

# Aumento de dados e suavização integrada para predição de séries temporais baseada em aprendizado de máquina

Fernando Alexandrino<sup>1,2</sup>, Carla Pacheco<sup>3</sup>, Diego Carvalho<sup>1</sup>, Eduardo Ogasawara<sup>1</sup>

<sup>1</sup>Centro Federal de Educação Tecnológica Celso Suckow da Fonseca - CEFET/RJ

<sup>2</sup>Instituto Federal de Educação, Ciência e Tecnologia de São Paulo - IFSP

<sup>3</sup>Pontifícia Universidade Católica do Rio de Janeiro - PUC-Rio

fernando.alexandrino@ifsp.edu.br, cpacheco@inf.puc-rio.br

d.carvalho@ieee.org, eogasawara@ieee.org

**Resumo.** A predição de séries temporais é uma tarefa importante da ciência de dados e possui aplicações práticas em diversos campos. Este artigo explora técnicas de aumento de dados e de suavização para melhorar a precisão e a robustez de modelos preditivos. Uma metodologia chamada de Augment+Smooth, que integra ambas as abordagens, é proposta para gerar novas instâncias a partir de termos recentes e bem comportados. Os experimentos demonstraram que ela resulta em melhorias significativas na previsão em um contexto extremo de baixa disponibilidade de dados. Portanto, o método proposto oferece uma base sólida para futuras pesquisas e aplicações na predição de séries temporais.

**Abstract.** Time series prediction is a crucial task in data science with practical applications across various fields. This paper explores data augmentation and smoothing techniques to enhance the accuracy and robustness of predictive models. We introduce Augment+Smooth, a novel method that integrates both data smoothing and augmentation to generate new instances from recent and well-behaved data. We conducted an experiment demonstrating that our approach significantly improves forecasting performance in contexts with extremely limited data availability. Therefore, the proposed method provides a solid framework for future research and applications in time series prediction.

## 1. Introdução

Uma série temporal (ST) é uma sequência ordenada de observações de um fenômeno ao longo do tempo, como o preço de uma ação, a temperatura de determinada região ou o deslocamento de um objeto. O estudo e análise das STs tem o objetivo de identificar padrões, tendências e comportamentos nos dados que as representam. Uma destas tarefas é a predição, que consiste em usar toda a informação disponível até um dado período para prever o valor da série no(s) período(s) seguinte(s), tendo aplicações nas mais diversas áreas da ciência, da indústria ou da economia [Box et al., 2015].

Tipicamente, o processo de predição de STs é feito em duas etapas [Salles et al., 2022]: (i) pré-processamento dos dados de entrada; e (ii) modelagem e predição em si. A primeira consiste em identificar características da série e aplicar transformações que podem ser necessárias pelos modelos preditivos ou que tendem a torná-los mais efetivos.

Por exemplo, alguns modelos assumem que a ST é estacionária, normalmente flutuando em torno de uma média constante. Embora essa propriedade seja de difícil verificação nos fenômenos do mundo real, algumas abordagens podem ser empregadas para transformar a série original em outra mais “comportada”, como a suavização. O processo de suavizar a ST atenua flutuações e ajusta componentes sazonais ou cíclicos, facilitando a identificação de tendências e comportamentos ao longo do tempo e aumentando a precisão dos modelos preditivos [Hyndman and Athanasopoulos, 2018]. Além da suavização, a etapa de pré-processamento inclui a normalização dos dados, que tem o propósito de convertê-los para uma escala comum antes de serem utilizados.

