

CAPÍTULO 3

ANÁLISE DE SÉRIES TEMPORAIS, POR MEIO DE LINGUAGEM R, E PREVISÃO DE ACIDENTES AUTOMOBILÍSTICOS APLICANDO DE ESTATÍSTICA E REDES NEURAS ARTIFICIAIS

Data de aceite: 03/04/2023

Márcio Mendonça

Universidade Tecnológica Federal do
Paraná
Programa de Pós-Graduação em
Engenharia Mecânica (PPGEM-CP)
UTFPR -CP
<http://lattes.cnpq.br/5415046018018708>

Francisco de Assis Scannavino Junior

Universidade Tecnológica Federal do
Paraná
Departamento Acadêmico de Engenharia
Elétrica UTFPR - Cornélio Procópio – PR
<http://lattes.cnpq.br/4513330681918118>

Kleber Romero Felizardo

Universidade Tecnológica Federal do
Paraná
Departamento Acadêmico de Engenharia
Elétrica UTFPR - Cornélio Procópio – PR
<http://lattes.cnpq.br/6914500968214052>

Kazuyochi Ota Junior

Mestrando - PPGEM-CP - Programa de
Pós-Graduação em Engenharia Mecânica
Cornélio Procópio – PR
<http://lattes.cnpq.br/3845751794448092>

Andressa Haiduk

Dimension Engenharia - Ponta Grossa –
PR
<http://lattes.cnpq.br/2786786167224165>

José Augusto Fabri

Universidade Tecnológica Federal do
Paraná Departamento Acadêmico
de Computação (DACOM) - Cornélio
Procópio – PR
<http://lattes.cnpq.br/1834856723867705>

Marcio Aurélio Furtado Montezuma

Universidade Tecnológica Federal do
Paraná - Departamento Engenharia
Mecânica (DAMEC) - Cornélio Procópio
– PR
<http://lattes.cnpq.br/2487283169795744>

Wagner Fontes Godoy

Universidade Tecnológica Federal do
Paraná Departamento Acadêmico de
Engenharia Elétrica (DAELE) - Cornélio
Procópio – PR
<http://lattes.cnpq.br/7337482631688459>

Henrique Cavalieri Agonilha

Universidade Filadélfia (Unifil) – Londrina
– PR
<http://lattes.cnpq.br/9845468923141329>

Luiz Henrique Geromel

IFSP Instituto Federal de São Paulo
-Campus Piracicaba - SP
<http://lattes.cnpq.br/3945660888230811>

Matheus Gil Bovolenta

Acadêmico - Universidade Tecnológica Federal do
Paraná Departamento Acadêmico de Engenharia Elétrica (DAELE)
<http://lattes.cnpq.br/1518815195539638>

Emanuel Ignacio Garcia

Acadêmico - Universidade Tecnológica Federal do
Paraná Departamento Acadêmico de Engenharia Elétrica (DAELE)
<http://lattes.cnpq.br/8501809850590859>

Lucas Botoni de Souza

Mestre - PPGEM-CP - Programa de Pós-Graduação
em Engenharia Mecânica Cornélio Procópio – PR
<http://lattes.cnpq.br/5938489268359300>

Gilberto Mitsuo Suzuki Trancolin

Mestrando - PPGEM-CP - Programa de Pós-Graduação
em Engenharia Mecânica Cornélio Procópio – PR
<http://lattes.cnpq.br/3352701154826935>

Fabio Rodrigo Milanez

Faculdades da Indústria - Senai Londrina
Departamento Engenharia Elétrica

RESUMO: Este artigo propõe uma análise de uma série temporal, a qual tem uma tendência, como será visto nos resultados estatísticos e nos resultados de uma análise de acidentes com uma base de dados nativa da linguagem R. A priori a decomposição da série não apresenta uma sazonalidade bem redefinida uma curva irregular e alguns resíduos. O que sugere pelo menos dois métodos de análise, um estatístico por meio de regressão, e outro utilizando Redes Neurais Artificiais que a própria linguagem define número de neurônios, entradas recorrentes. Além disso, utilizando Linguagem R foi empregada na análise da qualidade da produção por meio de algumas métricas essenciais nesse tipo de análise para poder comparar os dois modelos. Um parâmetro importante, está no fato de métodos estatísticos clássicos, além da curva de previsão, apresentam com a tendência apresentada pela série com valores máximos e mínimos com normalmente 95% de confiança, enquanto as redes neurais artificiais sou apresentam uma proposta de geração futura. Sugere-se pela base nativa da linguagem R, os dados são reais e foram separados em dois grupos um de teste para treino da rede neural e ajuste do modelo de regressão. Para a posteriori comparar as duas evoluções. De um modo geral, esse artigo pretende fundamentar e sugerir a aplicação da técnica na análise temporal aplicada em engenharia de segurança com uma análise dos resultados a partir das métricas clássicas. E finalmente, o trabalho se encerra com uma conclusão e endereça futuros trabalhos.

