

# The impact of window size on univariate time series forecasting using machine learning

João David Freitas<sup>1</sup>, Caio Ponte<sup>1</sup>, Rafael Bomfim<sup>1</sup>, Carlos Caminha<sup>2</sup>

<sup>1</sup> Programa de Pós-Graduação em Informática Aplicada, Unifor, Brasil

joaodavidfreitasc@edu.unifor.br

caioponte@unifor.br

bomfim@unifor.br

<sup>2</sup> Universidade Federal do Ceará, Brasil

caminha@ufc.br

**Abstract.** In the task of modeling time series prediction problems, the window size ( $w$ ) is a hyperparameter that defines the amount of time units that will be present in each example to be fed into a learning model. This hyperparameter is particularly important due to the need to make the learning model understand both long-term and short-term trends, as well as seasonal patterns, without making it sensitive to random fluctuations. In this article, our aim is to understand the effect that the window size has on the results of machine learning algorithms in univariate time series prediction problems. To achieve this goal, we used 40 time series from two distinct domains, conducting trainings with variations in the window size using four types of machine learning algorithms: Bagging, Boosting, Stacking, and a Recurrent Neural Network architecture. As a result, we observed that increasing the window size can lead to an improvement in the evaluation metric values until reaching a stabilization scenario, where further increasing the window size does not yield better predictions. The aforementioned stabilization occurred in both studied domains only when  $w$  values exceeded 100 time steps. We also observed that Recurrent Neural Network architectures do not outperform ensemble models in several univariate time series prediction scenarios.

CCS Concepts: • **Computing methodologies** → **Machine learning algorithms**.

Keywords: Machine learning, Sliding Window, Time Series, window size

## 1. INTRODUÇÃO

A previsão de séries temporais é uma tarefa importante com aplicações em muitas áreas, incluindo economia, finanças, saúde, engenharia, ciências sociais e climatologia [De Gooijer and Hyndman 2006]. Séries temporais podem ser definidas como sequências de termos ordenados ao longo do tempo, geralmente em intervalos regulares, e podem fornecer informações valiosas sobre as tendências, padrões e comportamentos de sistemas dinâmicos ao longo do tempo. Ao estudar séries temporais, pode-se encontrar informações ocultas e prever comportamentos futuros, viabilizando a tomada de decisão embasada e o aprimoramento da eficiência e eficácia. Analisar essas séries também pode ajudar a identificar mudanças em padrões sazonais ou em eventos extremos, como picos ou quedas bruscas em uma série, o que pode ser valioso para a prevenção ou mitigação de riscos.

Apesar de todas essas aplicações, a modelagem de um problema de previsão de séries temporais pode apresentar desafios para compreender como os hiperparâmetros afetam as previsões. Além de considerar a definição do tamanho da janela, é necessário explorar uma variedade de hiperparâmetros do modelo adotado, tais como taxa de aprendizagem, número de camadas ocultas, tamanho dos filtros, dentre outros. Além disso, é importante definir adequadamente as estratégias de previsão em múltiplos passos [Taieb et al. 2012] e o tamanho do horizonte de previsão [Hamzaçebi et al. 2009]. Outra

---

Copyright©2023 Permission to copy without fee all or part of the material printed in KDMiLe is granted provided that the copies are not made or distributed for commercial advantage, and that notice is given that copying is by permission of the Sociedade Brasileira de Computação.

consideração crucial é escolher transformações adequadas para lidar com possíveis não-linearidades, sazonalidades, tendências e outras características da série temporal [Salles et al. 2019]. Ou seja, a modelagem de séries temporais requer uma análise cuidadosa de vários fatores para garantir previsões precisas e confiáveis.

Especificamente em relação ao tamanho da janela ( $w$ ), esse hiperparâmetro é importante na previsão de séries temporais, pois ele define a quantidade de unidades de tempo que estarão presentes em cada exemplo no processo de treinamento de um modelo e o mesmo influencia na precisão e estabilidade da previsão [Azlan et al. 2019]. Escolher um tamanho de janela adequado é fundamental para capturar tendências de longo e curto prazo e padrões sazonais, sem tornar o modelo sensível a flutuações aleatórias ou levar a um superajuste. Muitas vezes, a escolha ideal do tamanho da janela precisa levar em conta a complexidade da série temporal, o tamanho do conjunto de dados, modelo utilizado e a capacidade computacional disponível.