A segunda etapa da predição de STs consiste na modelagem através de modelos estatísticos ou de aprendizado de máquina (AM) para prever observações futuras com base nos dados históricos. No primeiro caso, destaca-se o uso do modelo ARIMA (*Autoregressive Integrated Moving Average*) [Box et al., 2015, p. 19], que, embora seja altamente empregado, assume dependência temporal linear [Zhang, 2003]. Já o grupo de AM inclui uma vasta gama de modelos, com destaque para as redes neurais artificiais do tipo MLP (*Multi Layer Perceptron*) [Rumelhart et al., 1986], principalmente pela sua capacidade em lidar com padrões complexos [Zhang, 2003]. Em ambos os casos, um aspecto relevante a ser considerado é o volume disponível de dados que, quando insuficiente, dificulta o treinamento dos modelos e pode levá-los ao *overfitting*. Nestas situações, o aumento de dados (AD) é uma estratégia eficaz para melhorar o desempenho dos modelos de AM em previsões e análises de STs [Iglesias et al., 2023]. Ao usar transformações nos dados existentes, como inversão, adição de ruído e mudança de magnitude, aumenta-se a diversidade e a capacidade dos modelos em capturar padrões e variações.

Este trabalho apresenta uma metodologia para a combinação de AD e suavização na predição de STs, chamada de *Augment+Smooth*, oferecendo uma abordagem poderosa para modelos preditivos de AM. Enquanto o aumento de dados gera novas instâncias, enriquecendo as informações de treinamento, a suavização reduz ruídos e destaca padrões. A relevância desta proposta está em aprimorar a precisão e a robustez dos modelos de AM, capazes de fornecer previsões mais confiáveis, especialmente em contextos com poucos dados. A fim de verificar a eficácia da metodologia, foi realizado um experimento para a predição do consumo mundial de pesticidas, inspirado em Capistrano et al. [2023], no qual as STs possuem apenas 31 observações. Os resultados do *Augment+Smooth* revelaram melhora do desempenho preditivo médio de um modelo MLP com normalização adaptativa exponencial [Ogasawara et al., 2010] de aproximadamente 29% quando comparado às séries originais e de 14% quando empregado apenas o aumento de dados.

Além desta introdução, a pesquisa está organizada em mais cinco seções. As Seções 2 e 3 apresentam, em linhas gerais, as principais técnicas de aumento de dados e de suavização de STs, respectivamente. A Seção 4 detalha a metodologia proposta. Na Seção 5, são explicados os passos do experimento e os seus resultados. Finalmente, a Seção 6 destaca as conclusões do estudo e fornece perspectivas para trabalhos futuros.

## 2. Aumento de dados

O AD é uma abordagem que visa aumentar o tamanho e a diversidade de um conjunto de dados de STs através da criação de novas instâncias por meio de transformações ou modificações dos dados originais [Iglesias et al., 2023]. Essa técnica pode superar desafios

associados a pequenos conjuntos de dados e corroborar a robustez dos modelos. Entre as principais estratégias de AD, destacam-se [Iglesias et al., 2023]:

**Flip.** A técnica Flip (ou Rotation, para séries multivariadas) inverte os sinais de treinamento [Um et al., 2017], fornecendo amostras bastante diversificadas entre si, o que aumenta a capacidade de generalização do modelo preditivo.

**Jitter.** Outro método popular é Jitter, que adiciona ruído aos dados originais [Bishop, 1995]. O objetivo é aumentar a robustez do modelo contra flutuações e imprecisões na ST. Geralmente, o ruído gaussiano é utilizado [Iglesias et al., 2023], de maneira que a média corresponde à magnitude do ruído e o desvio padrão à deformação.

**Awareness.** A técnica Awareness funciona de forma semelhante ao Jitter, mas enfatiza os dados recentes. Sendo assim, observações antigas tendem a ser esquecidas, com sua influência diminuindo exponencialmente ao longo do tempo [Haykin, 2011]. Isso é particularmente útil no contexto de STs, onde informações recentes são mais relevantes.

**Scaling.** O Scaling consiste em alterar a magnitude dos dados, modificando seu intervalo (ou escala, daí o nome do método) ao mesmo tempo que mantém sua forma. Esse processo pode ser feito de maneira homogênea para toda a série ou empregando algum critério como o tempo ou a frequência das observações.

A Figura 1 exibe um exemplo ilustrativo destas técnicas. As subsequências geradas, representadas pelas linhas verdes, são introduzidas em janelas deslizantes. Vale ressaltar que o método escolhido depende do contexto específico da tarefa e da natureza dos dados em questão. Iglesias et al. [2023] discutem a aplicabilidade de cada abordagem.