PALAVRAS-CHAVE: Séries Temporais, análise e previsão e análise de lucros, histórico de dados reais.

TIME SERIES ANALYSIS, THROUGH R LANGUAGE, AND PREDICTION OF AUTOMOBILE ACCIDENTS APPLYING STATISTICS AND ARTIFICIAL NEURAL NETWORKS

ABSTRACT: This article proposes an analysis of a time series, which has a trend, as will be seen in the statistical results and the results of an analysis of accidents with a native database of the R language. A priori, the decomposition of the series does not show seasonality well redefined an irregular curve and some residues. This suggests at least two methods of analysis, one statistical through regression and another using Artificial Neural Networks, that the language defines the number of neurons and recurrent inputs. In addition, using the R language was used to analyse the quality of production through some essential metrics in this type of analysis to be able to compare the two models. An important parameter is the fact that classic statistical methods, in addition to the forecast curve, show the trend shown by the series with maximum and minimum values with average of 95% confidence. At the same time, artificial neural networks do not present a proposal for a future generation. It is suggested by the native base of the R language that the data are real and were separated into two test groups, one for training the neural network and adjusting the regression model. For a posteriori, compare the two evolutions. In general, this article intends to substantiate and suggest applying the technique applied in security engineering with an analysis of the results from the classic metrics. And finally, the work ends with a conclusion and addresses future work.

KEYWORDS: Time series, forecasting, and analysis of gain, accurate historical data.

1 | INTRODUÇÃO

A perspectiva de antever o futuro sempre encantou a humanidade. Saber o que vai acontecer antes mesmo que os primeiros sinais apareçam podem fornecer melhor uso dos efeitos benéficos ou até mesmo maléficos, os quais devemos tomar ações de controle para eventos futuros positivos, ou uma preparação prévia de possíveis efeitos adversos.

Possivelmente, mais importante que reconhecer os resultados é reconhecer o que pode interferir, favorável ou desfavoravelmente, nos processos contínuos, para permitir o planejamento (SOCIETY, 2016). Neste contexto, análise de séries temporais tem inúmeras aplicações em áreas financeiras (ZHANG; LI; YU, 2009).

Séries temporais tem sido utilizadas de forma significativa para geração e consumo de energia com até mais de um método, como será abordado nessa investigação científica devido a importância do tema, como por ser conferido no trabalho de (DEB et al., 2017), e aprendizado de máquina é uma das ferramentas usadas com frequência, como nesse trabalho, como pode ser conferido no trabalho de (PETRI et al., 2014).

De um modo geral, uma série temporal é uma sequência de realizações (observações) de uma variável ao longo do tempo. Dito de outra forma, é uma sequência de pontos (dados numéricos) em ordem sucessiva, geralmente ocorrendo em intervalos uniformes. Portanto, uma série temporal é uma sequência de números coletados em intervalos regulares durante

um período de tempo (NUSRATULLAH et al., 2015).

De acordo com a literatura, o parágrafo anterior a uma série temporal tem componentes e padrões, dentre esses podemos destacar elementos que serão explorados em uma análise: Tendência, Sazonalidade, Ciclo e Erro (restante). Esses elementos ficarão clarificados na análise na seção de desenvolvimento dessa investigação científica. O ciclo é um caso a parte de uma série temporal, pois, por não ter um comportamento definido, não é trivial. Entretanto, a tendência, a sazonalidade e, finalmente, o erro, são os dados que sobram após a decomposição da sazonalidade e da tendência das séries temporais.