Apesar da importância da definição do tamanho da janela, não existem muitos artigos que tratam do estudo do real impacto do tamanho da janela na previsão de séries temporais, utilizando diferentes modelos. Nesse contexto, este artigo se propõe a estudar o impacto do tamanho da janela na previsão das séries temporais. Para isso foram utilizados dois conjuntos de dados reais, um do varejo e um de mobilidade urbana, com um total de 20 séries temporais cada, totalizando 120280 unidades de tempo para serem modeladas em um problema de previsão. A avaliação será realizada utilizando quatro classes de algoritmos de aprendizado de máquina: Um algoritmo de *Bagging*; um de *Boosting*; um de *Stacking*; e uma arquitetura de Rede Neural Recorrente.

Este artigo está organizado da seguinte forma: na Seção 2, será exposto o estado da arte relacionado a esta pesquisa; na Seção 3 será detalhada a metodologia da pesquisa; na Seção 4, serão apresentados os resultados dos testes realizados; e, por fim, na Seção 5, serão explicitadas as conclusões desta pesquisa e serão discutidos trabalhos futuros.

## 2. ESTADO DA ARTE

Uma série temporal é uma coleção de observações tomadas em ordem cronológica, e a abundância de dados neste formato impulsiona a realização de pesquisas em diversas áreas. Diversas tarefas podem ser realizadas utilizando séries temporais, tais como: detecção de anomalia [Munir et al. 2018], clusterização [Aghabozorgi et al. 2015], classificação [Fawaz et al. 2019], segmentação [Lovrić et al. 2014] e previsão [Lim and Zohren 2021].

Vários estudos têm analisado o impacto da seleção de características na previsão. [Huber and Stuckenschmidt 2020] analisou a previsão de produtos do varejo com o adicional de passar para os modelos datas especiais e feriados, também analisaram o efeito do aumento do horizonte de previsão, de forma *single-step* e *multi-step*. [Abolghasemi et al. 2020] desenvolveram um modelo para prever a demanda por produtos, que leva em consideração eventos sistemáticos, como um produto em promoção ou mudança no clima. Ao final dos testes, o modelo do artigo teve um desempenho superior ao adicionar as novas características.

Um desafio comum na previsão de séries temporais é o tamanho do horizonte de previsão, que muitas vezes se limita a apenas um passo à frente. Isso limita a aplicação dos modelos em situações que exigem planejamento a longo prazo. De acordo com [Cheng et al. 2006], as previsões de longo prazo tendem a ter maiores erros, pois o viés e a variância das previsões passadas afetam as previsões futuras, resultando em um erro acumulado.

Em um estudo sobre previsão de preços de ações, [Shynkevich et al. 2017] testaram a combinação do horizonte de previsão e do tamanho da janela utilizando três modelos. A previsão foi tratada como um problema de classificação para determinar se as ações iriam subir ou cair. Os resultados indicaram que o tamanho ideal da janela é próximo do tamanho do horizonte de previsão.

Alguns estudos investigaram o impacto do tamanho da janela na previsão de séries temporais. [Bergström and Hjelm 2019] estudaram o efeito do tamanho da janela na previsão do índice da bolsa americana S&P500 utilizando *LSTM* e descobriram que o tamanho ótimo da janela é de dez dias. Já [Liu et al. 2022], utilizaram o modelo *Transformer* para testar previsões com janelas pequenas e horizontes de previsão maiores em quatro séries temporais de temperatura de transformadores elétricos, e seus experimentos indicaram que menores janelas em conjunto com maior horizonte de previsão podem levar a melhores previsões. Além disso, [Ughi et al. 2023] comparou o desempenho de modelos *Transformer* com modelos mais simples (como MLP) testando quatro variações de tamanho de janela para seis séries de diferentes domínios.

O presente artigo avança nas pesquisas do estudo de séries temporais ao fazer experimentos com mais de uma série temporal, em dois domínios diferentes e com quantidades de anotações diferentes. Também avança ao estudar o impacto do tamanho da janela em um cenário mais desafiador, onde o horizonte de previsão é maior. Isso é particularmente importante devido grande parte dos artigos na literatura realizarem experimentos com apenas um anotação a frente (*One-Step*).

### 3. METODOLOGIA

Esta seção detalha os conjuntos de dados utilizados e explica a modelagem dos dados realizada para conduzir esta pesquisa. Foram utilizados dois conjuntos de dados: um de embarques em linhas de ônibus e um de vendas de produtos do varejo. Estes dados possuem padrões distintos e foram modelados como entrada para alguns modelos de aprendizagem de máquina, com o intuito de prever a quantidade de embarques e de vendas vários passos à frente.