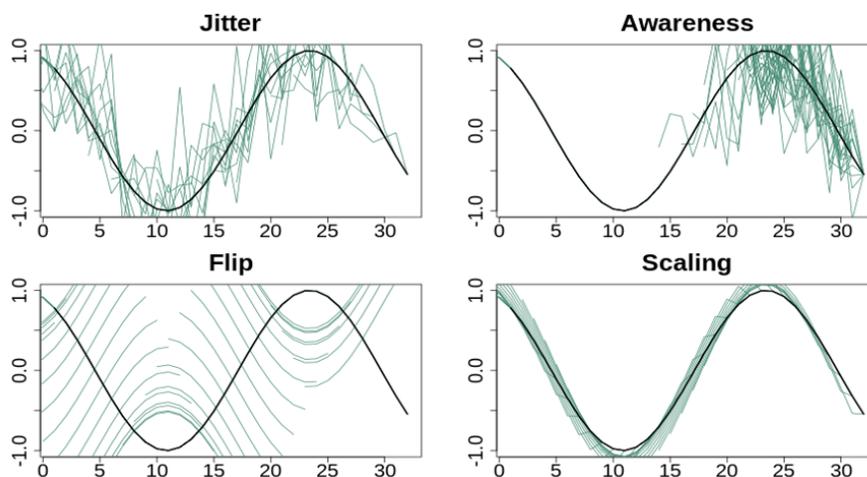


Figura 1. Aplicação das técnicas de AD (linhas verdes) a uma série cossenoidal

### 3. Suavização em STs

A suavização é utilizada para reduzir flutuações aleatórias em STs, facilitando a identificação de padrões, como tendências e sazonalidades. Ela não só melhora a clareza dos dados, mas também aumenta a precisão dos modelos preditivos, tornando-os mais robustos e confiáveis. As principais técnicas incluem [Hyndman and Athanasopoulos, 2018]:

**Suavização por Média Móvel.** A Suavização por Média Móvel (SMM) calcula a média das últimas  $k > 1$  observações, sendo útil em contextos de estacionariedade.

**Suavização Exponencial Simples.** A Suavização Exponencial Simples (SES) usa pesos que diminuem exponencialmente da observação mais recente para a mais antiga, sendo adequada para STs voláteis sem tendência ou sazonalidade.

**Suavização Exponencial Quadrática.** A Suavização Exponencial Quadrática (SEQ), também conhecida como Método de Tendência Linear de Holt, é uma extensão da SES que utiliza uma componente própria para capturar as tendências da ST.

**Suavização Lowess.** A Suavização Lowess (*Local Weighted Scatterplot Smoothing*) utiliza mínimos quadrados ponderados e um modelo polinomial de segunda ordem para suavizar os dados. Ela é adequada para padrões não lineares.

**Splines de Suavização Cúbica.** Os Splines de Suavização Cúbica (SSC) são uma técnica de suavização baseada em otimização [Hastie, 2017]. Considerando um conjunto de pontos  $\{x_i, y_i : i = 1, \dots, n\}$ , o *spline* cúbico é uma função de suavização composta por seções de polinômio cúbico, minimizando a curvatura total, i.e.,  $\int_{x_1}^{x_n} \{f''(x)\}^2 dx$ .

**Suavização por Kernel.** A Suavização por Kernel (KS) é uma técnica não paramétrica aplicada à estimativa de funções de densidade. Ela identifica estruturas em conjuntos de dados sem a necessidade de um modelo paramétrico específico [Wand and Jones, 1994]. O estimador de densidade de *kernel* univariado é dado por  $\hat{f}(x; h) = (nh)^{-1} \sum_{i=1}^n K\{(x - X_i)/h\}$ , onde  $K$  é a função kernel e  $h$  é a largura da banda.

Além destes métodos, convém destacar que a aplicação recursiva da normalização por diferenciação tende a produzir um resultado semelhante à suavização. Ao tomar as diferenças entre valores consecutivos da ST, são eliminadas as tendências de primeira ordem; repetindo o processo, elimina-se as tendências de segunda ordem; e assim sucessivamente até atingir a estacionariedade [Box et al., 2015, p. 160]. Como critério de parada, pode-se verificar a inexistência de *outliers* usando o intervalo interquartil (IQR).

