazonalidade conforme as expressões (47), (48) e (49).

Para determinar estes parâmetros é necessário adotar alguns critérios, tais como tipo de autocorrelação entre os dados e o custo de previsão. [Morettin e Toloi 2006] afirmam que um procedimento mais objetivo é selecionar o valor das constantes que forneçam o melhor modelo de ajustamento das observações já obtidas. Seguindo esta ideia, construímos um *script* em R que realiza o teste automático para cada uma das constantes.

O *script* funciona da seguinte maneira: os valores das constantes começam em 0 incrementando-as em 0,1 até o valor 1; a cada iteração, calcula-se o erro quadrático médio das previsões de acordo com os valores das constantes; no final são selecionadas as constantes que forneceram o menor erro quadrático médio.

Definidos os parâmetros para cada um dos métodos, utilizamos os pontos da série temporal para alimentar o modelo e, por fim, obter o ajustamento de cada um dos modelos.

3.3.3 RNA

Para o ajustamento dos modelos através da RNA procedeu-se da seguinte forma:

1. Organização e seleção da série temporal da base dados;
2. Divisão da série temporal em dois subconjuntos. Sendo o primeiro conjunto para realizar o treinamento da rede e o segundo subconjunto para realizar os testes;
3. Transformar a série temporal, normalizando-a para evitar a saturação da rede;
4. Selecionar, treinar e testar a rede;

5. Aplicar à rede neural treinada para finalizar o ajustamento do modelo e realizar previsões.

O primeiro passo desta sequência envolveu a seleção da série temporal da base dados do projeto ACCB-UESC conforme descrito na Seção 3.1. Após selecionada a série temporal objeto de estudo, o segundo passo abrange a divisão da série temporal em dois subconjuntos.

Esta divisão é necessária para realizar o treinamento da rede de modo supervisionado, em que o primeiro subconjunto foi destinado ao treinamento da rede e o segundo subconjunto para realizar a validação da rede. A série temporal contém um total de 144 pontos divididos em dois subconjuntos, em que o primeiro subconjunto contém 100 pontos, utilizados no treinamento da rede. O segundo subconjunto com 44 pontos foram empregados na validação.

Depois de selecionado os subconjuntos de treinamento e teste, o passo seguinte consistiu em realizarmos a transformação dos pontos da série temporal para evitar a saturação da rede. Esta transformação consistiu em normalizar os pontos da série temporal de forma que os valores estejam compreendidos no intervalo [0,1]. A normalização foi feita de acordo com a expressão

$$Z_t^- = \frac{Z_t - \min(Z)}{\max(Z) - \min(Z)}. \quad (81)$$

Em que:

Z_t^- = valor normalizado no instante t ;

Z_t = valor da série temporal no instante t ;

$\min(Z)$ = valor máximo da série temporal;

$\min(Z)$ = valor mínimo da série temporal.

Após o tratamento dos dados, o passo seguinte correspondente ao item 4 de nossa sequência, envolve a modelagem da rede neural. Para isto, foi necessário encontrar a melhor configuração e arquitetura de rede que forneça a melhor previsão. Em [Varfis e Versino 1990], se define que, para séries do tipo econômica, a metodologia adotada deve modelar a série temporal de forma a captar as componentes da mesma, considerando três grupos, são eles: valores passados da série, valores referentes a tendência e a sazonalidade.

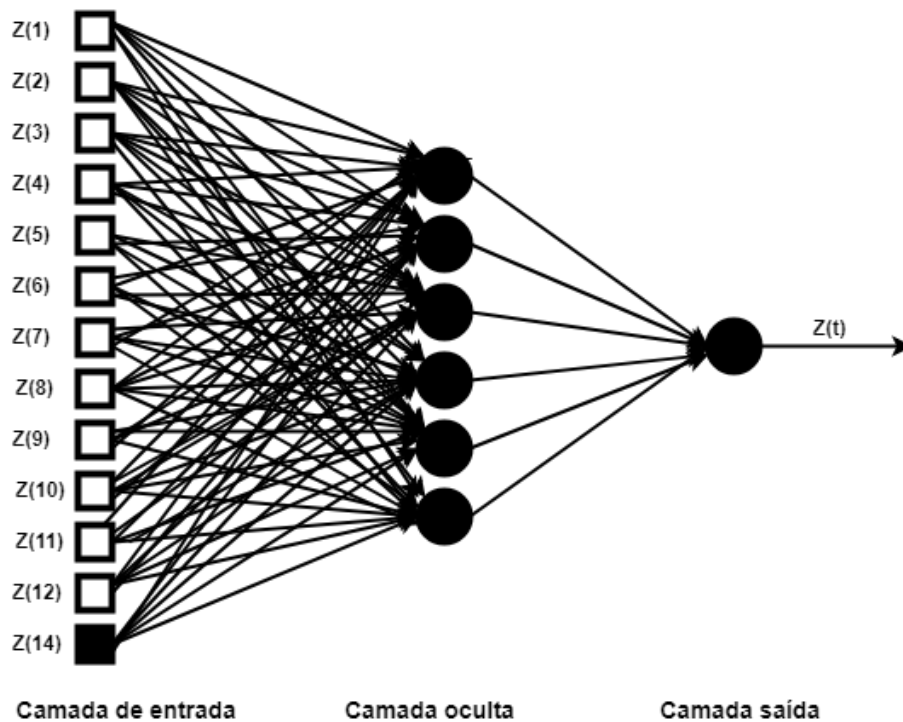
Já em [Fernandes et al. 1996], os autores propõem um modelo utilizando uma rede do tipo *multi-layer perceptron* para predição, estabelecendo na camada inicial deve conter neurônios com as observações passadas e neurônios para a modelagem do comportamento tendencial e sazonal. Na camada oculta [Fernandes et al. 1996] defen-

dem que deve ser composta por seis neurônios e a camada de saída composta por um neurônio que representa a predição.

Seguindo a metodologia proposta por [Fernandes et al. 1996], fizemos a modelagem da topologia da rede com a seguinte configuração: treze neurônios na camada inicial, isto é, doze neurônios correspondem a um ano de valores passados e um neurônio para modelagem do comportamento tendencial. A camada oculta é composta por seis neurônios e, por último, a camada final com um neurônio que é a configuração padrão.

A Figura 12 ilustra a topologia desta rede composta por camada inicial contendo 12 entradas referente a um ano de coleta da cesta básica e uma entrada (cor preta) referente a decima quarta coleta, ou seja, valor da cesta básica imediatamente ao que se quer prever. Esta configuração é responsável por modelar o comportamento tendencial da série. Seguindo, temos a camada oculta composta por seis neurônios e a camada final composta por um neurônio.

Figura 12 – Arquitetura da RNA para a série temporal do custo da cesta básica da cidade de Ilhéus, jan/2005 a dez/2016



Fonte: Elaborada pelo autor.

Como função de ativação fizemos o uso da função logística cujo resultado de saída assumirá sempre valores reais entre zero e um satisfazendo o intervalo de normalização dos pontos da série temporal definido na Eq. (81). A função logística tem sua

formulação matemática como

$$g(u) = \frac{1}{1 + e^{-\beta u}}, \quad (82)$$

onde β é uma constante real associada ao nível de inclinação da função logística frente a seu ponto de inflexão.

Para realizar o treinamento, utilizamos a estratégia de treinamento supervisionado que consistiu em dividir a série temporal, considerando cada amostra dos sinais de entrada, as respectivas saídas desejadas, ou seja, cada amostra de treinamento é composta pelos sinais de entradas e suas respectivas saídas. Assim, os pesos sinápticos e limiares foram então continuamente ajustados mediante a aplicação das ações comparativas, executadas pelo próprio algoritmo de aprendizagem, supervisionando a defasagem entre os valores previstos produzidos pela rede em relação àquelas desejadas, sendo essa diferença usada no procedimento de ajuste através do algoritmo *backpropagation*.

Como critério de parada de treinamento foi considerado o número de ciclos (100 iterações) juntamente com o erro tolerável de 10^{-3} . Após a modelagem da rede e o treinamento adequado, o passo seguinte foi então aplicar a rede neural desenvolvida para o problema proposto, obtendo o ajustamento do modelo e as previsões, ambos apresentados no capítulo de resultados.

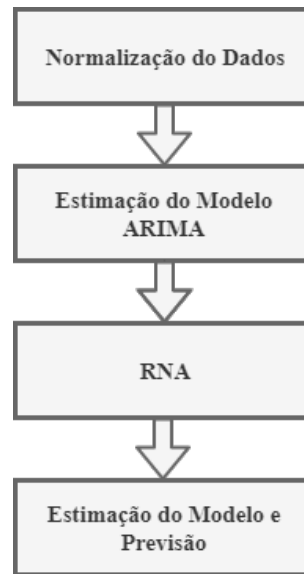
3.3.4 Modelos híbridos

As abordagens híbridas foram obtidas pela combinação de duas técnicas de modelagem, tendo como finalidade obter modelos mais poderosos e com menos deficiências. A construção dos modelos híbridos ocorreram de duas formas, sendo elas: modelo híbrido sequencial e modelo híbrido incorporado. O modelo sequencial envolve um modelo ARIMA acoplado a uma RNA; e o modelo incorporado envolve um sistema *neurofuzzy*.

3.3.4.1 ARIMA com RNA

Este sistema híbrido consiste em um modelo sequencial constituído de duas técnicas: a primeira técnica é o modelo ARIMA que trabalha com a estimação do modelo autorregressivo integrado de médias móveis, e a segunda técnica é uma RNA. Este modelo foi proposto por [Silva 2012], onde o autor define os passos para realizar a construção do modelo ARIMA da série temporal e, em seguida, utilizar-se este modelo estimado para alimentar uma RNA. A Figura 13 ilustra as etapas do modelo ARIMA com RNA.

Figura 13 – Etapas da abordagem ARIMA com RNA



Fonte: Elaborada pelo autor.

A Figura 13 apresenta os passos da construção do modelo. Iniciamos com a normalização dos pontos da série temporal conforme a eq. (81), de forma que os valores estejam compreendidos no intervalo entre 0 e 1, para evitar a saturação da rede neural. Após a normalização, a série normalizada foi utilizada para construir o modelo ARIMA. Os parâmetros do modelo ARIMA foram obtidos seguindo os passos descritos na Subseção 3.3.1. Feito a estimação do modelo ARIMA, o passo seguinte foi alimentar a rede neural usando a mesma configuração definida na Subseção 3.3.3 com a saída da primeira técnica.

3.3.4.2 Sistema neurofuzzy

Este é um sistema incorporado que combinou sistema de inferência *fuzzy* com uma RNA do tipo *feedforward*, resultando em um sistema *neurofuzzy*. A modelagem deste sistema resultou em um sistema com alto grau de hibridização, ao ponto que não é possível a separação entre as duas técnicas, ou seja, o grau de hibridização é tão elevado que se pode dizer que o primeiro subsistema contém o segundo, ou vice-versa.

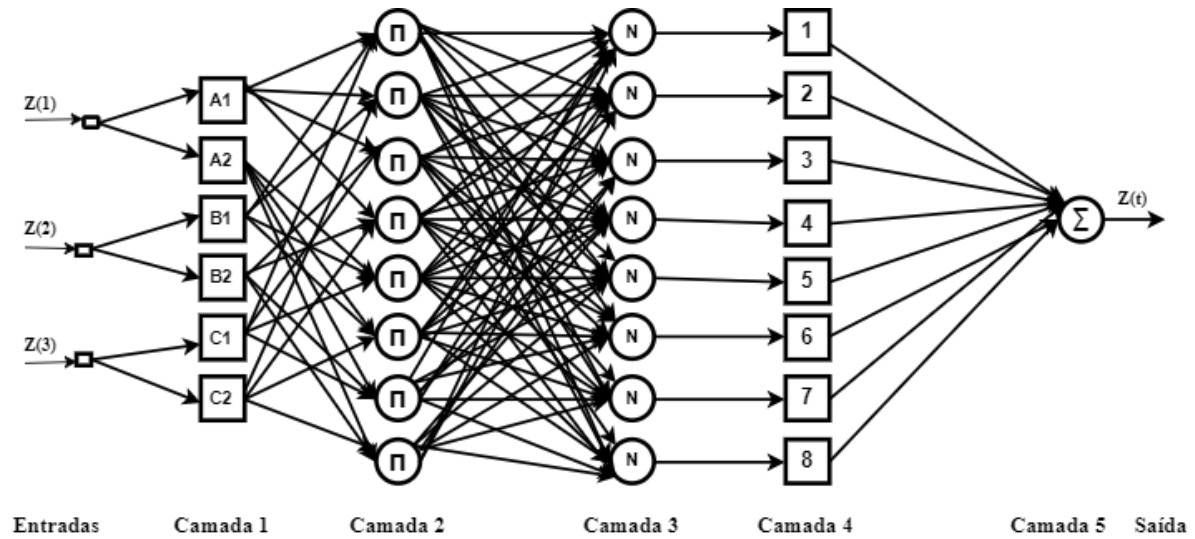
O ajustamento do modelo usando essa abordagem ocorreu através das etapas descritas a seguir.