#### 3.1 Conjunto de Dados

O primeiro conjunto de dados trata de séries temporais de vendas de produtos de uma loja do varejo, no setor de supermercado, localizado em Fortaleza (Ceará). Os dados foram obtidos de uma grande varejista, através de um projeto de pesquisa com parceira com a Universidade de Fortaleza. Obteve-se informações de vendas de vinte produtos da curva A (itens com maior contribuição no faturamento da loja) em um período de 02 de Janeiro/2017 a 30 de Abril/2019, totalizando cerca de 850 dias. As vendas dos produtos, por unidades ou por quilogramas, foram agrupadas por dias para cada um dos produtos analisados e montadas as séries temporais finais. Os identificadores dos produtos foram anonimizados. Os meses de Janeiro/2017 a Dezembro/2018 foram separados como treino e os meses de Janeiro/2019 a Abril/2019 como teste. A Figura 1(A) ilustra uma série temporal de venda por dia de um produto da curva A em uma amostra de 240 dias.

O outro conjunto de dados trata de séries temporais do número de embarques de passageiros nas vinte linhas de ônibus com maior uso dentro do sistema de transporte público na cidade de Fortaleza (Ceará). Os passageiros do sistema de ônibus possuem um *smart card* com o identificador do usuário e, toda vez que este cartão é usado, um registro de embarque é gravado. Os dados foram cedidos pela Prefeitura de Fortaleza e foram utilizados em outros artigos [Caminha et al. 2018][Ponte et al. 2018][Bomfim et al. 2020][Ponte et al. 2021]. Estes dados correspondem a um intervalo de 1<sup>o</sup> de Janeiro/2018 a 31 de Julho/2018 e os valores de número de embarque de passageiros foram agrupados por hora para cada uma das vinte linhas de ônibus com mais embarques na cidade. Desta forma, foram geradas vinte séries temporais com mais de 5000 horas de embarques para cada linha de ônibus. Os meses de Janeiro a Junho foram separados como treino e o mês de Julho como teste. A Figura 1(B) ilustra uma série temporal de embarque por hora de uma linha de ônibus, detalhando os padrões de sazonalidade que ocorrem nos veículos em uma amostra de cerca de 10 dias.

Antes de aplicar os métodos de previsão, foi realizada uma preparação adequada dos dados da série temporal. Após uma análise exploratória em ambos os conjuntos de dados, foi identificado que existiam dados não contíguos, ou seja, entradas de dados faltantes. Estes dados faltantes indicam

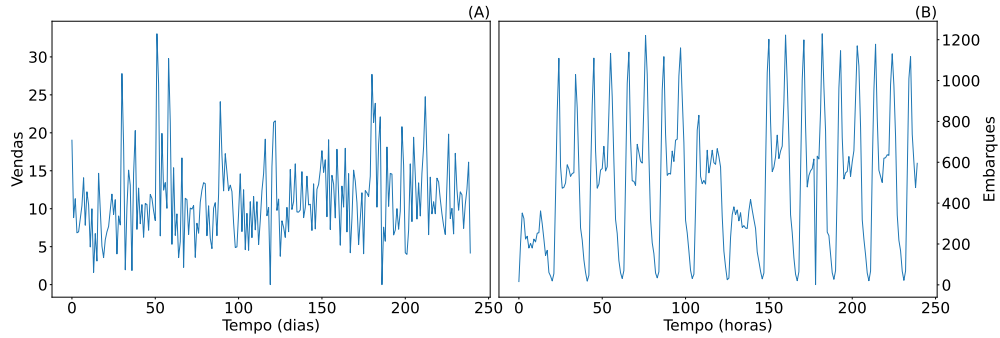


Fig. 1. Dois exemplos de séries temporais nos domínios estudados. Em (A), são ilustrados 240 dias de venda de um produto da curva A. O eixo y representa quantidade vendida desse produto e o eixo x o tempo, em dias. Em (B), são mostradas as primeiras 240 horas de dados da linha 42 de ônibus. O eixo y representa a quantidade de passageiros que embarcaram na linha em questão e o eixo x o tempo, em horas.

momentos em que não ocorrem embarque de passageiros em uma certa hora do dia ou não ocorreram vendas de um determinado produto no dia. Desta forma, estes momentos foram preenchidos com valores zeros, com o objetivo de não romper com a estrutura temporal da série.