A Figura 2 ilustra o efeito da suavização aplicada a uma ST. Diferentes abordagens de suavização podem ser adotadas, dependendo da natureza dos dados e do objetivo da análise (vide [Shumway and Stoffer, 2017] e [Hyndman and Athanasopoulos, 2018]).

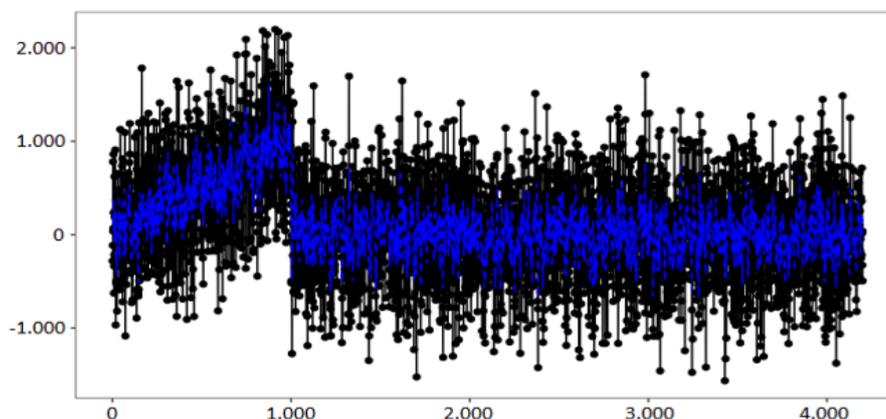
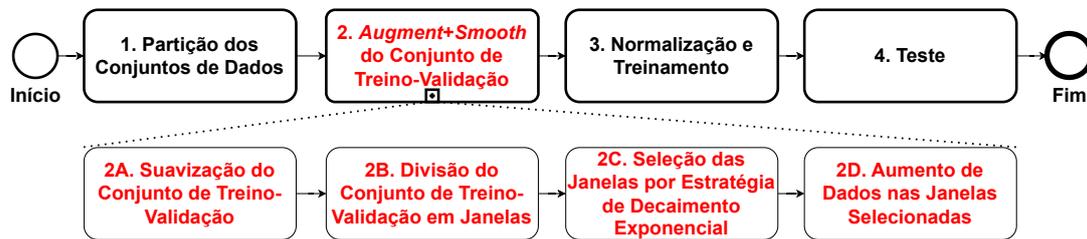


Figura 2. O efeito da suavização (linha azul) em uma ST

#### 4. Combinando aumento de dados com suavização

A metodologia proposta neste trabalho, chamada *Augment+Smooth*, combina técnicas de suavização com AD para melhorar a precisão e a robustez dos modelos preditivos de AM.

A Figura 3 apresenta o processo, cujos detalhes são explicados a seguir.



**Figura 3. Metodologia do Augment+Smooth**

**Passo 1: Partição dos Conjuntos de Dados.** No primeiro passo, a ST é dividida em dois conjuntos: (i) conjunto de treinamento-validação; e (ii) conjunto de teste. O primeiro compreende os dados de treinamento do modelo preditivo, assim como os dados de validação, que serão apresentados para o modelo treinado a fim de avaliar seu desempenho. Já o conjunto de teste contém os dados mais recentes e desconhecidos pelo modelo, utilizados para realizar as previsões e servir como base de mensuração e análise.

**Passo 2: Augment+Smooth.** Essa é a principal fase da metodologia e compreende:

- 2A. Suavização do Conjunto de Treinamento-Validação:** Suaviza toda a série de treinamento-validação para remover valores extremos utilizando algum dos métodos de suavização descritos na Seção 3, resultando em uma série mais bem comportada. A verificação de *outliers* pode ser feita pelo critério do IQR.
- 2B. Divisão em Janelas:** Divide o conjunto de treinamento-validação em janelas deslizantes de tamanho igual.
- 2C. Escolha de Janelas por Estratégia de Decaimento Exponencial:** As janelas são selecionadas para AD com base em uma estratégia de decaimento exponencial, onde as mais recentes têm uma maior probabilidade de serem escolhidas.
- 2D. Aumento de Dados nas Janelas Selecionadas:** Aplica uma determinada técnica de AD apenas nas janelas selecionadas. No caso do Jitter, por exemplo, se uma janela for selecionada mais de uma vez, ruídos diferentes serão adicionados, gerando dados artificiais distintos e enriquecendo o conjunto de treinamento.