Primeiro realizamos a divisão da série temporal em dois subconjuntos. Sendo o primeiro subconjunto para realizar o treinamento da rede e o segundo subconjunto para realizar o teste. Dos 144 pontos que compõe a série temporal, 100 pontos formam o primeiro subconjunto para realizar o treinamento do sistema e os outros 44 pontos formam o segundo subconjunto para realizar a validação do sistema utilizando o

treinamento supervisionado. Ressalta-se que é a mesma divisão de dados proposta nas RNAs anteriores.

Após a divisão da série temporal, o passo seguinte foi selecionar, treinar e testar a rede. A arquitetura selecionada foi o sistema *neurofuzzy* ANFIS por apresentar na literatura sucesso nas aplicações de previsão e aproximação de funções. A Figura 14 ilustra a arquitetura ANFIS utilizada com três entradas.

Figura 14 – Sistema neurofuzzy com três entradas para a série temporal do custo da cesta básica da cidade de Ilhéus jan/2005 a dez/2016



Fonte: Elaborada pelo autor.

Ao alimentar o sistema *neurofuzzy* ilustrado na Figura 14 com os pontos da série temporal, definidos por $Z(1)$, $Z(2)$ e $Z(3)$, pois é um sistema de três entradas, o mesmo realiza uma série de etapas em cada camada até se obter o ponto $Z(t)$ que corresponde ao ajustamento. A primeira etapa, referente a camada 1, que transforma cada ponto da série de entrada em variáveis do universo *fuzzy*. Para isto, foi usado função de pertinência do tipo gaussiana que possui o formato do tipo seno com imagem no intervalo $[0, 1]$,

$$\mu_{A_i}(Z(i)) = \frac{1}{1 + \left[\left(\frac{Z(i) - c_i}{a_i}\right)^2\right]^{b_i}} \quad (83)$$

onde $Z(i)$ corresponde a entrada do sistema para $i = 1, 2, 3$. A segunda etapa consistiu em aplicar a t-norma nos valores resultante das saídas da camada anterior. A t-norma, utilizada foi o produto do grau de ativação de cada variável de entrada conforme,

$$T(A, B, C) = w_j = \mu_{A_i}(Z(1))\mu_{B_i}(Z(2))\mu_{C_i}(Z(3)), i = 1, 2, 3. j = 1, \dots, 8. \quad (84)$$

A próxima etapa, executada na camada 3, realiza a normalização dos graus de ativação das regras resultantes da aplicação da t-norma, onde cada nó executa a função normalização, a qual é utilizada como um pré-processamento para a defuzzificação. A normalização foi dada por

$$\bar{w}_j = \frac{w_j}{w_1 + \dots + w_j}, j = 1, \dots, 8. \quad (85)$$

A etapa seguinte, correspondente a camada 4, contém funções adaptativas que realiza o cálculo com os parâmetros de consequência do sistema advindo da camada 3. Este cálculo é ajustado pela expressão

$$\bar{w}_j f_j = \bar{w}_j (p_j Z(1) + q_j Z(2) + h_j Z(3) + r_j), j = 1, \dots, 8. \quad (86)$$

A última etapa, associada com a camada 5, tem como resultado a saída crisp do sistema correspondente a previsão. Esta etapa de defuzzificação facilita-se pelo cálculo das camadas 3 e 4, e sua formulação é descrita pela expressão

$$Z(t) = \sum_{j=1}^8 \bar{w}_j f_j = \frac{\sum_{j=1}^8 w_j f_j}{\sum_{j=1}^8 w_j}. \quad (87)$$

O processo de aprendizado do sistema foi realizado em duas etapas iterativas até atingir o critério de parada. A primeira etapa consistiu em ajustar os parâmetros dos consequentes pelo método de MQO (Mínimos Quadrados Ordinários¹), enquanto os antecedentes permanecem fixos. A segunda etapa consistiu em realizar o ajuste dos parâmetros dos antecedentes pelo método do GD (*Gradient Decrescent*) [Luenberger et al. 1984], enquanto os consequentes permaneciam fixos.

A Tabela 2 apresenta resumidamente os dados de treinamento do sistema, sendo que, a rede composta por 34 neurônios, 32 parâmetros lineares, 18 parâmetros não-lineares e 8 regras *fuzzy*.

3.4 Previsão

A palavra previsão sugere conhecer um fenômeno e ou as consequências de um comportamento antes de que ele possa ser observado ou medido. Para isto, os procedimentos de previsão utilizados na prática variam muito, podendo ser simples e intuitivos ou mais quantitativos e complexos. Na análise de séries temporais, a maioria dos métodos de previsão baseiam-se na ideia de que as observações passadas contêm

¹Veja detalhes do método MQO em [Aguirre 2004]

Tabela 2 – Resumo de treinamento do sistema neurofuzzy para o preço total da cesta em Ilhéus (Período de 01/2005 a 12/2016)

Resumo de Treinamento do Sistema Neurofuzzy	
Total de Neurônios	34
Total de Parâmetros Lineares	32
Total de Parâmetros Não-Lineares	18
Total de Pontos de Treinamento	100
Total de Pontos de Validação	44
Total de Regras Fuzzy	8
Total de Épocas	200

informações sobre o padrão de comportamento da série. Assim, os modelos de previsão para obter os valores futuros utilizam os valores passados da própria série.

Quanto ao intervalo de tempo futuro, a previsão se classifica em previsões de curto prazo, médio prazo e longo prazo. A definição de curto, médio e longo prazo varia de acordo com o intervalo de tempo no qual a série é obtida. Para séries temporais obtidas mensalmente, considera-se a curto prazo a previsão de um até cinco meses a frente, a médio prazo previsões entre seis meses a três anos a frente e a longo prazo a previsão acima de três anos.

Adotamos então que faremos previsões de curto e médio prazo, sendo que, para a previsão a curto prazo consideramos o período de 4 meses e para a previsão a médio prazo foram considerados dois períodos, sendo o primeiro correspondente a 6 meses de previsão e o segundo período correspondente a 12 meses de previsão. A previsão a longo prazo foi desprezada neste trabalho devido a limitação dos dados da série temporal.

A série temporal corresponde ao período de janeiro de 2005 a dezembro de 2017. Os pontos da série temporal correspondente ao intervalo de janeiro de 2005 a dezembro de 2016 formam utilizados para realizar os ajustamentos dos modelos. Após os ajustes, foi então realizado as previsões para o ano de 2017, divididas em três períodos correspondente a quatro, seis e doze meses respectivamente. Os pontos correspondente ao intervalo de janeiro de 2017 a dezembro de 2017 foram usados para realizar a validação das previsões obtidas.

Os ajustamentos dos modelos que descrevem uma série não conduz, necessariamente, a um procedimento ou fórmula de previsão. É necessário especificar uma função-perda, além do modelo de ajustamento, para se chegar ao procedimento. As funções de perda são usadas para avaliar o desempenho dos métodos de previsão. O desempenho dos modelos está relacionado com a diferença entre os valores observados e previstos. Assim, a ideia é que o modelo escolhido minimize as medidas baseada na função-perda, ou seja, escolhamos como melhor modelo o que apresentou as menores

medidas de acurácia baseada na função-perda.

Como o objetivo deste estudo é realizar a comparação entre os métodos de previsão, optou-se por utilizar quatro medidas específicas para avaliar o erro de previsão, a saber: Erro Quadrático Médio (MSE - *Mean Squared Error*), Erro Percentual absoluto Médio (MAPE - *Mean Absolute Percentage Error*), Erro percentual absoluto médio simétrico (SMAPE - *Symmetric Mean Absolute Percentage Error*) e o Erro Máximo (*Erro Max*) [Hyndman e Koehler 2006]. As suas expressões são descritas conforme

$$MSE = \frac{\sum_{t=1}^n e_t^2}{n}, \quad (88)$$

$$MAPE = \frac{\sum_{t=1}^n \left| \frac{e_t}{Z_t} \right|}{n} 100, \quad (89)$$

$$SMAPE = \frac{\sum_{t=1}^n \frac{|e_t|}{\left(\frac{|Z_t| + |Z_t^*|}{2} \right)}}{n} 100, \quad (90)$$

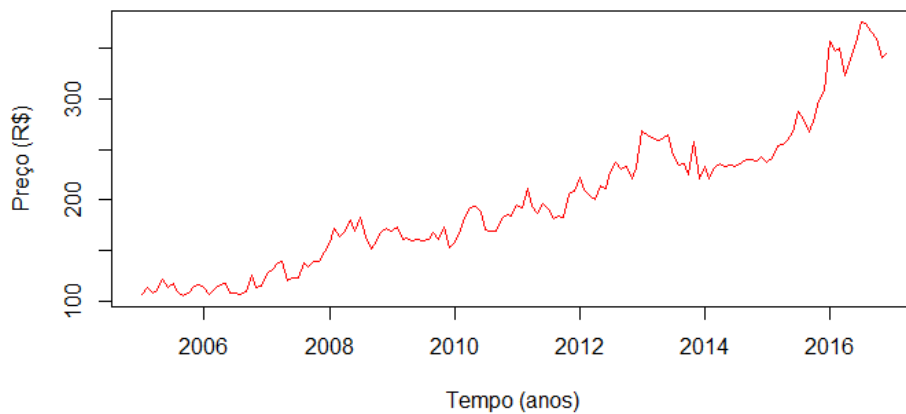
$$ErroMax = \max\left\{ \left| \frac{e_t}{Z_t} \right| \right\} 100. \quad (91)$$

Sendo t o instante da série temporal, n tamanho da série temporal, Z_t valor da série temporal no instante t , Z_t^* valor de previsão e e_t o erro de previsão definido por $e_t = |Z_t - Z_t^*|$.

4 Resultados e discussão

Assim como qualquer análise de séries temporais, o primeiro passo foi a construção de seu gráfico que revelará características importantes, como tendência, sazonalidade, variabilidade, observações atípicas, etc. A Figura 15 ilustra o comportamento da série do custo total da cesta básica da cidade de Ilhéus, no período correspondente ao intervalo de janeiro de 2005 a dezembro de 2016.

Figura 15 – Custo da cesta básica da cidade de Ilhéus, jan/2005 a dez/2016



Fonte: Elaborada pelo autor

Fazendo uma análise visual da Figura 15 podemos observar que a série apresenta comportamento crescente caracterizando a presença da componente tendencial. Quanto a componente sazonal, apenas com a análise visual, não é possível afirmar a sua presença. A seguir serão apresentados os resultados da análise da série temporal e da construção dos modelos estatísticos.