A tendência pode subir ou descer, linearmente ou como uma exponencial. Entretanto, a falta da tendência, ou muito curta durante a série a torna não estacionária. Já a sazonalidade são padrões que ocorrem de forma recorrente na série com formas muito similares. Quando esses não são recorrentes, os ciclos aumentam ou diminuem na frequência na série, mas não em padrões, como no caso da sazonalidade (WARREN LIAO, 2005).

Em estatística, econometria, matemática aplicada e processamento de sinais, uma série temporal é uma coleção de observações feitas sequencialmente ao longo do tempo. Em modelos de regressão linear com dados *cross-section* a ordem das observações é irrelevante para a análise, em séries temporais a ordem dos dados é fundamental. Uma característica muito importante deste tipo de dados é que as observações vizinhas são dependentes e o interesse é analisar e modelar essa dependência (ANTUNES; CARDOSO, 2015). Um exemplo clássico da literatura, de modo específico uma biblioteca da linguagem R, é o lucro trimestral da *Johnson Jhonson* de 1960 até 1980 como mostra a Figura 1.

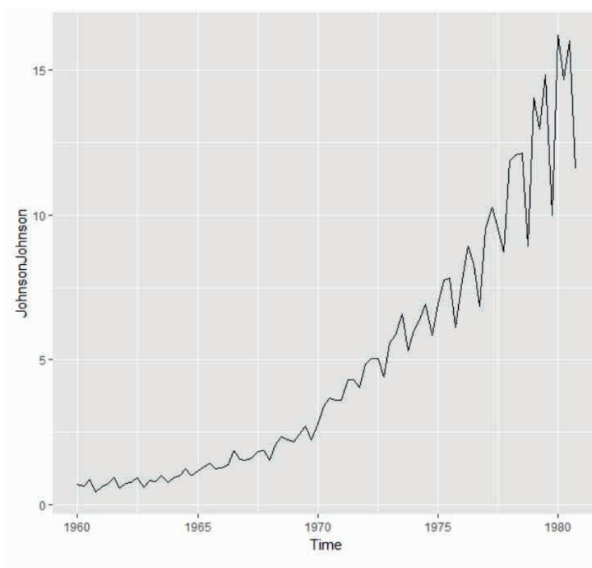


Figura 1: Lucro Trimestral da JhonsonJhonson de 1960 a 1980.

Fonte: adaptado de ZHANG; LI; YU, 2009.

Para uma análise mais acurada na série apresentada na figura 1 seria necessário decompor a mesma, entretanto já é possível observar uma tendência de crescimento em praticamente todo o período da mesma e uma aparente sazonalidade próximo aos anos 80.

As séries temporais existem nas mais variadas áreas de aplicação, como: finanças, marketing, economia, seguros, demografia, ciências sociais, meteorologia, energia, epidemiologia, entre outras (NUSRATULLAH et al., 2015). Outra característica importante é que, nesse exemplo e nos resultados dessa pesquisa, a série temporal é invariável, entretanto, existem séries temporais multivariáveis (SOCIETY, 2016)(PEDRYCZ; JASTRZEBSKA; HOMENDA, 2016).

O objetivo e a motivação desse trabalho estão na inferência dos lucros de uma empresa de transporte de carga com a interferência da Pandemia. A priori, foi constatado que o aumento dos lucros da empresa deu-se devido ao transporte de alguns insumos imprescindíveis, adubos e fertilizantes. Entretanto, decidiu-se fazer uma análise mais acurada por meio de séries temporais com histórico de dados antes da pandemia e comparar com os resultados no período, os quais pareceriam ser promissores. Por simplicidade, utilizou-se a planilha do Excel, que faz uso de métodos matemáticos como ARIMA.

Entretanto, Redes Neurais Artificiais (RNA) (DA SILVA; SPATTI; FLAUZINO, 2010), como também outras técnicas computacionais inteligentes como *Fuzzy Cognitive Maps* (FCM) ou Mapas Cognitivos Fuzzy (PEDRYCZ; JASTRZEBSKA; HOMENDA, 2016). Observa-se que é possível empregar Séries Temporais por meio de aplicação de Técnicas Computacionais Inteligentes, porém, esse não é o escopo desse trabalho. E, que os dados empregados são verídicos. A tabela 1 é explicativa e fornece e clarifica a logística da metodologia a ser discutida nessa pesquisa.