**Passo 3: Normalização e Treinamento.** Após o *Augment+Smooth*, prepara-se o conjunto de treinamento com a normalização. As principais abordagens são: a normalização *minmax*, que converte a escala dos dados para, usualmente, o intervalo  $[0, 1]$ , e pode ser feita tanto para a série inteira quanto para cada janela; a diferenciação, que computa as diferenças entre valores consecutivos na ST, conforme mencionado anteriormente; e a normalização adaptativa, que envolve o cálculo da média móvel de cada janela, a remoção de *outliers* em relação a essas médias e, por fim, a normalização *minmax* dos dados restantes em cada janela. A normalização adaptativa exponencial [Ogasawara et al., 2010] é um caso particular desta última que utiliza a média móvel exponencial, criando uma sequência de pesos que decrescem exponencialmente do mais recente para o mais antigo.

Após a normalização, o modelo é treinado e os hiperparâmetros são ajustados.

**Passo 4: Teste.** No último passo, o desempenho do modelo preditivo é avaliado no conjunto de teste. É fundamental que os dados deste conjunto utilizem os mesmos parâmetros aplicados no treinamento, bem como a mesma técnica de normalização.

A Figura 4 ilustra a aplicação do *Augment+Smooth* ao exemplo utilizado na Seção 2. Ao remover *outliers* e enfatizar dados recentes bem comportados, a metodologia aumenta a capacidade dos modelos de capturar variações e adaptar-se a novas condições.

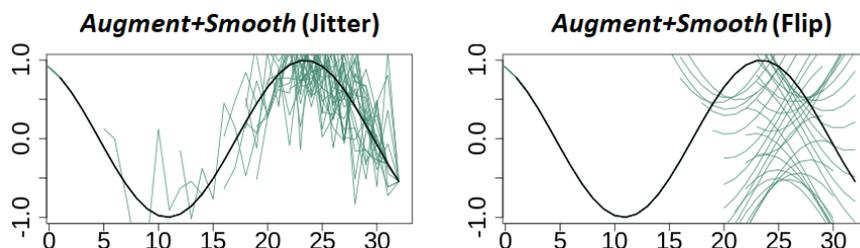


Figura 4. Aplicação do *Augment+Smooth* (em verde) a uma série cossenoidal

## 5. Experimento

Esta seção avalia a metodologia proposta em uma aplicação para a predição do consumo mundial de pesticidas, concentrando-se nos dez países com maior volume de uso destas substâncias. As STs deste *dataset* foram obtidas do FAOSTAT<sup>1</sup> e compreendem apenas 31 observações cada, referentes aos anos de 1990 a 2020, constituindo um cenário extremo de baixa disponibilidade de dados. Nenhum tipo de tratamento foi realizado.

Definiu-se o período de 1990 a 2015 como conjunto de treinamento-validação e os anos seguintes como conjunto de teste. As predições foram feitas um passo à frente usando a metodologia *rolling-origin* [Hyndman and Athanasopoulos, 2018], o que resultou em 50 cenários experimentais (cinco períodos para cada série). Além do *Augment+Smooth* (com Jitter e a abordagem de suavização por diferenciação), o experimento foi replicado usando apenas aumento de dados nas STs (com a técnica *Awareness*, pela semelhança com o *Augment+Smooth*) e também usando as séries originais, sem nenhum tipo de suavização ou AD. Em todos os casos foi empregada a normalização adaptativa exponencial devido à sua característica de enfatizar os dados mais recentes, tornando a comparação mais equilibrada, e o MLP como modelo preditivo. A implementação foi desenvolvida em R<sup>®</sup> através do pacote DAL Toolbox<sup>2</sup> e os hiperparâmetros empregados foram: *sliding window* = {8}, *input size* = {3, ..., 7}, *nhid* = {1, ..., 10}, *decay* = {0.0, 0.1, ..., 1.0} e *maxit* = {1.000}, cada qual submetido a uma busca sistemática por meio de *grid search*. Os modelos finais resultaram das melhores combinações identificadas durante esse processo, não sendo escopo deste trabalho a investigação dos impactos individuais de cada hiperparâmetro no desempenho das previsões. Por fim, os resultados foram submetidos ao teste de significância estatística de Wilcoxon [Mason et al., 2003].