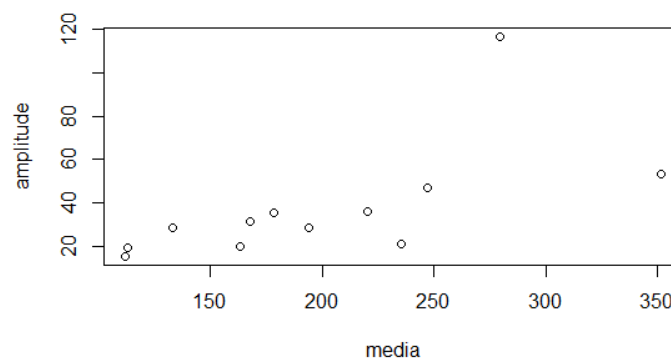
4.1 Métodos estatísticos

Nesta seção serão apresentados os resultados obtidos através da modelagem dos métodos estatísticos. Primeiro serão apresentados os resultados obtidos através da modelagem ARIMA, juntamente com a análise da série temporal, identificando as principais componentes presentes. Por último, serão apresentados os resultados obtidos através da modelagem com os métodos de suavização.

4.1.1 Análise da série e modelagem ARIMA

A primeira análise feita na série temporal consistiu em identificar a necessidade ou não de uma transformação não-linear logarítmica para estabilizar a variância e tornar o efeito sazonal aditivo. O resultado do teste pode ser visualizado na Figura 16 que ilustra graficamente a amplitude \times média. Nele podemos observar que os pontos estão distribuídos na horizontal, isto é, o comportamento no eixo das abscissas que representa a média é independente do eixo das ordenadas que representa a amplitude e, portanto, não há necessidade de aplicar a transformação.

Figura 16 – Amplitude \times média do custo da cesta básica da cidade de Ilhéus, jan/2005 a dez/2016



Fonte: Elaborada pelo autor

A próxima análise feita foi a identificação da componente tendencial através do teste do sinal ou teste Cox-Stuart e através do diagnóstico das funções de autocorrelação e autocorrelação parcial. Na Tabela 3 é apresentado o resultado do teste, em que T_2 representa o total de pares com sinal "+" e c representa o total de pares formados pelas pontos da série temporal.

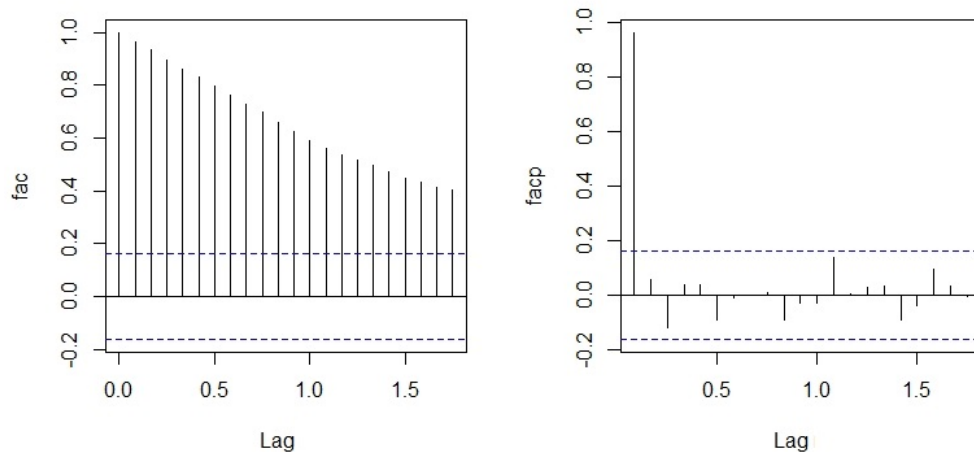
Tabela 3 – Resultado do teste Cox-Stuart

Teste Cox-Stuart	
T_2	72
c	72
p-valor	2.2e-16

Note que, a estatística T_2 coincide com o número total de pares c , portanto, rejeitamos a hipótese nula de que a série não tem tendência. Já a Figura 17 corresponde a representação gráfica das funções de autocorrelação e autocorrelação parcial. Nelas podemos observar que a f_{ac} amostral não decresce rapidamente para zero, apresentando

um decaimento lento ao longo do eixo das abscissas indicando a não-estacionariedade da série e, portanto, a série apresenta tendência como comprovado através do teste do Cox-Stuart.

Figura 17 – Autocorrelação e autocorrelação parcial amostrais do custo da cesta básica da cidade de Ilhéus, jan/2005 a dez/2016



Fonte: Elaborada pelo autor

Como identificado na análise anterior, a série apresenta a componente tendencial. Logo, foi necessário tomar a diferença da mesma para torná-la estacionária. A Figura 18 apresenta a série do custo da cesta básica da cidade de Ilhéus, de janeiro de 2005 a dezembro de 2016, acompanhada de sua primeira diferença.

Para verificar a veracidade da estacionariedade da série, foi aplicado o teste de raiz unitária ADF. A Tabela 4 apresenta o resultado do teste ADF. Comparando o valor de t com o valor crítico τ , temos que $t < \tau$, rejeitando assim a hipótese nula $H_0 : \pi = 0$, ou seja, a série não possui raiz unitária e, portanto, é estacionária.

Tabela 4 – Teste ADF para primeira diferença da série temporal do custo da cesta básica da cidade de Ilhéus, jan/2005 a dez/2016.

Diferença	Valor do teste	Valor crítico	Decisão do teste
1	$t = -13,638$	$\tau = -2,58$	$H_1 : \pi \neq 0$

Analisando a Figura 17 podemos observar que a fac não apresenta comportamento senoidal, indicando assim, que a série não apresenta sazonalidade. Já o teste considerando o modelo sazonal determinístico é apresentado na Figura 19, onde a série original é representada pela linha sólida e a série dessazonalizada representada pela linha tracejada. Note que, não há muita diferença da série original da série livre de sazonalidade, isto é, a sazonalidade presente não afeta na análise da série temporal.

Figura 18 – (a) Série temporal do custo da cesta básica da cidade de Ilhéus, jan/2005 a dez/2016 (b) Diferença do custo da cesta básica.

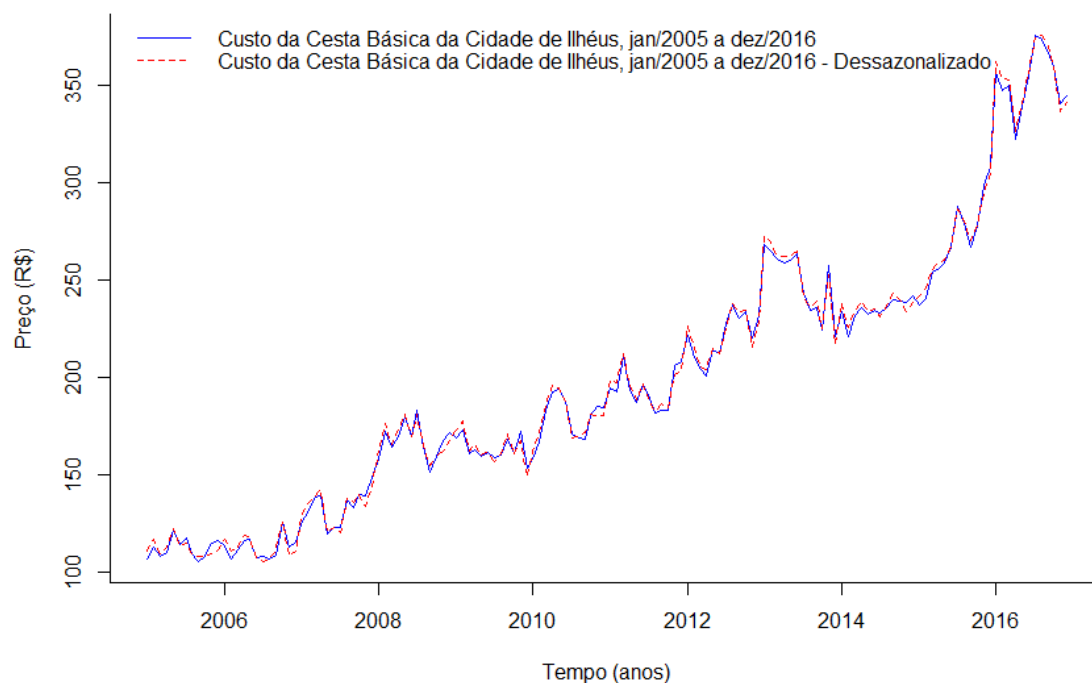


Fonte: Elaborada pelo autor

O processo de estimação dos parâmetros do modelo $ARIMA(p, d, q)$ baseou-se na análise das funções de autocorrelações e autocorrelações parciais estimadas. A Figura 20 apresenta as funções de autocorrelação e autocorrelação parcial, onde podemos observar que os dados são autocorrelacionados, tornando o modelo *Box-Jenkins* viável para a modelagem da série temporal, além disso, podemos observar que na função autocorrelação fac apresenta dois picos, nos indicando o coeficiente de autocorrelação igual a 2 ($p = 2$) e na função de autocorrelação parcial apresenta três picos significativos, levando-nos a estimar coeficiente de autocorrelação parcial igual a 3 ($q = 3$). O parâmetro d refere-se a quantidade de vezes em foi tomada a diferença da série até tornar-se estacionária. Como discutido anteriormente, foi necessário apenas tomar a primeira diferença para tornar a série estacionária, ou seja, temos $d = 1$. Com isso, temos como resultado o modelo $ARIMA(2, 1, 3)$.

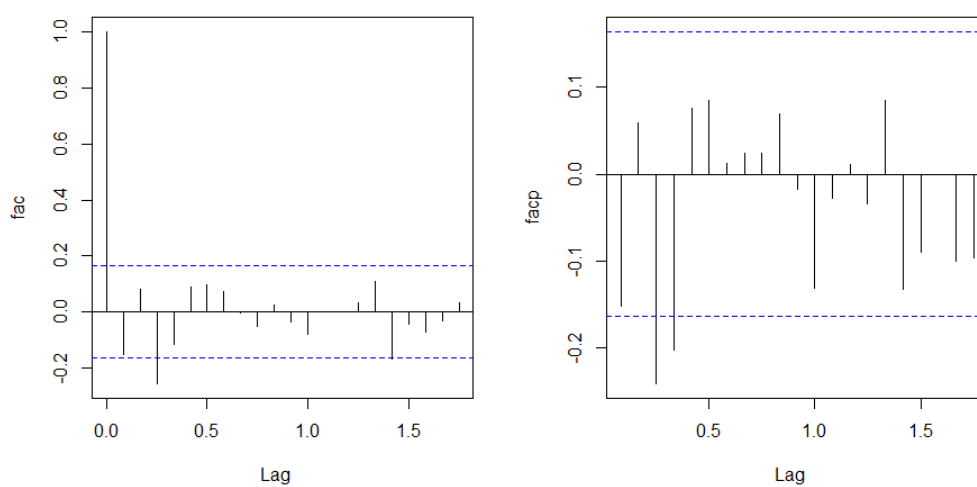
O estudo dos resíduos através da análise das $facs$, da análise do histograma e sua representação em papel de probabilidade, mostraram que os mesmos são ruído branco. A Figura 21 apresenta o correlograma da fac dos resíduos do modelo. Nela podemos observar que não há coeficiente de autocorrelação fora do intervalo de confiança, isto

Figura 19 – Modelo livre de sazonalidade do custo da cesta básica da cidade de Ilhéus, jan/2005 a dez/2016



Fonte: Elaborada pelo autor

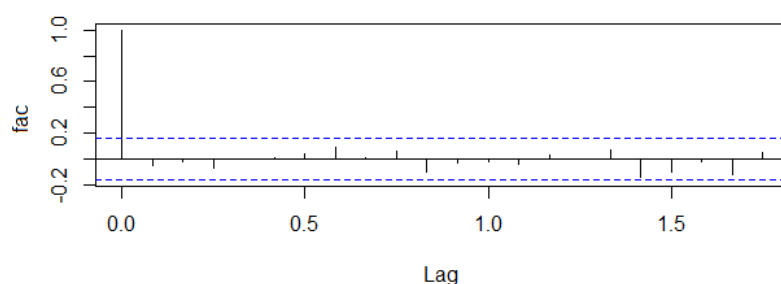
Figura 20 – Funções de autocorrelação e autocorrelação parcial do processo estacionário do custo da cesta básica da cidade de Ilhéus, jan/2005 a dez/2016



Fonte: Elaborada pelo autor

é, não há indícios de que determinados padrões presentes na série não tenham sido capturados pelo modelo.