Passo 1	Série temporal acidentes da linguagem R
Passo 2	Análise da decomposição da série, tendência sazonalidade, resíduos
Passo 3	Separação dos dados de teste e treinamento
Passo 4	Comparativo das duas previsões com o grupo de teste
Passo 5	Análise da qualidade das previsões por meio das principais métricas

Tabela 1 – Ontologia da pesquisa realizada

Este trabalho está organizado da seguinte forma: na seção 2, fundamentos de séries temporais são brevemente apresentados, por se tratar de uma técnica não trivial e com muitas variações e técnicas, como por exemplo os resultados dessa pesquisa, que sugerem que a série é não temporal por apresentar sazonalidade e crescimento.

Já a seção 3 apresenta o desenvolvimento, ou seja, a ferramenta empregada e os dados históricos reais como supracitado da empresa e, ainda, mostra os resultados e as suas respectivas inferências. A seção 4 conclui e endereça futuros trabalhos. E, finalmente,

a seção 5, que se destina as referências utilizadas.

2 | FUNDAMENTOS DE SÉRIES TEMPORAIS

A análise de dados experimentais que foram observados em diferentes pontos com o tempo, leva a problemas novos e únicos na modelagem e inferência estatística. A correlação óbvia introduzida pela amostragem de pontos adjacentes com o tempo, pode restringir severamente a aplicabilidade de muitos métodos estatísticos convencionais tradicionalmente dependentes da suposição de que estas as observações são independentes e distribuídas de forma idêntica.

A abordagem sistemática pela qual alguém vai responder às questões matemáticas e estatísticas colocadas por essas correlações de tempo são comumente chamadas de análise de série temporal.

O impacto da análise de série temporal em aplicações científicas pode ser parcialmente documentado pela produção de uma lista abreviada dos diversos campos em que podem surgir problemas importantes de séries temporais. Por exemplo, muitas séries temporais ocorrem no campo da economia, onde estamos continuamente expostos a cotações do mercado de ações ou números mensais de desemprego. Cientistas sociais seguem séries de populações, como taxas de natalidade ou matrículas escolares. Um epidemiologista pode estar interessado no número de casos de gripe observados durante algum período. Na medicina, medições de pressão arterial rastreadas ao longo do tempo podem ser úteis para avaliar medicamentos usados no tratamento da hipertensão. Imagens de ressonância magnética de padrões de séries temporais de ondas cerebrais podem ser usadas para estudar como o cérebro reage a certos estímulos sob várias condições. Em um ambiente cada vez mais competitivo, em especial após o conceito das indústrias 4.0. produzir a quantidade necessária é um dos fatores fundamentais para manter a sustentabilidade da empresa. Ao se tratar de previsão de crescimento de vendas, consumo de água, de faturamento ou de do número de automóveis, os benefícios vão além da previsibilidade para o setor automobilístico, por exemplo, pois pode auxiliar também aos setores supracitados e os públicos em tomadas tomada de decisão.

Em outras palavras, A perspectiva de antever o futuro sempre encantou a humanidade. Saber o que vai acontecer antes mesmo de os primeiros sinais se manifestarem pode propiciar melhor aproveitamento dos efeitos benéficos de eventos futuros ou até mesmo uma preparação antecipada de eventuais efeitos adversos (ANTUNES; CARDOSO, 2015).

3 | DESENVOLVIMENTO

Duas técnicas distintas foram empregadas para analisar e comparar as previsões de acidentes com morte automobilística, o que nessa base não distingue se é em rodovia

ou em cidades nos Estados Unidos, base de dados da linguagem R.

Dentre as técnicas a primeira, após a análise da decomposição, a seguir. O método de regressão linear R.L., mais simplista comparado ao aprendizado de máquinas, por meio de redes neurais artificiais R.N.A.