### 3.2 Modelagem e Treinamento de Modelos

Utilizou-se o conceito de janelas deslizantes no processo de modelagem do problema de previsão de série temporal. A técnica de janelas deslizantes trata de uma abordagem em que o dado da série temporal é dividido em segmentos menores e de tamanho constante ( $w$ ). O termo deslizante refere-se ao processo de deslocamento da janela ao longo da série e com um determinado passo ( $p$ ) para a direita, permitindo a construção de um conjunto de dados de treinamento.

As janelas deslizantes tratam-se de um artifício para transformar a série temporal em um conjunto de dados rotulado, em que cada janela contém um conjunto de observações ocorridas no passado da série temporal e que consideramos a entrada para os modelos de previsão. A observação logo após o final da janela é definida como o valor alvo que se deseja prever dado a janela passada. Desta forma, está técnica é bastante aderente a métodos de Aprendizado de Máquinas supervisionados de regressão para realizar previsão em séries temporais.

A Figura 2 ilustra o processo da geração dos exemplos de entradas e saídas que serão utilizadas para o treinamento de modelos de IA. Uma série temporal diária, em azul, como exemplo, será transformada em um conjunto de dados supervisionado. Uma janela de tamanho  $w$  é aplicada, gerando a primeira amostra de entrada do conjunto e que possui  $w$  valores, representando as características (*features*) que devem ser aprendidas pelos modelos. Estas características permitem que os modelos aprendam, a partir de padrões e tendências presentes em observações passadas, as dependências temporais das séries. O dia logo após a janela representa a variável alvo que se deseja prever. Em seguida, desloca-se a janela  $p$  passos à direita, a fim de se obter novas amostras para o conjunto de dados e repetindo o processo até percorrer toda a série. Dependendo de  $w$  e  $p$  irá existir uma sobreposição nos dados, reforçando a descoberta do padrão dos dados e aumentando o número de amostras de treinamento.

A parte final do fluxo da Figura 2 ilustra o processo de treinamento de modelos de Aprendizado de Máquina a partir dos dados rotulados das etapas anteriores. Modelos baseados em *ensembles* e em Redes Neurais foram utilizados para conduzir a previsão do número de embarque por hora e do número de produtos vendidos por dia. O objetivo principal desta pesquisa consiste em avaliar como  $w$  impacta na qualidade das previsões em séries temporais univariáveis realizadas por estes modelos. Desta forma, o valor de  $w$  será variado e calculado um erro associado a este  $w$ . Para os dados de mobilidade,  $w$  começará em 24 horas e será incrementado de 24 em 24 horas, até atingir um  $w$  máximo

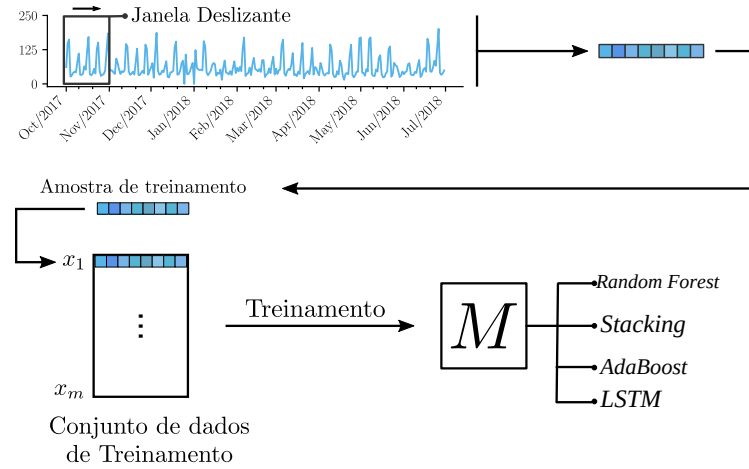


Fig. 2. Ilustração da Janela Deslizante com dados do Varejo. Na ilustração temos uma série temporal de vendas, a criação da amostra de treinamento, o conjunto de dados de treinamento com todas as amostras e o treinamento do modelos  $M$ , podendo ser *Random Forest*, *Stacking*, *AdaBoost* e *LSTM*.

de 360 horas. Para os dados do varejo,  $w$  começará em uma sequência de 7, 14, 21 e 30 dias e, a partir de então, será incrementado de 30 em 30 dias, até atingir um  $w$  máximo de 365 dias. Devido a sazonalidade de alguns produtos do varejo, podendo apresentar comportamentos característicos de forma semanal, quinzenal e mensal, decidiu-se escolher esses dias iniciais de forma mais personalizada. Foi escolhido um valor de  $p = 1$  para ambos os domínios.