Figura 21 – Funções de autocorrelação dos resíduos do modelo ARIMA(2,1,3) para o custo da cesta básica da cidade de Ilhéus, jan/2005 a dez/2016



Fonte: Elaborada pelo autor

A Figura 22 ilustra graficamente o histograma dos resíduos juntamente com a probabilidade normal. O histograma sugere uma distribuição aproximadamente gaussiana, indicando assim a normalidade. Já a probabilidade normal apresenta pontos que seguem praticamente a linha reta com um pequeno afastamento nas caudas, sugerindo também uma distribuição normal.

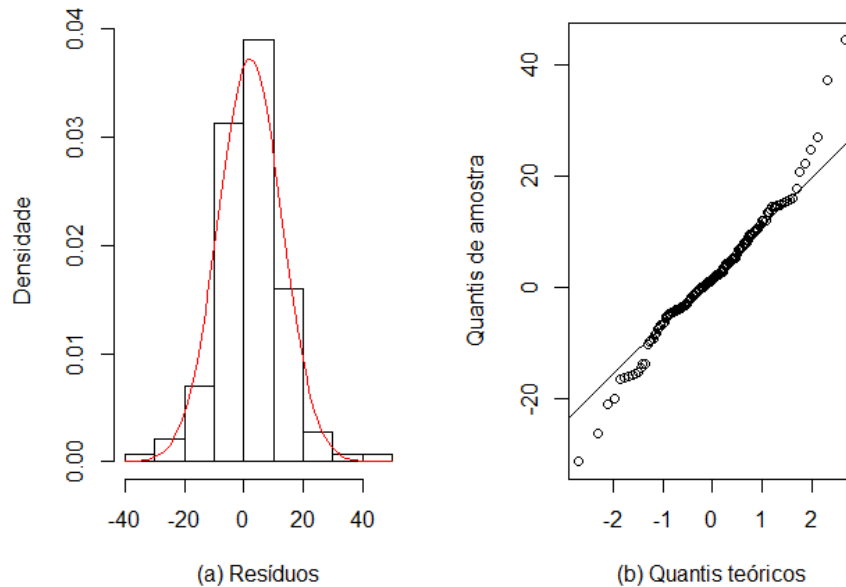
A partir da análise e testes apresentados nesta seção podemos afirmar que o modelo ARIMA(2, 1, 3) captura apropriadamente o comportamento da série e pode ser utilizado nas próximas etapas da pesquisa.

4.1.2 Ajustamento dos modelos

O ajuste do modelo autorregressivo integrado de médias móveis ARIMA(2, 1, 3) ao conjunto de dados da série temporal é apresentado na Figura 23. O modelo mostrou ter flexibilidade ao se ajustar a série temporal, conseguindo acompanhar o comportamento tendencial de forma eficiente. Entretanto, [Martins e Werner 2014] afirmam que este modelo tende a perder a capacidade de captura dos padrões ao lidar com séries temporais grandes, além da grande dificuldade em se adaptar a inclusão de novos dados ao modelo.

O primeiro ajustamento obtido foi do método de SMMS. Este método tem como principal vantagem a simplicidade de implementação e a flexibilidade de lidar com séries com um número pequeno de observações devido a variação da média móvel de acordo com o padrão encontrado. A Figura 24 apresenta o ajustamento do modelo SMMS considerando o valor $r = 3$ para calcular a média móvel, onde a linha sólida representa a série original e a linha tracejada representa o ajustamento do modelo.

Figura 22 – Análise de resíduos do modelo ARIMA(2,1,3) para o custo da cesta básica da cidade de Ilhéus, jan/2005 a dez/2016. (a) Histograma dos resíduos, (b) Probabilidade normal



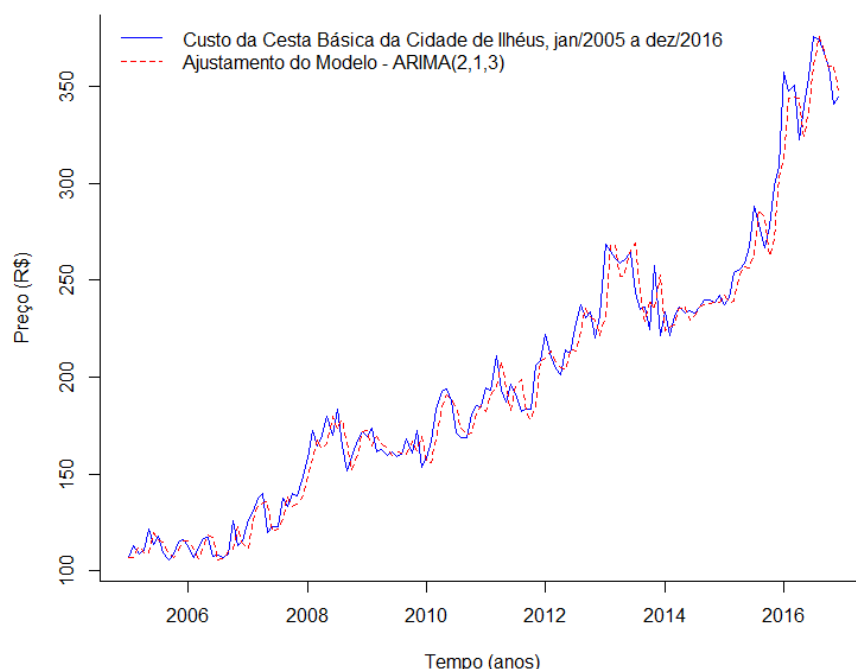
Fonte: Elaborada pelo autor

A Figura 25 representa o ajustamento do modelo através do método SES considerando o valor da constante de suavização $\alpha = 0.610$, onde a linha sólida representa a série original e a linha tracejada representa o ajustamento do modelo. Mesmo o método sendo indicado para séries temporais que não apresenta tendência, o método produziu um ajustamento do modelo em que conseguiu acompanhar o comportamento tendencial da série.

O método de SH é indicado para séries temporais que apresenta a componente tendencial, o que torna o método mais compatível ao nosso problema. A Figura 26 ilustra o ajustamento do modelo considerando as constantes de suavização como $A = 0.35$ e $C = 0.4$, em que a linha sólida representa série temporal e a linha tracejada representa o ajustamento do modelo. Note que, assim como o método SES, este método também conseguiu acompanhar o comportamento tendencial presente.

O modelo de SHW é indicado para séries temporais com comportamento mais geral, que apresentem a componente tendencial e sazonal. A grande vantagem desse método é a capacidade de atribuir pesos a média móvel, modelar a tendência e a sazonalidade. Este método é dividido em dois modelos preditivos, sendo o primeiro indicado para séries com comportamento sazonal aditivo e o segundo para séries temporais com comportamento sazonal multiplicativo. As Figuras 27 e 28 ilustram os

Figura 23 – Ajustamento do modelo ARIMA(2,1,3) para o custo da cesta básica da cidade de Ilhéus, jan/2005 a dez/2016



Fonte: Elaborada pelo autor

ajustamentos dos modelos para a modelagem usando método aditivo e multiplicativo respectivamente. Para os modelos aditivos foram consideradas as seguintes constantes de suavização: $A = 0.658$, $B = 0.2$ e $C = 0.659$. Já para os modelos multiplicativos foram consideradas as seguintes constantes de suavização: $A = 0.344$, $B = 0.584$ e $C = 0.365$.

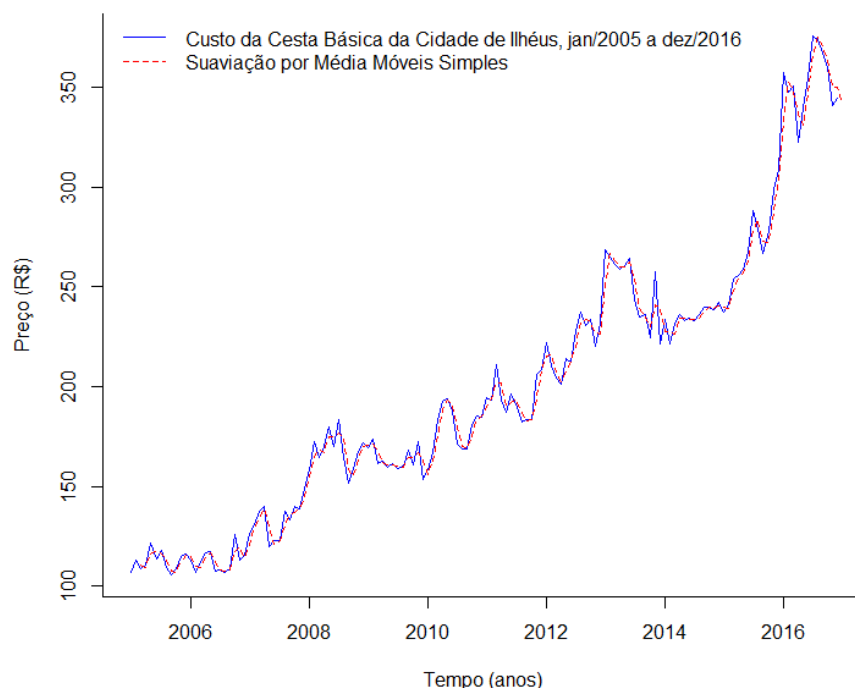
Analisando as Figuras 27 e 28, podemos perceber que o ajuste de Holt-Winters sazonal aditivo (Figura 27) conseguiu acompanhar melhor o comportamento da série do que o ajuste usando o modelo Holt-Winters sazonal multiplicativo (Figura 28).

Na Tabela 5 são apresentadas as medidas de acurácia dos ajustamentos dos modelos dos métodos estatísticos. Nela podemos observar que todos os métodos foram bastante competitivos ao realizar o ajustamento do modelo, destacando-se o modelo de SHW aditivo que apresentou melhor ajustamento em termo de acurácia, mostrando que o modelo de ajustamento conseguiu modelar com eficiência as duas componentes das constantes de suavização composta pela tendência e sazonalidade.

4.1.3 Previsão

Nesta subseção apresentamos os resultados das previsões usando os modelos estatísticos, onde realizamos previsões a curto prazo e a médio prazo conforme indicado

Figura 24 – Ajustamento do modelo SMMS para o custo da cesta básica da cidade de Ilhéus, jan/2005 a dez/2016



Fonte: Elaborada pelo autor

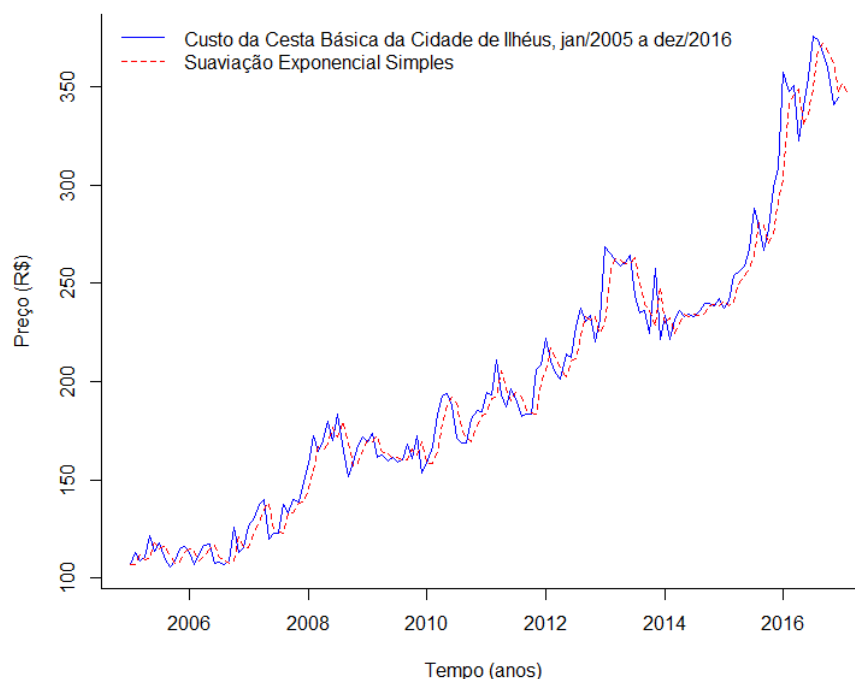
Tabela 5 – Acurácia dos ajustamentos dos modelos estatísticos para o preço total da cesta em Ilhéus (Período de 01/2005 a 12/2016)

Métodos	MSE	MAPE	SMAPE	ErroMax
ARIMA	71,0	3,62	3,6	17,2
Média Móveis Simples	90,5	2,88	2,8	14,2
Exponencial Simples	72,7	3,48	3,4	15,7
Exponencial de Holt	127	4,65	4,6	20,5
Holt-Winters Aditivo	55,3	1,94	1,8	14,00
Holt-Winters Multiplicativo	78,3	1,84	1,8	18,1

na Seção 3.4. Na Tabela 6 são apresentadas as previsões nos períodos de quatro, seis e doze meses.