A Tabela 1 apresenta os resultados consolidados do conjunto de teste usando a métrica SMAPE (*Symmetric Mean Absolute Percentage Error*) para cada abordagem. Como é possível observar, as previsões usando o *Augment+Smooth* foram consistentemente melhores em relação às séries originais e ao *Awareness*, mesmo este último produzindo um efeito semelhante no que consiste ao AD. Isso acontece porque o aumento na metodologia proposta leva em consideração a série mais bem comportada, como consequência da etapa de suavização, fazendo com que o novo dado gerado seja relevante em termos de

<sup>1</sup><https://www.fao.org/faostat/en/>. Dados coletados em 22/03/2024.

<sup>2</sup><https://CRAN.R-project.org/package=daltoolbox>

informação e de probabilidade de ocorrência no fenômeno real. Por outro lado, ao utilizar apenas a técnica de AD, não há garantias de que a transformação aconteça sobre um dado não-outlier; assim, o dado sintético pode não ser tão relevante para o modelo preditivo.

**Tabela 1. SMAPE médio (%) por país**

País	Séries Originais	Séries Aumentadas	<i>Augment+Smooth</i>
Argentina	16,97 ± 1,82	13,12 ± 1,62	<b>4,72 ± 2,70</b>
Austrália	12,66 ± 4,59	12,30 ± 4,52	<b>12,11 ± 4,51</b>
Brasil	8,29 ± 0,57	12,96 ± 2,79	<b>8,11 ± 0,61</b>
Canadá	11,76 ± 0,68	<b>8,11 ± 1,34</b>	11,41 ± 1,10
China	14,93 ± 11,90	8,83 ± 2,81	<b>8,01 ± 3,15</b>
Estados Unidos	1,01 ± 3,32	0,66 ± 3,60	<b>0,56 ± 4,16</b>
França	24,87 ± 1,58	11,45 ± 5,13	<b>9,13 ± 1,16</b>
Índia	7,41 ± 1,63	11,05 ± 1,43	<b>5,40 ± 1,35</b>
Itália	3,37 ± 5,77	3,36 ± 5,43	<b>3,35 ± 4,48</b>
Rússia	21,67 ± 0,15	<b>19,84 ± 0,19</b>	24,49 ± 0,14
Média	12,29 ± 8,59	10,17 ± 6,09	<b>8,73 ± 6,86</b>

Na Tabela 2, foi realizado um teste pareado de Wilcoxon comparando os SMAPEs obtidos pelas previsões de cada abordagem. Dados os *p*-valores inferiores a 0,05 e os respectivos *effect sizes*, o desempenho global do *Augment+Smooth* em relação às séries originais e ao *Awareness* atingiu superioridade estatística de magnitude grande e moderada, respectivamente, demonstrando o potencial do método proposto. Em termos locais, apenas no caso da Itália houve empate técnico entre todas as abordagens, embora a variabilidade tenha sido menor utilizando o *Augment+Smooth*.

**Tabela 2. Análise de significância estatística**

Teste pareado	<i>p</i> -valor	<i>Effect Size</i>	Magnitude
<i>Augment+Smooth</i> x Séries Originais	0,0002536	0,518	Grande
<i>Augment+Smooth</i> x Séries Aumentadas	0,0131100	0,352	Moderada

## 6. Considerações finais

O sucesso das predições em séries temporais depende da qualidade e do volume disponível de dados. Ao explorar as principais técnicas tanto de suavização, que visa reduzir ruídos e destacar padrões importantes, quanto de aumento de dados, que busca reforçar a diversidade e a quantidade de dados de treinamento, este trabalho propôs a metodologia *Augment+Smooth*, que combina ambas as abordagens para elevar a precisão e a robustez dos modelos de previsão. Juntas, essas técnicas podem mitigar problemas de *overfitting* e melhorar a capacidade dos modelos de aprendizado de máquina em capturar variações e adaptar-se a novas condições, ainda que não tenham acontecido historicamente.