Os modelos utilizados serão três do tipo *ensemble* (*Random Forest*, *AdaBoost* e baseado em *Stacking*) e uma rede neural (*LSTM*). No *Random Forest* [Breiman 2001] e no *AdaBoost* [Freund et al. 1996] foram utilizados os parâmetros padrões do *Scikit-learn* [Pedregosa et al. 2011]. Já na rede neural *LSTM* [Hochreiter and Schmidhuber 1997] foi utilizada implementação do *Tensorflow* [Abadi et al. 2016], com os parâmetros  $neurons = 200$ ,  $batch\_size = 32$ , e com função de ativação *ReLU*,  $epochs = 200$ , validação de 20% e otimizador “Adam” com learning rate de 0.0001. O *Stacking* [Wolpert 1992], possui como *estimators*: *Ridge Regressor* [Hoerl and Kennard 1970], *SVR* [Drucker et al. 1996], *Random Forrest*, *Gradient Boosting* [Friedman 2001] e *KNN* [Cover and Hart 1967] e Regressão Linear [Galton 1889] como *final estimator*, todos em suas versões *default*.

### 3.3 Previsão Múltiplos Passos à Frente

O processo de inferência em séries temporais múltiplos passos à frente, ou previsão de longo prazo, consiste em prever os próximos  $h$  valores no futuro. Algumas abordagens de previsão podem ser realizadas como Iterativa, Direta, MIMO (*Multiple Input Multiple Output*) ou combinações entre estas estratégias [Taieb et al. 2012]. Nesta pesquisa, utilizou-se a abordagem Iterativa, em que para estimar os valores futuros para um horizonte  $h$  de tempo, faz-se uma previsão por vez. A Figura 3 ilustra o processo de previsão, em que o modelo  $M$  recebe como entrada uma janela de tamanho  $w$  (inicialmente indexados de 1 a  $t$ ) e produz  $\hat{y}_{t+1}$  como saída. Em seguida, o processo é repetido adicionando a previsão mais recente,  $\hat{y}_{t+1}$ , na nova entrada para produzir a segunda inferência,  $\hat{y}_{t+2}$  (desconsidera-se o valor mais antigo da entrada anterior, mantendo sempre janelas de tamanho  $w$ ). Dependendo do número  $h$  (um número suficientemente grande) e  $w$ , existirá um momento que apenas valores estimados serão usados como entradas, em vez dos dados reais.

A avaliação dos resultados foi feita utilizando o Erro Percentual Médio Simétrico (*SMAPE*) [Makridakis 1993], escolhido pelo fato de ser um erro percentual, uma vez que o *dataset* de varejo contém unidades diferentes (e.g. produtos vendidos por unidade e produtos vendidos no peso). O *SMAPE* também é um erro médio geométrico sendo ideal para comparar o desempenho de múltiplos modelos

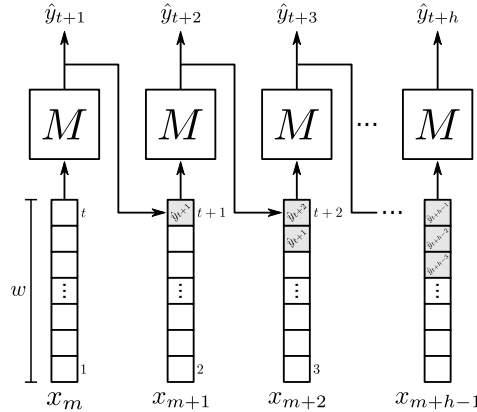


Fig. 3. Esquema da previsão de múltiplos passos à frente, onde a previsão do modelo é utilizada nos passos futuros da janela, fazendo com que o modelo preveja os  $h$  períodos futuros.

e em um alto número de previsões, de acordo com [Kreinovich et al. 2014].

#### 4. RESULTADOS E DISCUSSÃO

A Figura 4 ilustra o resultado dos experimentos seguindo a metodologia detalhada na Seção 3 deste artigo. Foram realizadas previsões de quarenta séries temporais, utilizando quatro modelos de aprendizado de máquina. Em (A) é possível observar os valores de  $SMAPe$  para diversos tamanhos de janela,  $w$ , em experimentos realizados com dados do varejo. Houve uma queda entre todos os modelos para  $w \leq 30$ . Para valores de  $w$  superiores, observa-se que os modelos mantiveram valores  $SMAPe$  com menor variação. No caso da *LSTM* observou-se uma maior variação de  $SMAPe$  para  $w > 30$ .