Nas Tabelas 7 e 8 são apresentados os erros de previsão para os métodos estatísticos para previsão a curto e médio prazo, correspondendo a quatro e seis meses de previsão respectivamente. Analisando as medidas de acurácias, podemos notar que o método SES apresentou previsões mais acuradas nos dois períodos. Este desempenho é justificado devido a flexibilidade permitida pela constante de suavização α que dá pesos maiores às observações mais recentes, fazendo com que a previsão fique alerta a mudanças no nível da série. Entretanto, como a previsão é realizada utilizando apenas a

Figura 25 – Ajustamento do modelo SES para o custo da cesta básica da cidade de Ilhéus, jan/2005 a dez/2016



Fonte: Elaborada pelo autor

observação mais recente e o valor de α , fez com que o método perdesse a capacidade de adquirir o padrão das observações mais antigas, fazendo com que o método perdesse o poder preditivo para previsões a médio prazo.

Na Tabela 9 são apresentadas as medidas de acurácia para as previsões a médio prazo, correspondendo a doze meses de previsão. Nela, podemos notar que o método SH apresentou melhor poder preditivo devido a capacidade em lidar com séries temporais que apresentam comportamento tendencial. Ao contrário do método SES, que suaviza apenas um só nível, o método SH utiliza uma nova constante de suavização para modelar a tendência da série, levando o método a produzir previsões a médio prazo mais acuradas.

A Figura 29 ilustra graficamente os resultados apresentados na tabela de previsão anterior. Embora os métodos de suavização apresentaram melhores resultados em termo de acurácia, conforme as Tabelas 7, 8 e 9, os mesmo não conseguiram acompanhar o comportamento oscilatório da série. Já o modelo *Box-Jenkins* apresentou melhores resultados nas previsões a curto prazo para o período de quatro meses, além de acompanhar o comportamento oscilatório da série temporal.

Tabela 6 – Previsões dos métodos estatísticos para o preço total da cesta em Ilhéus (Período de 01/2017 a 12/2017)

Índice	Valor Real	ARIMA	SMMS	SES	SH	SHWA	SHWM
1	341,74	344,58	342,70	345,53	353,41	338,24	342,37
2	339,22	345,54	347,4	346,07	349,45	336,22	314,53
3	345,29	347,07	346,55	345,91	346,45	334,31	321,92
4	357,61	328,91	345,16	345,96	342,96	332,38	323,01
5	342,11	317,58	347,04	345,94	339,48	330,49	337,58
6	336,46	316,47	345,86	345,95	336,00	328,56	339,74
7	351,03	310,62	346,10	345,95	332,52	326,61	335,44
8	328,45	309,38	346,45	345,95	329,04	324,67	330,65
9	308,69	312,58	345,98	345,95	325,55	322,70	321,40
10	319,36	309,04	346,28	345,95	322,07	320,74	312,71
11	325,24	313,70	346,21	345,95	318,59	318,81	314,01
12	313,25	306,32	346,13	345,95	315,11	316,90	320,58

Tabela 7 – Medidas de acurácias das previsões dos métodos estatísticos para o preço total da cesta em Ilhéus (Período de 01/2017 a 04/2017)

Medidas	ARIMA	SMMS	SES	SH	SHWA	SHWM
SME	218,717	56,107	49,349	114,202	194,590	588,328
MAPE	2,808	1,634	1,641	2,715	3,035	5,976
SMAPE	2,8	1,8	1,8	2,6	3,0	6,2
ErroMax	8,025	3,481	3,258	4,097	7,055	9,675

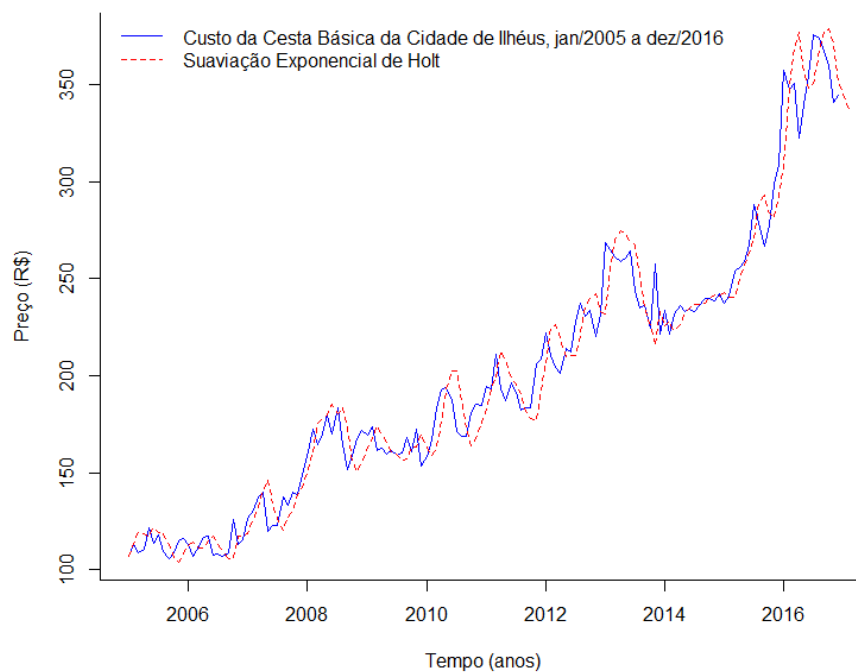
Tabela 8 – Medidas de acurácias das previsões dos métodos estatísticos para o preço total da cesta em Ilhéus (Período de 01/2017 a 06/2017)

Medidas	ARIMA	SMMS	SES	SH	SHWA	SHWM
SME	312,697	56,181	50,353	77,103	162,633	397,431
MAPE	4,058	1,795	1,750	1,913	2,981	4,368
SMAPE	4,18	1,8	1,77	1,9	3,0	4,4
ErroMax	8,025	3,481	3,258	4,097	7,055	9,675

Tabela 9 – Medidas de acurácia das previsões dos métodos estatísticos para o preço total da cesta em Ilhéus (Período de 01/2017 a 12/2017)

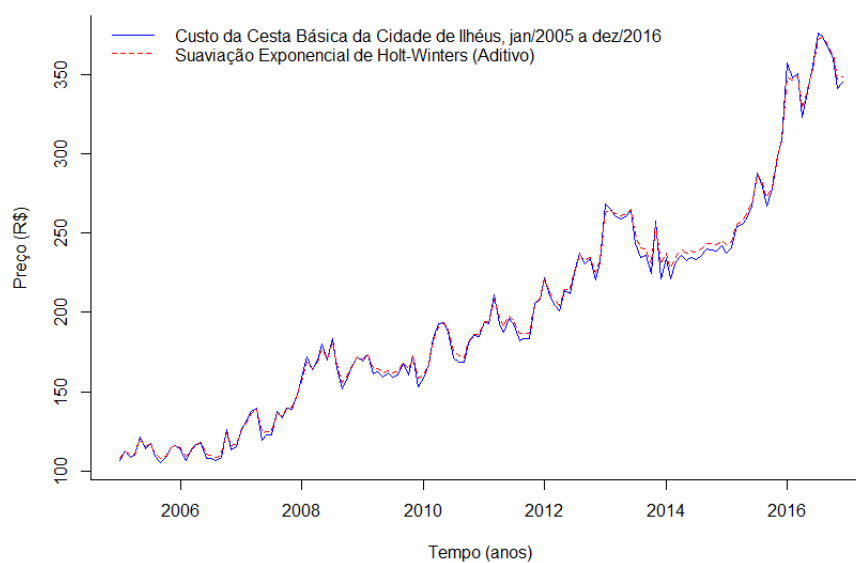
Medidas	ARIMA	SMMS	SES	SH	SHWA	SHWM
SME	347,970	360,121	352,310	95,517	153,272	251,507
MAPE	4,327	4,592	4,540	2,180	2,842	3,610
SMAPE	4,47	4,42	4,4	2,2	2,89	3,69
ErroMax	11,511	12,080	12,070	5,461	7,055	9,675

Figura 26 – Ajustamento do modelo SH para o custo da cesta básica da cidade de Ilhéus, jan/2005 a dez/2016



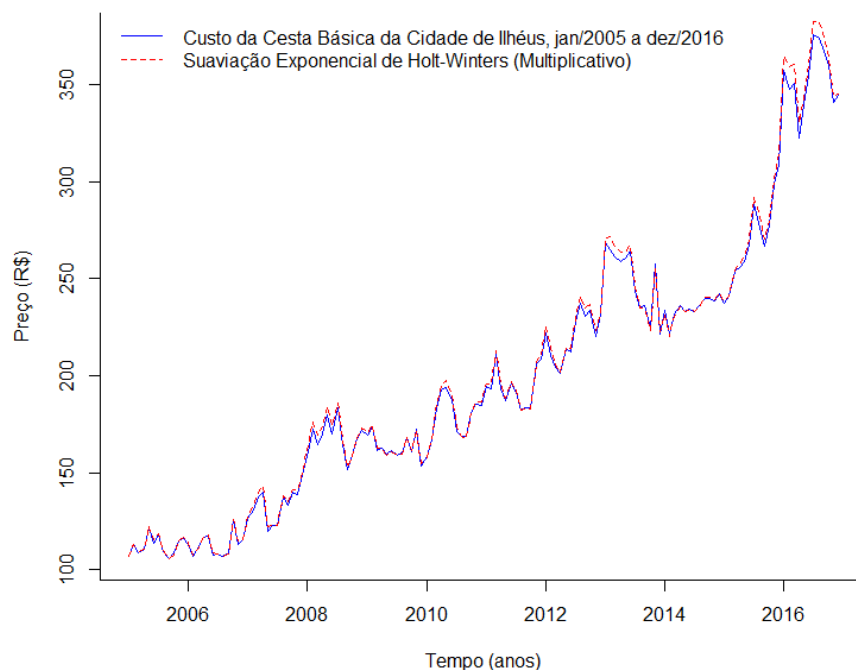
Fonte: Elaborada pelo autor

Figura 27 – Ajustamento do modelo SHW aditivo para o custo da cesta básica da cidade de Ilhéus, jan/2005 a dez/2016



Fonte: Elaborada pelo autor

Figura 28 – Ajustamento do modelo SHW multiplicativo para o custo da cesta básica da cidade de Ilhéus, jan/2005 a dez/2016



Fonte: Elaborada pelo autor

4.2 Aprendizagem de máquina

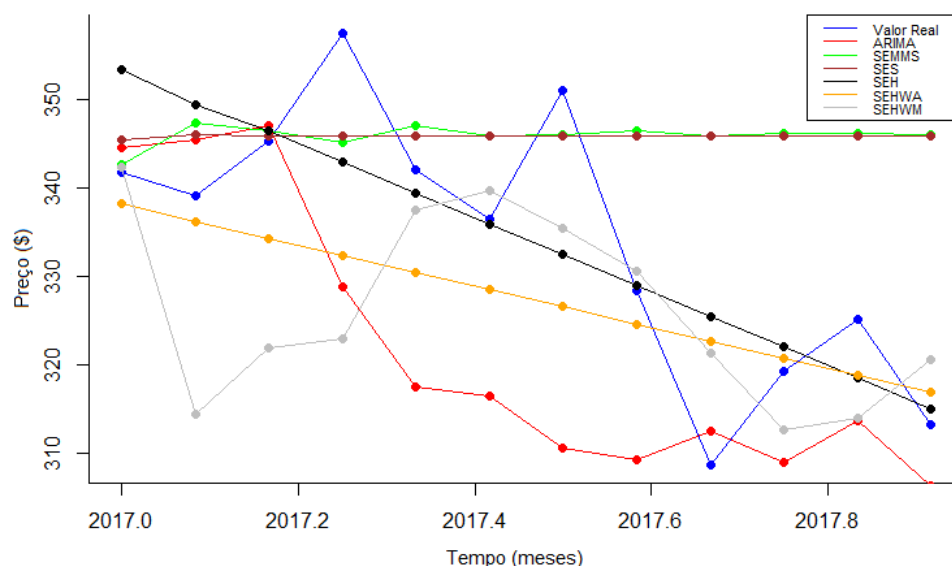
Nesta seção serão apresentados os resultados dos ajustamentos dos modelos das abordagens utilizando RNA e dos modelos híbridos, assim como as previsões realizadas com eles.