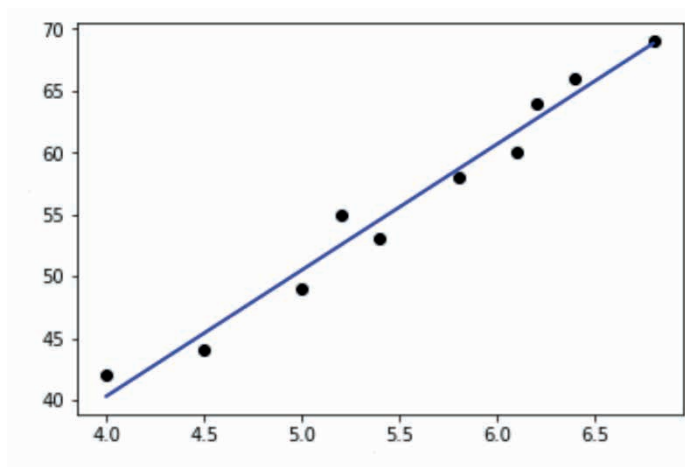


Figura 2: Regressão linear clássica R.L..

A figura 2 mostra uma regressão linear clássica, na qual uma reta busca achar um ponto médio dentre os pontos do experimento, os métodos dos mínimos quadrados são usados na literatura para tal ajuste.

Já para a aplicação das redes neurais consiste em definir a topologia da rede, o número de neurônios na entrada, o número de neurônios com atraso que irão variar de acordo com a complexidade da série temporal, caso a casa. Nesse trabalho a linguem R já define automaticamente esses parâmetros. A Figura 3 mostra um modelo clássico dessa ferramenta. E, não é escopo do trabalho detalhar as ferramentas, e sim aplicar e discorrer sobre os resultados

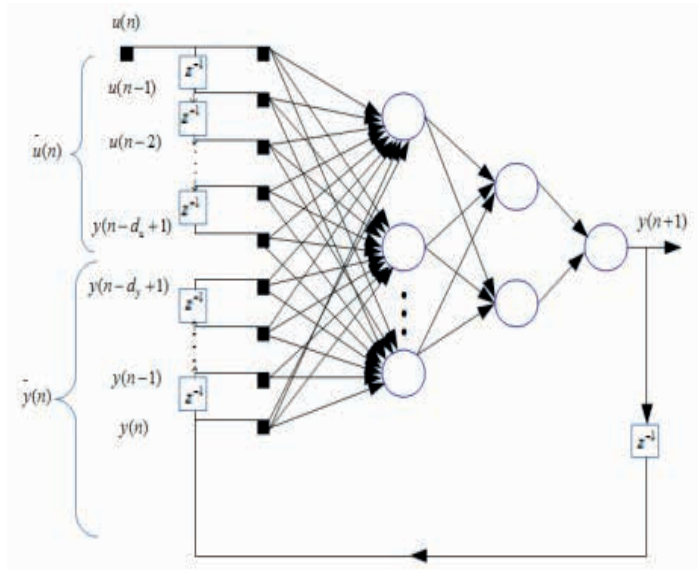


Figura 3: Previsão por meio de redes neurais artificiais R.N.A.