Na Figura 4(B) é possível ver o desempenho dos modelos para o dado da mobilidade. É perceptível que a previsão apresentou melhorias à medida que o tamanho da janela foi aumentado, tendo a melhora se tornado estável após um determinado tamanho da janela, para alguns modelos. O *Random Forest* foi o modelo que se saiu melhor com esse conjunto de dados, mas a *LSTM* foi o modelo que mais teve uma melhora de *performance* (redução do  $SMAPe$ ), com o aumento do  $w$ . O *AdaBoost* foi o modelo que teve a menor variação da melhora.

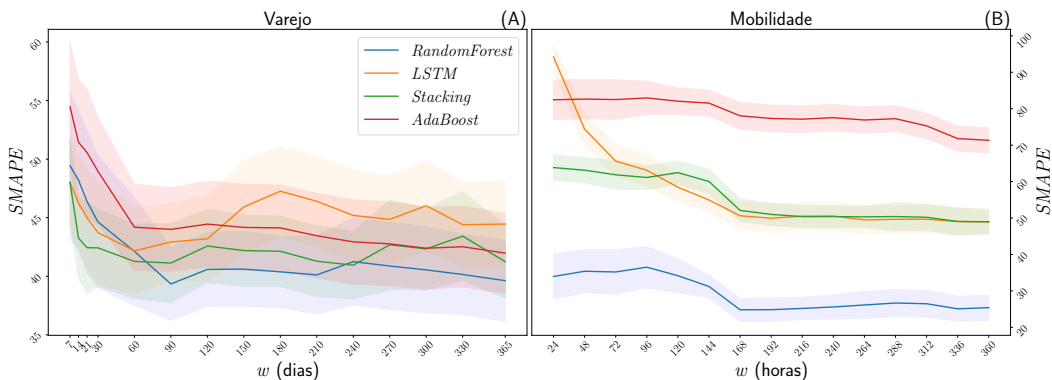


Fig. 4. Desempenho dos modelos, em (A) os resultados da varejo e em (B) os resultados do mobilidade, onde a linha azul representa o desempenho do *Random Forest*, a linha laranja representa o desempenho da *LSTM*, a linha verde representa o desempenho do *Stacking* e a linha vermelha do *AdaBoost*. As sombras, representam o erro médio padrão de cada modelo. No eixo  $x$  está o tamanho da janela e no eixo  $y$  o valor do  $SMAPe$

Analisando a trajetória das curvas dos erros da mobilidade, todos os modelos têm uma redução considerável no erro, em  $w$  igual a 168, onde é possível que os modelos capturem o padrão no comportamento semanal de cada linha.

Em ambos os conjuntos de dados, foi possível observar um impacto positivo nas previsões ao aumentar o tamanho da janela. Após um determinado tamanho de  $w$ , o erro se mantém constante. No entanto, é importante ponderar se vale a pena usar tamanhos de janela muito maiores, uma vez que o aumento do tamanho da janela também aumenta o tamanho dos dados de entrada e, conseqüentemente, o tempo de treinamento dos modelos. No conjunto de dados de mobilidade, o *Adaboost* apresentou o maior erro e a menor redução do erro ao aumentar o tamanho da janela. Isso pode estar relacionado ao fato de que o horizonte de previsão é maior do que no conjunto de dados de varejo.

## 5. CONCLUSÃO

Este artigo teve como foco compreender o impacto do tamanho da janela no desempenho de algoritmos de aprendizado de máquina em problemas de previsão de séries temporais univariáveis. Através de experimentos realizados em 40 séries temporais de dois domínios distintos, utilizando variações no tamanho da janela em quatro algoritmos de aprendizado de máquina, algumas descobertas importantes surgiram. Primeiramente, observou-se que aumentar o tamanho da janela pode levar a melhorias nos valores das métricas de avaliação até um ponto de estabilização. Além desse limite, aumentar ainda mais o tamanho da janela não resulta em previsões melhores.