4.2.1 Ajustamentos dos modelos

No ajuste do modelo RNA utilizou-se a série temporal como entrada da rede com a finalidade da mesmo capturar e aprender as principais e através das mesmas, obter um modelo com a capacidade de generalização. A Figura 30 ilustra o ajustamento do modelo, onde a linha sólida representa os dados reais e a linha tracejada representa o ajustamento do modelo usando as características descritas na Subseção 3.3.3.

A junção de técnicas tende a construir modelos mais robustos com a absorção das características individuais de cada modelo. A Figura 31 ilustra o ajustamento do modelo híbrido que combinou RNA com ARIMA. A capacidade de generalização da RNA juntamente com a capacidade de gerar previsões mais acuradas a curto prazo do modelo ARIMA contribuiu para gerar um ajustamento do modelo mais preciso do que

Figura 29 – Previsão dos métodos estatísticos para o custo da cesta básica da cidade de Ilhéus, jan/2017 a dez/2017



Fonte: Elaborada pelo autor

o ajuste usando estas técnicas individualmente.

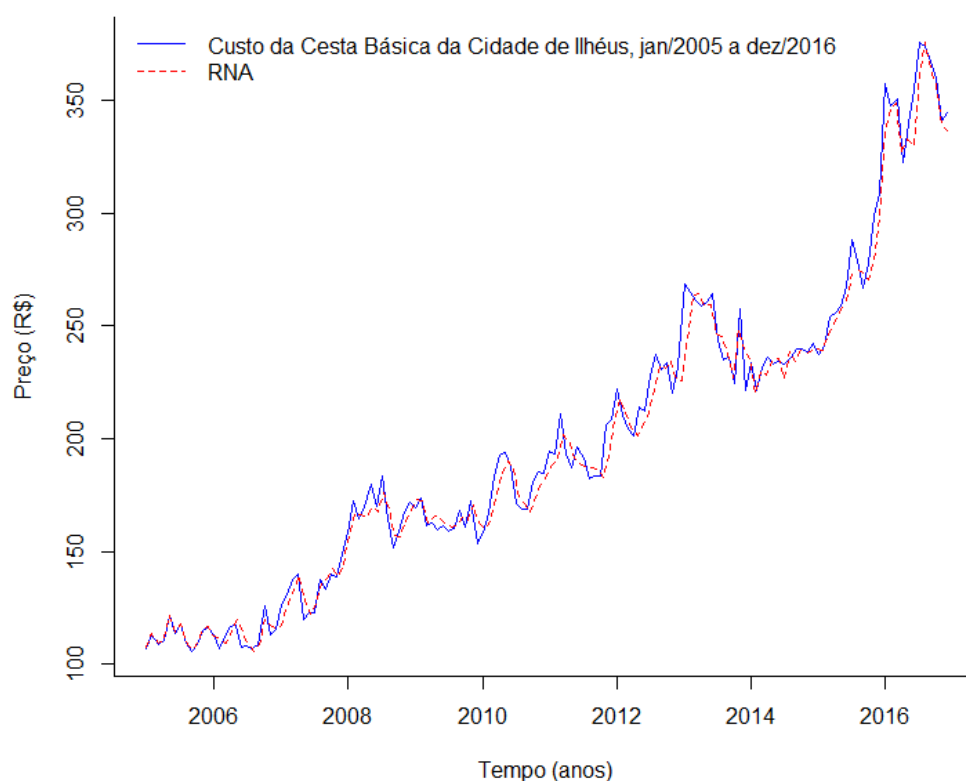
A segunda abordagem combinou RNA com um sistema de inferência *fuzzy*. Uma das principais vantagens dessa abordagem está vinculada à modelagem das funções de pertinência para cada entrada, o que possibilita manter uma relação entre as entradas e suas respectivas saídas. A Figura 32 ilustra o ajustamento do modelo do sistema *neurofuzzy*.

Na Tabela 10 são apresentadas as medidas de acurácia para os ajustamentos dos métodos de aprendizagem de máquina, com destaque para o modelo híbrido resultante da combinação ARIMA com RNA, que obteve melhor ajustamento do modelo em termos de acurácia. A combinação dessas duas abordagens (ARIMA com RNA), resultou em um modelo com a absorção das características individuais de cada um dos modelos, o que possibilitou um melhor ajustamento do modelo.

Tabela 10 – Acurácia dos ajustamentos dos modelos para o preço total da cesta em Ilhéus (Período de 01/2005 a 12/2016)

Métodos	MSE	MAPE	SMAPE	ErroMax
RNA	33,3	2,05	2,0	9,67
RNA com ARIMA	11,5	1,122	1,0	27,3
Sistema Neurofuzzy	17,5	1,50	1,4	6,08

Figura 30 – Ajustamento do modelo RNA para o custo da cesta básica da cidade de Ilhéus, jan/2005 a dez/2016



Fonte: Elaborada pelo autor

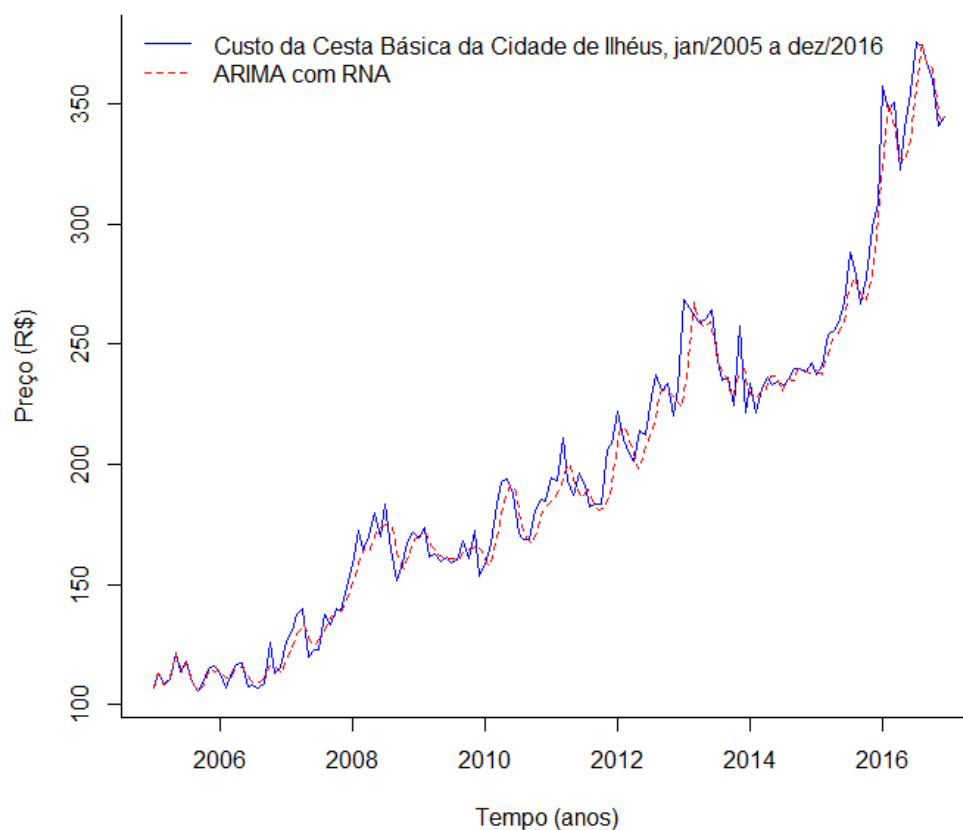
4.2.2 Previsão

As abordagens híbridas e RNA produziram previsões mais acuradas do que as previsões usando os métodos estatísticos, além de todos fornecerem previsões que acompanham o comportamento oscilatório da série temporal. Na Tabela 11 são apresentadas as previsões a curto e médio prazo para os períodos de quatro, seis e doze meses respectivamente para as abordagens RNA, RNA com ARIMA (H1) e sistema *neurofuzzy* (H2).

A Tabela 12 apresenta as medidas de acurácia para a previsão no período de quatro meses. Note que, neste período de previsão, a abordagem híbrida (H1) apresentou melhores resultados em termos de acurácia devido a capacidade de generalização da RNA juntamente com o poder preditivo a curto prazo apresentado pelo modelo ARIMA.

Nas Tabelas 13 e 13 são apresentados as medidas de acurácia para as previsões a médio prazo no período de seis e doze meses, respectivamente. Nelas podemos observar

Figura 31 – Ajustamento do modelo híbrido ARIMA com RNA para o custo da cesta básica da cidade de Ilhéus, jan/2005 a dez/2016



Fonte: Elaborada pelo autor

que, o sistema *neurofuzzy* produziu previsões mais acuradas e, além disso, podemos observar que o sistema *neurofuzzy* tende a diminuir o erro na medida em que aumenta a quantidade de previsão. Este resultado é justificado pela habilidade de associar a capacidade de aprendizado e de tolerância a falhas da RNA com a interpretabilidade dos sistemas *fuzzy* ao lidar com dados imprecisos.

Na Figura 33 são ilustrados os resultados obtidos pelos métodos de aprendizagem de máquina. Nela podemos observar que ao contrário dos métodos estatísticos, os métodos de aprendizagem de máquina produziram previsões que conseguiram acompanhar o comportamento da série ao longo do tempo, tornando-os bons modelos preditivos. A abordagem ARIMA com RNA apresentou melhores resultados nas previsões no período de quatro meses, perdendo eficácia a medida em que aumenta o número de meses de previsão. Já o sistema *neurofuzzy* não apresentou boa eficiência para previsões imediatas quando comparados com a abordagem ARIMA com RNA, entretanto, superou todos os métodos na previsão a médio prazo.