Através de um experimento inicial para a predição do consumo global de pesticidas, com séries de 31 observações, foi possível verificar a efetividade do *Augment+Smooth* em potencializar as previsões de um modelo MLP, alcançando resultados melhores quando comparados com as séries originais ou quando o aumento de dados foi realizado sem a suavização. Em termos de desempenho global, a metodologia proposta foi estatisticamente superior às demais abordagens, além de ter vencido em oito dos dez países analisados.

Pesquisas futuras podem expandir o domínio de aplicação para outros *datasets* e classificadores de desempenho, bem como investigar diferentes combinações de técnicas para cada fase do método, superando desafios associados a pequenos conjuntos de dados e promovendo a generalização eficaz dos modelos de aprendizado de máquina para cenários do mundo real.

## Agradecimentos

Os autores agradecem ao CNPq, CAPES e FAPERJ pelo apoio parcial a esta pesquisa.

## Referências

- Bishop, C. M. (1995). Training with Noise is Equivalent to Tikhonov Regularization. *Neural Computation*, 7(1):108–116.
- Box, G. E. P., Jenkins, G. M., Reinsel, G. C., and Ljung, G. M. (2015). *Time Series Analysis: Forecasting and Control*. John Wiley & Sons.
- Capistrano, B., Chen, L., Ribeiro, M., Pacheco, C., Lobosco, D., Quadros, J., Barreto, M. I., and Ogasawara, E. (2023). Desafios na Predição do Consumo de Pesticidas em Escala Global Usando Aprendizado de Máquina. In *Anais do Brazilian e-Science Workshop (BreSci)*, pages 33–38. SBC.
- Hastie, T. J. (2017). *Generalized Additive Models*. Routledge.
- Haykin, S. O. (2011). *Neural Networks and Learning Machines*. Pearson Education.
- Hyndman, R. J. and Athanasopoulos, G. (2018). *Forecasting: principles and practice*. OTexts.
- Iglesias, G., Talavera, E., González-Prieto, Á., Mozo, A., and Gómez-Canaval, S. (2023). Data Augmentation techniques in time series domain: a survey and taxonomy. *Neural Computing and Applications*, 35(14):10123 – 10145.
- Mason, R. L., Gunst, R. F., and Hess, J. L. (2003). *Statistical Design and Analysis of Experiments: With Applications to Engineering and Science*. John Wiley & Sons, 2nd. edition.
- Ogasawara, E., Martinez, L. C., De Oliveira, D., Zimbrão, G., Pappa, G. L., and Mattoso, M. (2010). Adaptive Normalization: A novel data normalization approach for non-stationary time series. In *Proceedings of the International Joint Conference on Neural Networks*.
- Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088):533 – 536.
- Salles, R., Pacitti, E., Bezerra, E., Porto, F., and Ogasawara, E. (2022). TSPred: A framework for nonstationary time series prediction. *Neurocomputing*, 467:197 – 202.
- Shumway, R. H. and Stoffer, D. S. (2017). *Time Series Analysis and Its Applications: With R Examples*. Springer.
- Um, T. T., Pfister, F. M., Pichler, D., Endo, S., Lang, M., Hirche, S., Fietzek, U., and Kulic, D. (2017). Data augmentation of wearable sensor data for Parkinson’s disease monitoring using convolutional neural networks. In *ICMI 2017 - Proceedings of the 19th ACM International Conference on Multimodal Interaction*, pages 216 – 220.
- Wand, M. P. and Jones, M. C. (1994). *Kernel Smoothing*. CRC Press.
- Zhang, G. P. (2003). Time series forecasting using a hybrid arima and neural network model. *Neurocomputing*, 50:159–175.