Além disso, o estudo revelou que as arquiteturas de Redes Neurais Recorrentes não superaram os modelos *ensemble* em diversos cenários de previsão de séries temporais univariáveis. Embora as Redes Neurais Recorrentes sejam conhecidas por sua capacidade de capturar dependências sequenciais e padrões de longo prazo, os modelos *ensemble* apresentaram desempenho similar ou superior em diferentes Tamanhos da Janela. Como as séries temporais univariáveis contêm apenas uma única variável ao longo do tempo, é natural que haja menos informações disponíveis para o modelo aprender e capturar padrões complexos. As Redes Neurais Recorrentes, devido à sua capacidade de lidar com dependências sequenciais, podem ser especialmente eficazes em previsões de séries temporais com uma grande quantidade de dados multivariáveis. Nesses casos, o modelo tem a oportunidade de explorar correlações entre diferentes variáveis e capturar nuances mais sutis nos padrões temporais. No entanto, em séries temporais univariáveis, essa vantagem pode ser minimizada devido à falta de variabilidade nos dados. Por outro lado, os modelos *ensemble* aproveitam a diversidade de diferentes algoritmos ou amostras dos dados para melhorar o desempenho preditivo. Essa abordagem pode ajudar a mitigar os desafios apresentados pelas séries temporais univariáveis devido à sua natureza limitada. Ao combinar várias perspectivas e estratégias de aprendizado, os modelos *ensemble* conseguem compensar a falta de dados multivariáveis, resultando em previsões equivalentes ou mais precisas.

## REFERENCES

- ABADI, M., BARHAM, P., CHEN, J., CHEN, Z., DAVIS, A., DEAN, J., DEVIN, M., GHEMAWAT, S., IRVING, G., ISARD, M., ET AL. Tensorflow: a system for large-scale machine learning. In *Osd* (2016). Vol. 16. Savannah, GA, USA, USENIX Association, USA, pp. 265–283, 2016.
- ABOLGHASEMI, M., BEH, E., TARR, G., AND GERLACH, R. Demand forecasting in supply chain: The impact of demand volatility in the presence of promotion. *Computers & Industrial Engineering* vol. 142, pp. 106380, 2020.
- AGHABOZORGI, S., SHIRKORSHIDI, A. S., AND WAH, T. Y. Time-series clustering—a decade review. *Information systems* vol. 53, pp. 16–38, 2015.
- AZLAN, A., YUSOF, Y., AND MOHSIN, M. F. M. Determining the impact of window length on time series forecasting using deep learning. *International Journal of Advanced Computer Research* 9 (44): 260–267, 2019.
- BERGSTRÖM, C. AND HJELM, O. Impact of time steps on stock market prediction with lstm, 2019.
- BOMFIM, R., PEI, S., SHAMAN, J., YAMANA, T., MAKSE, H. A., ANDRADE JR, J. S., LIMA NETO, A. S., AND FURTADO, V. Predicting dengue outbreaks at neighbourhood level using human mobility in urban areas. *Journal of the Royal Society Interface* 17 (171): 20200691, 2020.
- BREIMAN, L. Random forests. *Machine learning* vol. 45, pp. 5–32, 2001.