Tabela 11 – Previsões obtidas pelos métodos RNA e as abordagens híbridas para o preço total da cesta em Ilhéus (Período de 01/2017 a 12/2017)

Índice	Valor Real	RNA	H1	H2
1	341,74	346,00	345,47	338,49
2	339,22	339,56	340,49	336,33
3	345,29	342,45	342,91	342,11
4	357,61	336,99	352,48	352,82
5	342,11	332,59	349,52	337,87
6	336,46	338,45	337,59	333,68
7	351,03	335,48	345,73	347,40
8	328,45	323,71	338,51	325,02
9	308,69	322,87	317,92	306,87
10	319,36	330,71	315,35	319,09
11	325,24	328,33	320,43	324,33
12	313,25	315,00	310,73	312,01

Tabela 12 – Medidas de acurácia das previsões para o preço total da cesta em Ilhéus (Período de 01/2017 a 04/2017)

Medidas	RNA	H1	H2
SME	91,018	11,877	13,470
MAPE	1,884	0,897	1,021
SMAPE	1,9	0,8	1,02
ErroMax	5,767	1,434	1,340

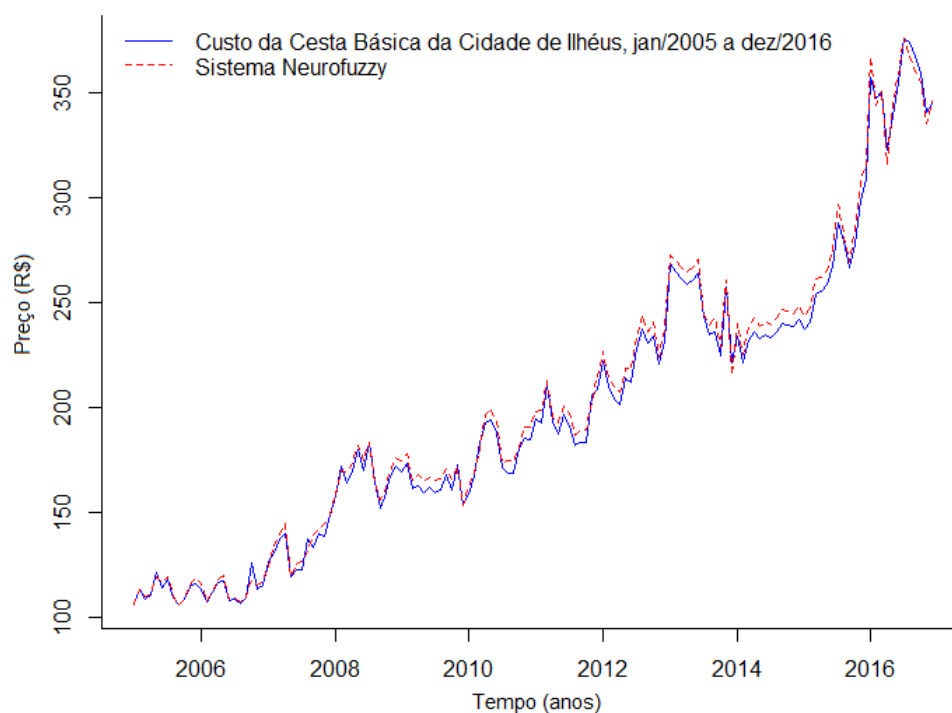
Tabela 13 – Medidas de acurácias das previsões para o preço total da cesta em Ilhéus (Período de 01/2017 a 06/2017)

Medidas	RNA	H1	H2
SME	91,017	20,052	12,947
MAPE	1,884	1,162	1,0215
SMAPE	1,8	1,0	1,0
ErroMax	5,766	2,165	1,340

Tabela 14 – Medidas de acurácia das previsões para o preço total da cesta em Ilhéus (Período de 01/2017 a 12/2017)

Medidas	RNA	H1	H2
SME	96,073	31,698	9,030
MAPE	2,237	1,507	0,797
SMAPE	2,2	1,4	0,8
ErroMax	5,7660	3,062	1,340

Figura 32 – Ajustamento do modelo híbrido sistema neurofuzzy para o custo da cesta básica da cidade de Ilhéus, jan/2005 a dez/2016



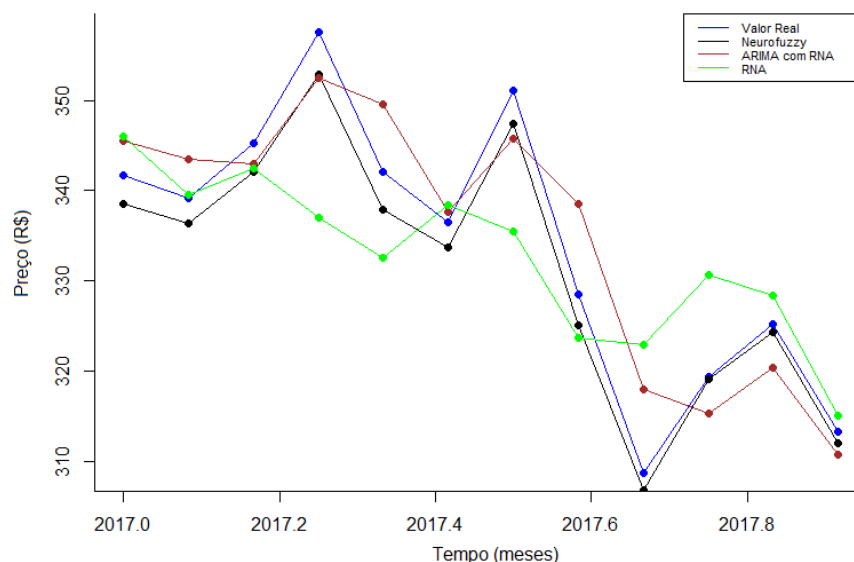
Fonte: Elaborada pelo autor

4.3 Comparação entre os modelos

Nesta seção comparamos o desempenho de predição dos modelos construídos nesta pesquisa, na previsão da série temporal do custo total da cesta básica da cidade de Ilhéus. A Figura 34 ilustra os erros quadráticos médios das previsões realizadas nos períodos de quatro e seis meses para previsão a curto prazo e doze meses para previsão a médio prazo.

Analisando a Figura 34, podemos constatar que entre os métodos de suavização, os métodos SMMS e SES apresentaram melhores desempenhos nas previsões de quatro e seis meses. Ainda entre os métodos de suavização, o SH mostrou ser o mais estável ao realizar previsões nos três períodos e, os que apresentaram piores desempenhos foram SHW aditivo e multiplicativo. Entre os métodos ARIMA e RNA, a RNA mostrou ser mais eficaz nos três períodos de previsão, aumentando o erro na medida em que aumenta também o número de previsões. Já abordagem ARIMA com RNA aprendeu melhores resultados nas previsões no período de quatro meses, perdendo eficiência à medida em que aumenta o número de previsões. Por outro lado, o sistema *neurofuzzy*

Figura 33 – Previsões dos métodos de aprendizagem de máquina para o Preço total da cesta em Ilhéus (Período de 01/2017 a 12/2017)



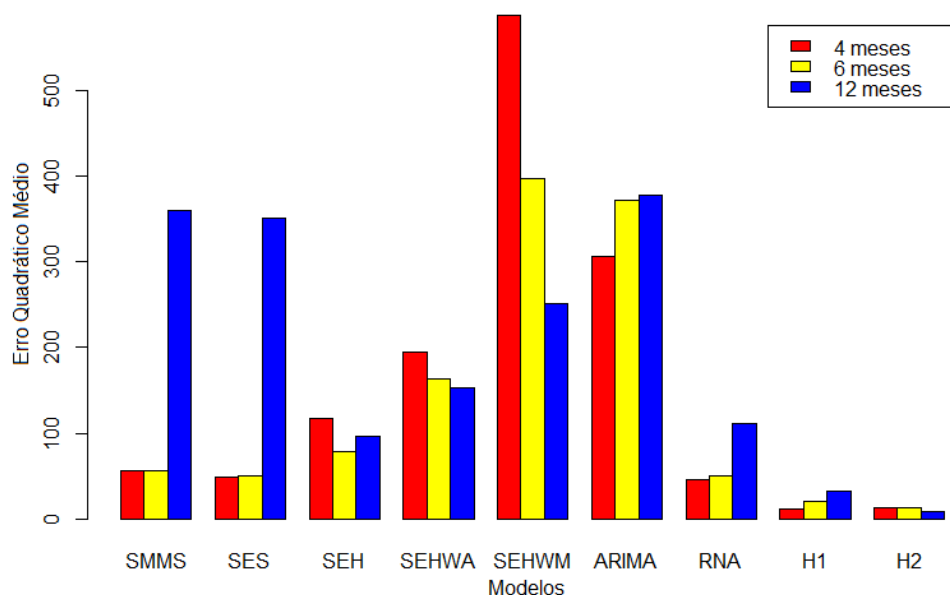
Fonte: Elaborada pelo autor

não apresentou boa eficiência para previsões imediatas quando comparado à abordagem RNA com ARIMA, porém, superou todos os métodos na previsão a médio prazo.

A Figura 35 ilustra o erro quadrático médio de cada modelo agrupado por períodos previsão. Nela podemos observar que, para o período de quatro meses, o modelo híbrido RNA com ARIMA apresentou melhor desempenho e o que apresentou pior desempenho foi o método SHW multiplicativo. Para os períodos de seis e doze meses, o modelo híbrido sistema *neurofuzzy* apresentou melhor desempenho, enquanto que o pior desempenho no período de seis meses foi obtido pelo método SHW multiplicativo e para o período de doze meses, o pior desempenho foi obtido pelo modelo ARIMA. Logo, de acordo com as Figuras 34 e 35, podemos perceber que as abordagens híbridas são mais eficazes ao realizar previsões a curto e médio prazo. Esta conclusão confirma os resultados presentes na literatura sobre a previsão de séries temporais usando abordagens híbridas.

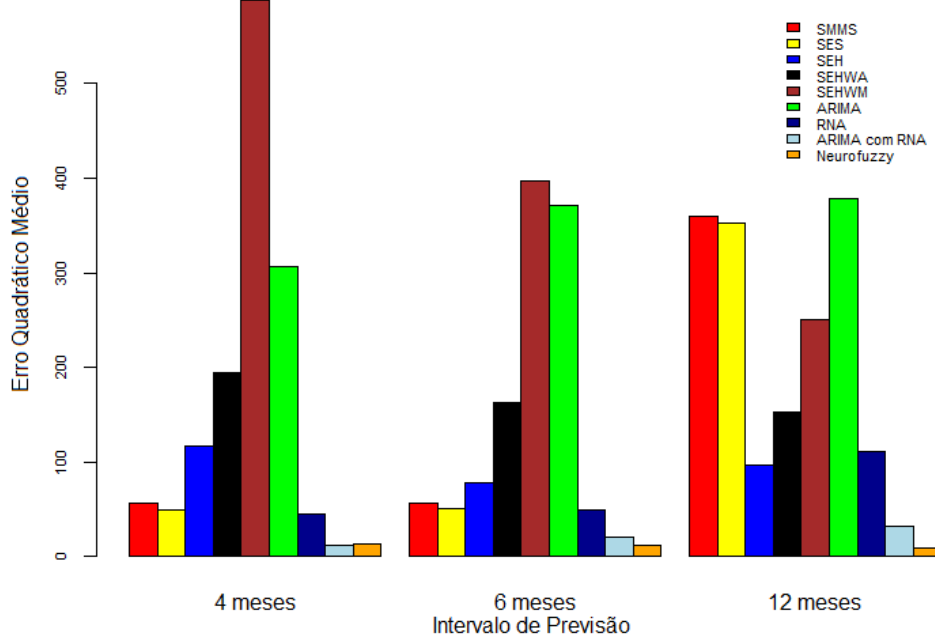
Ao compararmos os modelos constituídos nesta pesquisa, consideramos a medida de acurácia de erro quadrático médio. Destacamos que, conclusões semelhantes em relação ao desempenho dos métodos são obtidas se consideramos algumas das outras medidas de acurácia empregadas neste trabalho.

Figura 34 – Erro quadrático médio das previsões do custo da cesta básica da cidade de Ilhéus, jan/2017 a dez/2017



Fonte: Elaborada pelo autor

Figura 35 – Erros quadráticos médios agrupados nos períodos de quatro, seis e doze meses de previsões para o custo da cesta básica da cidade de Ilhéus, jan/2017 a dez/2017



Fonte: Elaborada pelo autor

5 Conclusões e trabalhos futuros

Neste trabalho buscou-se conhecer entre diversas técnicas estatísticas e de inteligência artificial, qual ou quais fornecem melhor modelo preditivo de séries temporais de dados socioeconômico do projeto ACCB. Especificamente foram considerados métodos estatísticos de suavização, um modelo ARIMA, um modelo de RNA e dois modelos híbridos que combinou ARIMA com RNA e sistema *fuzzy* com RNA.

Os métodos de suavização mostraram-se bastante competitivos em relação ao ajustamento do modelo, entretanto, os mesmos mostraram ser métodos deficientes em termo de acurácia ao realizar previsões longas. Sendo assim, os métodos de suavização, mostraram-se ser uma opção válida para realizar previsão um mês após o ajustamento do modelo, sendo vantajosos pela simplicidade computacional.