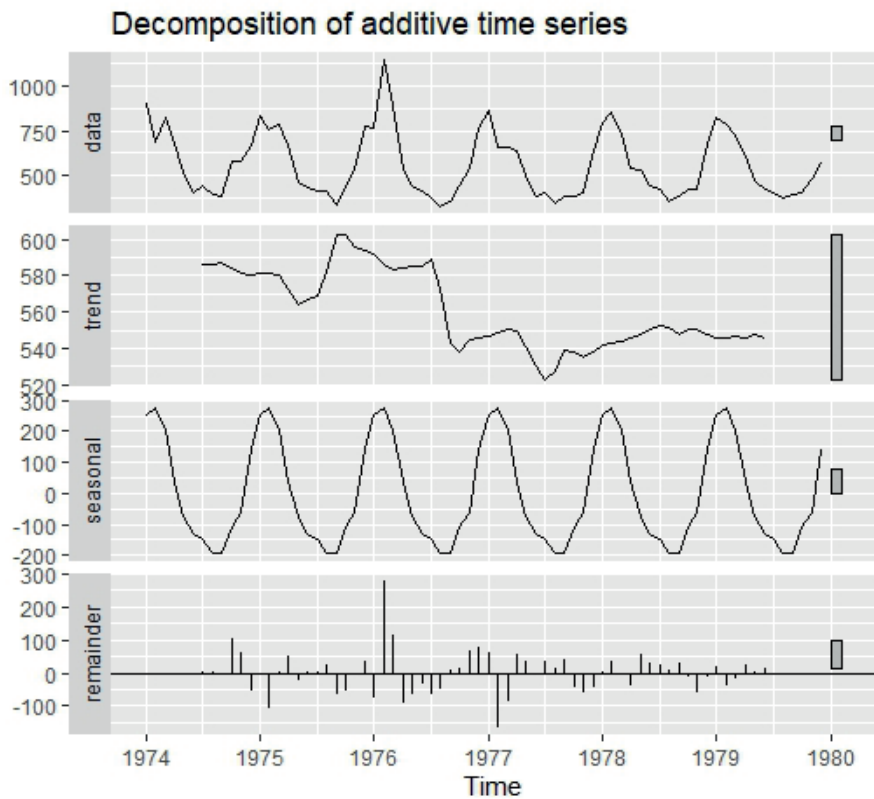


Figura 4: Elementos análise temporal.

Algumas considerações para diminuir acidentes dessa magnitude em países de modo geral, em especial para o Brasil.

O trânsito brasileiro é regulamentado pela Lei nº 9.503, de 23 de setembro de 1997: **o Código de Trânsito Brasileiro (CTB)**. Algumas sugestões podem ser elencadas para atenuar esse problema de cunho social severo.

Algumas dicas podem ser aplicadas para ao menos evitar colisões no trânsito

1. Obedeça à distância segura. ...
2. Utilize setas e pisca alerta. ...
3. Respeite as faixas horizontais. ...
4. Fique atento às sinalizações. ...
5. Controle a sua velocidade. ...
6. Descanse sempre que precisar. ...
7. Faça revisão no carro periodicamente. ...
8. Conheça procedimentos de segurança.
9. Utilize cinto de segurança, esse último item atenuou muito nos EUA quando foi obrigatório.

As tabelas 2 e 3 mostram as métricas que quantificam a qualidade da predição. Nesse trabalho, as métricas foram aplicadas nos testes, e quanto menor o valor da métrica melhor é a qualidade da predição.

A rigor as métricas empregadas foram ME. *Mean Error*, erro médio, a R.N.A. teve um resultado significativamente menor que a regressão linear, como foram encontrados valores negativos, nesse aspecto a R.L. teve melhor resultado (mais negativo, menor).

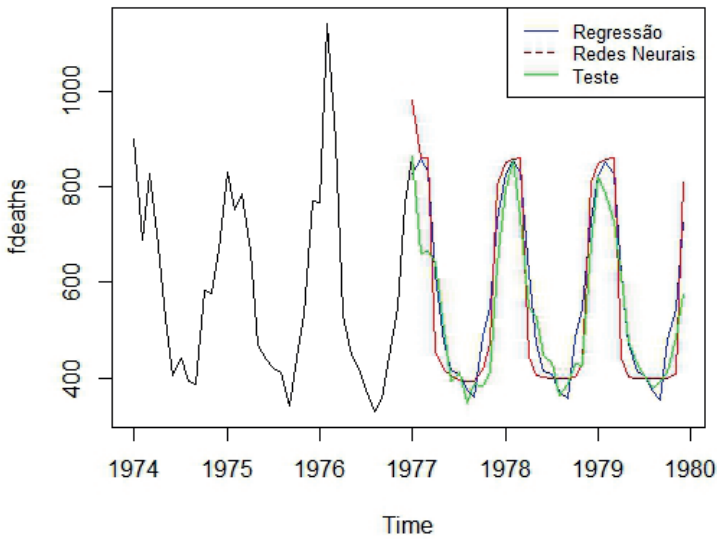


Figura 5:Decomposição da Série temporal.

Fonte: Autoria Própria.

	ME	RMSE	MAE	MPE	MAPE
Training set	0.0000	70.5901	47.1296	-0.9287	7.5113
Test set	-36.1944	73.9892	54.4167	-6.4682	10.3276

Tabela 2 - Resultados para o método da regressão

	ME	RMSE	MAE	MPE	MAPE
Training set	0.03362	48.3835	34.7471	-0.9832	6.4230
Test set	-19.8185	99.7224	74.6505	-2.5178	12.8565

Tabela 3 - Resultados para o método das RNA

Já a métrica RMSE *root mean square error*, erro médio quadrado R.L. teve valor menor que a R.N.A., o que confere melhor desempenho também. Já a métrica MAE *mean Absolute error*, teve menor resultado para a técnica de R.L. A métrica MPE *mean percentage error*, média percentual de erro a técnica R.L. também teve resultado menor; assim como a MAPE *mean Absolute percentagem error*, erro absoluto percentual foi menor.