- CAMINHA, C., FURTADO, V., PINHEIRO, V., AND PONTE, C. Graph mining for the detection of overcrowding and waste of resources in public transport. *Journal of Internet Services and Applications* 9 (1): 1–11, 2018.
- CHENG, H., TAN, P.-N., GAO, J., AND SCRIPPS, J. Multistep-ahead time series prediction. In *Advances in Knowledge Discovery and Data Mining: 10th Pacific-Asia Conference, PAKDD 2006, Singapore, April 9-12, 2006. Proceedings 10*. Springer, Singapore, pp. 765–774, 2006.
- COVER, T. AND HART, P. Nearest neighbor pattern classification. *IEEE transactions on information theory* 13 (1): 21–27, 1967.
- DE GOOLIER, J. G. AND HYNDMAN, R. J. 25 years of time series forecasting. *International journal of forecasting* 22 (3): 443–473, 2006.
- DRUCKER, H., BURGESS, C. J., KAUFMAN, L., SMOLA, A., AND VAPNIK, V. Support vector regression machines. *Advances in neural information processing systems* vol. 9, pp. 155–161, 1996.
- FAWAZ, H. I., FORESTIER, G., WEBER, J., IDOUMGHAR, L., AND MULLER, P.-A. Adversarial attacks on deep neural networks for time series classification. In *2019 International Joint Conference on Neural Networks (IJCNN)*. IEEE, Budapest, pp. 1–8, 2019.
- FREUND, Y., SCHAPIRE, R. E., ET AL. Experiments with a new boosting algorithm. In *icml*. Vol. 96. Citeseer, New Jersey, pp. 148–156, 1996.
- FRIEDMAN, J. H. Greedy function approximation: a gradient boosting machine. *Annals of statistics* 29 (5): 1189–1232, 2001.
- GALTON, F. *Natural inheritance*. Vol. 42. Macmillan, London, 1889.
- HAMZAÇEBI, C., AKAY, D., AND KUTAY, F. Comparison of direct and iterative artificial neural network forecast approaches in multi-periodic time series forecasting. *Expert systems with applications* 36 (2): 3839–3844, 2009.
- HOCHREITER, S. AND SCHMIDHUBER, J. Long short-term memory. *Neural computation* 9 (8): 1735–1780, 1997.
- HOERL, A. E. AND KENNARD, R. W. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* 12 (1): 55–67, 1970.
- HUBER, J. AND STUCKENSCHMIDT, H. Daily retail demand forecasting using machine learning with emphasis on calendric special days. *International Journal of Forecasting* 36 (4): 1420–1438, 2020.
- KREINOVICH, V., NGUYEN, H. T., AND OUNCHAROEN, R. How to estimate forecasting quality: A system-motivated derivation of symmetric mean absolute percentage error (smape) and other similar characteristics. *Departmental Technical Reports (CS)*, 2014.
- LIM, B. AND ZOHREN, S. Time-series forecasting with deep learning: a survey. *Philosophical Transactions of the Royal Society A* 379 (2194): 20200209, 2021.
- LIU, Y., WANG, Z., YU, X., CHEN, X., AND SUN, M. Memory-based transformer with shorter window and longer horizon for multivariate time series forecasting. *Pattern Recognition Letters* vol. 160, pp. 26–33, 2022.
- LOVRIĆ, M., MILANOVIĆ, M., AND STAMENKOVIĆ, M. Algorithmic methods for segmentation of time series: An overview. *Journal of Contemporary Economic and Business Issues* 1 (1): 31–53, 2014.
- MAKRIDAKIS, S. Accuracy measures: theoretical and practical concerns. *International journal of forecasting* 9 (4): 527–529, 1993.
- MUNIR, M., SIDDIQUI, S. A., DENGEL, A., AND AHMED, S. Deepant: A deep learning approach for unsupervised anomaly detection in time series. *Ieee Access* vol. 7, pp. 1991–2005, 2018.
- PEDREGOSA, F., VAROQUAUX, G., GRAMFORT, A., MICHEL, V., THIRION, B., GRISEL, O., BLONDEL, M., PRETTENHOFER, P., WEISS, R., DUBOURG, V., ET AL. Scikit-learn: Machine learning in python. *the Journal of machine Learning research* vol. 12, pp. 2825–2830, 2011.
- PONTE, C., CARMONA, H. A., OLIVEIRA, E. A., CAMINHA, C., LIMA, A. S., ANDRADE JR, J. S., AND FURTADO, V. Tracing contacts to evaluate the transmission of covid-19 from highly exposed individuals in public transportation. *Scientific Reports* 11 (1): 24443, 2021.
- PONTE, C., MELO, H. P. M., CAMINHA, C., ANDRADE JR, J. S., AND FURTADO, V. Traveling heterogeneity in public transportation. *EPJ Data Science* 7 (1): 1–10, 2018.
- SALLES, R., BELLOZE, K., PORTO, F., GONZALEZ, P. H., AND OGASAWARA, E. Nonstationary time series transformation methods: An experimental review. *Knowledge-Based Systems* vol. 164, pp. 274–291, 2019.
- SHYNKEVICH, Y., MCGINNITY, T. M., COLEMAN, S. A., BELATRECHE, A., AND LI, Y. Forecasting price movements using technical indicators: Investigating the impact of varying input window length. *Neurocomputing* vol. 264, pp. 71–88, 2017.
- TAIEB, S. B., BONTEMPI, G., ATIYA, A. F., AND SORJAMAA, A. A review and comparison of strategies for multi-step ahead time series forecasting based on the nn5 forecasting competition. *Expert systems with applications* 39 (8): 7067–7083, 2012.
- UGHI, R., LOMURNO, E., AND MATTEUCCI, M. Two steps forward and one behind: Rethinking time series forecasting with deep learning. *arXiv preprint arXiv:2304.04553*, 2023.
- WOLPERT, D. H. Stacked generalization. *Neural networks* 5 (2): 241–259, 1992.