Os métodos ARIMA e RNA mostraram ser uma opção válida na análise de séries temporais, apresentando resultados superiores em relação aos métodos de suavização. Entretanto, o método ARIMA tem como grande dificuldade a determinação do ajuste do modelo, um processo de passos iterativos que necessita na maioria das vezes de conhecimento em alguma ferramenta computacional. Já o modelo RNA mostrou ser superior em relação ao método ARIMA, tanto no ajustamento do modelo como na previsão.

Já as abordagens híbridas mostraram ser uma opção satisfatória na análise de séries temporais, apresentando resultados superiores nas etapas de previsão em relação aos outros métodos. O modelo híbrido que combinou ARIMA com RNA, tem como principal dificuldade a estimação do modelo ARIMA adequado, sendo necessário várias etapas iterativas para obter o modelo apropriado. Já o modelo híbrido sistema *neurofuzzy*, tem por principal limitação a restrição a redes *feedforward*.

Os dois modelos híbridos mostraram ser eficientes na realização de previsão, em que o modelo híbrido ARIMA com RNA mostrou ser mais eficiente ao realizar previsões imediatas, apresentando previsões mais acuradas até o quarto mês de previsão. Já o sistema *neurofuzzy* apresentou ser uma opção válida para a abordagem ao problema de previsão de séries temporais, apresentando previsões mais acuradas para predições longas, sendo o modelo mais eficiente em termo de acurácia com relação aos métodos de estatística e de redes neurais.

Os trabalhos futuros desta pesquisa se orienta ao aperfeiçoamento do modelo *neurofuzzy*. Especificamente considerar a troca da t-norma e da função de pertinência e estudar sua influência na previsão. Adicionalmente, verificar o desempenho deste tipo de modelo em previsão de longo prazo acima de 36 meses.

Referências

- ABREU, T.; ARAÚJO, K. A.; LOPES, M. L.; LOTUFO, A. D. P. Metodologia híbrida utilizando os modelos arima e redes neurais artificiais para previsão de cargas elétricas. In: **Congresso Brasileiro de Automática. Campina Grande**. [S.l.: s.n.], 2012.
- AGUIRRE, L. A. **Introdução à identificação de sistemas—Técnicas lineares e não-lineares aplicadas a sistemas reais**. [S.l.]: Editora UFMG, 2004.
- ARNOLD, L. **Random dynamical systems**. [S.l.]: Springer Science & Business Media, 2013.
- AZEVEDO, R. R. de; SILVA, J. M. da; GATSIOS, R. C. Análise crítica dos modelos de previsão de série temporal com base no icms estadual. **Revista de Gestão, Finanças e Contabilidade**, v. 7, n. 1, p. 164–184, 2017.
- BARBOSA, E. C.; SÁFADI, T.; NASCIMENTO, M.; NASCIMENTO, A. C. C.; SILVA, C. H. O.; MANULI, R. **Metodologia amor Box & Jenkins para previsão de temperatura média mensal da cidade de Bauru (SP)**. 2015. 104–117 p.
- BEDREGAL, B. R. C.; TAKAHASHI, A. Interval t-norms as interval representations of t-norms. In: **FUZZ-IEEE**. [S.l.: s.n.], 2005. p. 909–914.
- BOX et al. **Time series analysis: forecasting and control**. [S.l.]: John Wiley & Sons, 2015.
- BRASIL. **Decreto Lei n 399, de 30 de outubro de 1938. Diário Oficial da República Federativa do Brasil**. [S.l.]: Poder Executivo Brasília, DF, 1938.
- BROOKS, C. **Introductory econometrics for finance**. [S.l.]: Cambridge university press, 2014.
- CAMPOS, R. J. **Previsão de séries temporais com aplicações a séries de consumo de energia elétrica**. Tese (Doutorado) — Dissertação de Mestrado, Programa de Pós-Graduação em Engenharia Elétrica da ... , 2008.
- CASSAR, V. B. **Direito do trabalho**. [S.l.]: Impetus Niterói, 2011.
- CHAHIN, A. **E-gov. br: a próxima revolução brasileira: eficiência, qualidade e democracia: o governo eletrônico no Brasil e no mundo**. [S.l.]: FINANCIAL TIMES BR, 2004.
- FAUSTINO, C. P. **Previsão de séries temporais via modelos baseados em regras**. Tese (Doutorado) — Dissertação de mestrado, UNIFEI, Itajubá MG, 2011.
- FAYYAD, U.; PIATETSKY-SHAPIO, G.; SMYTH, P. From data mining to knowledge discovery in databases. **AI magazine**, v. 17, n. 3, p. 37, 1996.
- FEDERAL, G. **Controladoria-Geral da União: Portal da Transparência**. 2004. <<http://transparencia.gov.br/>>. Accessed: 2017-01-05.
- FEDERAL, G. **SPED: sistema público de escrituração digital**. 2007. <<http://sped.rfb.gov.br/>>. Accessed: 2019-01-10.

- FERNANDES, L. G. L.; PORTUGAL, M. S.; NAVAUX, P. O. Previsão de séries de tempo: redes neurais artificiais e modelos estruturais. **Pesquisa e Planejamento Econômico**, v. 26, n. 2, p. 253–276, 1996.
- FERRER, F.; SANTOS, P. **e-Government: o governo eletrônico no Brasil**. [S.l.]: Editora Saraiva, 2004.
- HEBB, D. O. et al. **The organization of behavior**. [S.l.]: New York: Wiley, 1949.
- HODGKIN, A. L.; HUXLEY, A. F. A quantitative description of membrane current and its application to conduction and excitation in nerve. **The Journal of physiology**, Wiley Online Library, v. 117, n. 4, p. 500–544, 1952.
- HYNDMAN, R. J.; KHANDAKAR, Y. et al. **Automatic time series for forecasting: the forecast package for R**. [S.l.]: Monash University, Department of Econometrics and Business Statistics, 2007.
- HYNDMAN, R. J.; KOEHLER, A. B. Another look at measures of forecast accuracy. **International journal of forecasting**, Elsevier, v. 22, n. 4, p. 679–688, 2006.
- JANG, J.-S. Anfis: adaptive-network-based fuzzy inference system. **IEEE transactions on systems, man, and cybernetics**, IEEE, v. 23, n. 3, p. 665–685, 1993.
- JANG, J.-S. R.; SUN, C.-T.; MIZUTANI, E. Neuro-fuzzy and soft computing-a computational approach to learning and machine intelligence [book review]. **IEEE Transactions on automatic control**, IEEE, v. 42, n. 10, p. 1482–1484, 1997.
- KLEMENT, E. P.; MESIAR, R.; PAP, E. **Triangular norms**. [S.l.]: Springer Science & Business Media, 2013. v. 8.
- LLC, M. **Departamento Intersindical de Estatística e Estudos Socioeconômicos-DIEESE**. 1955. <<https://www.dieese.org.br>>. Accessed: 2019-01-10.
- LUENBERGER, D. G.; YE, Y. et al. **Linear and nonlinear programming**. [S.l.]: Springer, 1984. v. 2.
- LÜTKEPOHL, H.; XU, F. The role of the log transformation in forecasting economic variables. **Empirical Economics**, Springer, v. 42, n. 3, p. 619–638, 2012.
- MAKRIDAKIS, S.; WHEELWRIGHT, S. C.; MCGEE, V. E. **Metode dan aplikasi peramalan**. Jakarta: Erlangga, 1999.
- MARGARIDO, M. A.; ANEFALOS, L. C. Testes de raiz unitária e o software sas. **Agricultura em São Paulo**, v. 46, n. 2, p. 19–45, 1999.
- MARTINS, V. L. M.; WERNER, L. Comparação de previsões individuais e suas combinações: um estudo com séries industriais. **Produção, Porto Alegre**, SciELO Brasil, v. 24, n. 3, p. 618–627, 2014.
- MATTOSO, J. Mínimo salário mínimo. **Artigo publicado no Jornal do Economista**, v. 133, 1988.
- MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. **The bulletin of mathematical biophysics**, Springer, v. 5, n. 4, p. 115–133, 1943.

- MEDEIROS, R. Á. O. de; WINKELER, B. G. N.; VILLANUEVA, J. M. M.; RODRIGUEZ, Y. P. M.; TAVARES, E. C. Previsão de demanda a médio prazo aplicada em dados reais do sistema de distribuição: uma comparação entre rna e lógica fuzzy. **Revista Principia**, Directory of Open Access Journals, v. 1, n. 31, p. 75–85, 2016.
- MENGER, K. Statistical metrics. **Proceedings of the National Academy of Sciences**, National Acad Sciences, v. 28, n. 12, p. 535–537, 1942.
- MIRANDA, A. P.; CORONEL, D. A.; VIEIRA, K. M. Previsão do mercado futuro do café arábica utilizando redes neurais e métodos econométricos. **Estudos do CEPE**, p. 66–98, 2013.
- MORETTIN, P. A.; TOLOI, C. Análise de séries temporais. In: **Análise de séries temporais**. [S.l.: s.n.], 2006.
- MÜLLER, B.; REINHARDT, J.; STRICKLAND, M. T. **Neural networks: an introduction**. [S.l.]: Springer Science & Business Media, 2012.
- OLIVEIRA, H.; CALDEIRA, A.; MACHADO, M.; SOUZA, R.; TANSCHKEIT, R. Inteligência computacional aplicada a administração, economia e engenharia em matlab. **Rio de Janeiro, Thompson**, 2007.
- PEDRYCZ, W.; GOMIDE, F. **An introduction to fuzzy sets: analysis and design**. [S.l.]: Mit Press, 1998.
- PRIESTLEY, M. B. Spectral analysis and time series. Academic press, 1981.
- SAMOHYL, R. W.; ROCHA, R.; MATTOS, V. L. D. d. Utilização do método de holt-winters para previsão do leite entregue às indústrias catarinenses. 2001.
- SATO, R. C. Gerenciamento de doenças utilizando séries temporais com o modelo arima. **Einstein (São Paulo)**, Instituto Israelita de Ensino e Pesquisa Albert Einstein, 2013.
- SCHEFFER, D.; SOUZA, A. M.; ZANINI, R. R. Utilização de modelos arima para previsão da arrecadação de icms do estado do rio grande do sul. **XVII Simpósio de Pesquisa Operacional e Logística da Marinha-SPOLM. Anais... São Paulo**, 2014.
- SILVA, I. d.; SPATTI, D. H.; FLAUZINO, R. A. Redes neurais artificiais para engenharia e ciências aplicadas. **São Paulo: Artliber**, v. 23, n. 5, p. 33–111, 2010.
- SILVA, T. A. d. A. Previsão de cargas elétricas através de um modelo híbrido de regressão com redes neurais. Universidade Estadual Paulista (UNESP), 2012.
- SOUZA, A. P. d. et al. Análise comparativa de métodos de previsão de séries temporais através de modelos estatísticos e rede neural artificial. Pontifícia Universidade Católica de Goiás, 2012.
- SOUZA, J. A. V. et al. Aplicação de redes neuronais na previsão de vendas para retalho. 2011.
- SOUZA, F. J. V. d.; MELO, M. M. D. d.; ARAÚJO, F. R. d.; SILVA, M. C. d.; ARAÚJO, A. O. Práticas de governança eletrônica: um estudo nos sítios das capitais brasileiras. 2013.

TEAM, R. C. et al. R: A language and environment for statistical computing. Citeseer, 2013.

TEIXEIRA, R. d. A.; CORRÊA, M.; ASSIS, J. d. O. Redes neuro-fuzzy: algoritmo de treinamento e aplicações. **Principium Online: Iniciação Científica no Unileste-MG, Coronel Fabriciano**, v. 1, n. 3, p. 152–168, 2007.

UESC, B. M.; UESC, G. S.; PIRES, M. M. Custo da cesta básica oficial: levantamento e análise dos preços em nível de varejo nas cidades de ilhéus e itabuna.

VARFIS, A.; VERSINO, C. **Univariate Economic Time Series Forecasting**. [S.l.]: Cambridge University Press, 1990.