Desse modo, apesar do método de regressão linear ser mais simples obteve resultados melhores do que a técnica de aprendizado de máquina, em especial redes neurais artificiais nos dados de teste, com as principais métricas de desempenho, esse resultado já foi sugerido com a comparação das curvas em uma avaliação visual, na qual a curva da R.L. se apresentou mais próxima da curva real de teste.

De acordo com o parágrafo anterior duas considerações devem ser feitas. Primeiramente, não se pode generalizar que a predição em R.L. é melhor do que a R.N.A.,

foi melhor sim de acordo com os resultados das métricas para esse caso específico. E o código de redes neurais em R é uma “caixa” preta e determina parâmetros como número de neurônios e de entradas em atraso de forma automática. Possivelmente, com uma batelada de opções poder-se-ia ter melhores resultados.

4 | CONCLUSÃO

Apesar dos resultados e análise, ainda serem iniciais, foram satisfatórios, pois as duas técnicas empregadas tiveram resultados visualmente diferentes, e até mesmo a análise quantificada foi possível por meio das métricas observar melhores resultados para o problema proposto, com sério impacto e importância social, que a regressão linear apresentou resultados mais convincentes, e como foi supracitado, essa conclusão não deve ser generalizada, e ficou demonstrado que a complexidade do método não pode garantir melhores resultados, como era de se esperar com técnicas de *machine learning*, redes neurais artificiais nessa pesquisa.

Futuros trabalhos endereçam aumento no intervalo de análise para tentar identificar com mais acurácia, com intervalos de treinos diferentes, e mais opções de configuração para técnicas de *machine learning*, em especial. Outra possível investigação futura é de se aplicar modelo ARIMA, por meio da função `Auto_Arima` da linguagem R, nesse caso a aplicação da suavização exponencial, parte do pressuposto que os dados mais recentes terem maior influência na série temporal, a qual poderá contribuir na melhoria do modelo.

REFERÊNCIAS

- ANTUNES, J. L. F.; CARDOSO, M. R. A. Uso da análise de séries temporais em estudos epidemiológicos. **Epidemiologia e Serviços de Saúde**, v. 24, n. 3, p. 565–576, set. 2015.
- DA SILVA, I. N.; SPATTI, D. H.; FLAUZINO, R. A. **Redes neurais artificiais para engenharia e ciências aplicadas: curso prático**. 1. ed. São Paulo: ARTLIBER, 2010.
- DEB, C. et al. A review on time series forecasting techniques for building energy consumption. **Renewable and Sustainable Energy Reviews**, v. 74, n. January, p. 902–924, 2017.
- NUSRATULLAH, K. et al. **Detecting changes in context using time series analysis of social network**. 2015 SAI Intelligent Systems Conference (IntelliSys). **Anais...IEEE**, nov. 2015Disponível em: <<http://ieeexplore.ieee.org/document/7361265/>>
- PEDRYCZ, W.; JASTRZEBSKA, A.; HOMENDA, W. Design of fuzzy cognitive maps for modeling time series. **IEEE Transactions on Fuzzy Systems**, v. 24, n. 1, p. 120–130, 2016.
- PETRI, I. et al. A modular optimisation model for reducing energy consumption in large scale building facilities. **Renewable and Sustainable Energy Reviews**, v. 38, p. 990–1002, 2014.

SOCIETY, R. S. Review Author (s): M . Farrow Review by : M . Farrow Source : Journal of the Royal Statistical Society . Series C (Applied Statistics), Vol . 30 , No . 2 Published by : Wiley for the Royal Statistical Society Stable URL : <http://www.jstor.org/stable/2. v. 30, n. 2, p. 30–31, 2016>.

WARREN LIAO, T. Clustering of time series data - A survey. **Pattern Recognition**, v. 38, n. 11, p. 1857–1874, 2005.

ZHANG, H.; LI, W.; YU, Q. Multifractality of financial time series. **FBIE 2009 - 2009 International Conference on Future BioMedical Information Engineering**, n. i, p. 237–239, 2009.