



UNIVERSIDADE FEDERAL DO MARANHÃO
Curso de Ciências da Computação

KAIO YUKIO GONÇALVES VIEIRA GUEDES

ANÁLISE DE SÉRIES TEMPORAIS: Conceitos e aplicações

São Luís - MA
2025



KAIO YUKIO GONÇALVES VIEIRA GUEDES

ANÁLISE DE SÉRIES TEMPORAIS: Conceitos e aplicações

Trabalho de Conclusão de Curso
apresentado ao Curso de Bacharelado em
Ciência da Computação da Universidade
Federal do Maranhão, como requisito para
obtenção do grau de Bacharel em Ciência
da Computação.

Orientador: Prof. Dr. Ivo José da Cunha Serra

São Luís - MA
2025

Ficha gerada por meio do SIGAA/Biblioteca com dados fornecidos pelo(a) autor(a).
Diretoria Integrada de Bibliotecas/UFMA

Gonçalves Vieira Guedes, Kaio Yukio.

Análise de Séries Temporais : conceitos e aplicações /
Kaio Yukio Gonçalves Vieira Guedes. - 2025.
53 p.

Orientador(a): Ivo José da Cunha Serra.

Curso de Ciência da Computação, Universidade Federal do
Maranhão, São Luís - Ma, 2025.

1. Redes Neurais Lstm. 2. Shap. 3. Análise Espacial.
I. da Cunha Serra, Ivo José. II. Título.

KAIO YUKIO GONÇALVES VIEIRA GUEDES

ANÁLISE DE SÉRIES TEMPORAIS: Conceitos e aplicações

Trabalho de Conclusão de Curso
apresentado ao Curso de Bacharelado em
Ciência da Computação da Universidade
Federal do Maranhão, como requisito para
obtenção do grau de Bacharel em Ciência
da Computação.

Aprovado em _____ de _____ de _____.

Banca Examinadora

Prof. Dr. Ivo José da Cunha Serra

Orientador

Profª. Dra. Simara Vieira da Rocha

Examinador

Prof. Me. Carlos Eduardo Portela Serra de Castro

Examinador

AGRADECIMENTOS

Agradeço à minha família, em especial à minha mãe, pelo constante incentivo à minha educação e por sua participação ativa ao longo da minha trajetória acadêmica. Estendo esse agradecimento à minha avó, cuja presença e apoio foram relevantes em minha formação.

Sou grato também aos meus amigos, cuja convivência contribuiu positivamente durante esse período, e à minha namorada, pelo apoio e compreensão ao longo da jornada.

Por fim, agradeço ao meu professor e orientador, Ivo, pela orientação precisa, disponibilidade e suporte constante durante o desenvolvimento deste trabalho.

Prever é muito difícil, especialmente o futuro.

Niels Bohr

RESUMO

A análise de séries temporais tem ganhado destaque diante da crescente disponibilidade de dados cronológicos em contextos complexos, como saúde pública, comércio e mobilidade urbana. Este estudo teve como objetivo principal analisar os principais conceitos e metodologias aplicadas à modelagem de séries temporais, destacando sua aplicabilidade prática por meio da comparação de estudos de caso em diferentes áreas, como saúde pública, varejo e mobilidade urbana. A execução do estudo ocorreu por meio de uma abordagem qualitativa e exploratória, utilizando três estudos de caso baseados em artigos científicos: (1) previsão de surtos de dengue com redes LSTM e variáveis climáticas e espaciais defasadas; (2) previsão de demanda no varejo com métodos tradicionais e algoritmos de machine learning; e (3) análise espaço-temporal de acidentes de trânsito urbanos com técnicas de geoprocessamento (GIS). Como contribuição, o trabalho oferece uma análise comparativa entre abordagens clássicas e modernas, evidenciando a superioridade dos modelos LSTM em ambientes de alta variabilidade.

Palavras-chave: Redes neurais LSTM. SHAP. Previsão preditiva. Análise espacial.

ABSTRACT

Time series analysis has gained prominence due to the increasing availability of chronological data in complex contexts such as public health, commerce, and urban mobility. This study aimed to analyze the main concepts and methodologies applied to time series modeling, highlighting their practical applicability through the comparison of case studies in different areas, including public health, retail, and urban mobility. The research was conducted using a qualitative and exploratory approach, based on three case studies drawn from scientific articles: (1) forecasting dengue outbreaks using LSTM networks with lagged climatic and spatial variables; (2) demand forecasting in retail using traditional methods and machine learning algorithms; and (3) spatiotemporal analysis of urban traffic accidents using Geographic Information System (GIS) techniques. As a contribution, the study offers a comparative analysis between classical and modern approaches, evidencing the superior performance of LSTM models in high-variability environments.

Keywords: LSTM neural networks. SHAP. Predictive forecasting. Spatial analysis.

LISTA DE ABREVIATURAS E SIGLAS

AR	Autoregressive (Autorregressivo)
ARFIMA	Autoregressive Fractionally Integrated Moving Average
ARIMA	Autoregressive Integrated Moving Average
ANN	Artificial Neural Network (Rede Neural Artificial)
FIGARCH	Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity
GARCH	Generalized Autoregressive Conditional Heteroskedasticity
GIS	Geographic Information System (Sistema de Informação Geográfica)
GRU	Gated Recurrent Unit
LIME	Local Interpretable Model-agnostic Explanations
LSTM	Long Short-Term Memory
MA	Moving Average (Média Móvel)
ML	Machine Learning (Aprendizado de Máquina)
MLR	Multiple Linear Regression (Regressão Linear Múltipla)
RNN	Recurrent Neural Network (Rede Neural Recorrente)
SHAP	SHapley Additive exPlanations
SVM	Support Vector Machine (Máquina de Vetores de Suporte)

LISTA DE SÍMBOLOS

Y_t	Valor da série temporal no tempo t
ε_t	Termo de erro aleatório (ruído branco) no tempo t
φ	Parâmetro de autorregressão (modelo AR)
θ	Parâmetro de média móvel (modelo MA)
p	Ordem do modelo autorregressivo (AR)
q	Ordem do modelo de médias móveis (MA)
d	Ordem de diferenciação (modelo ARIMA)
∇	Operador de diferença
μ	Média da série temporal
σ^2	Variância
h_t	Variância condicional no tempo t (modelo <i>GARCH</i>)
α	Parâmetro de impacto do choque (modelo <i>GARCH</i>)
β	Parâmetro de persistência da variância (modelo <i>GARCH</i>)
L	Operador de defasagem (<i>lag operator</i>)
$f(\cdot)$	Função de ativação em redes neurais
x_i	Valor da i -ésima variável preditora
ϕ_i	Valor shap atribuído à i -ésima variável preditora

SUMÁRIO

1 INTRODUÇÃO	12
1.1 OBJETIVO GERAL	13
1.2 OBJETIVOS ESPECÍFICOS	13
1.3 ORGANIZAÇÃO DOS CAPÍTULOS	13
2 FUNDAMENTAÇÃO TEÓRICA	14
2.1 CONCEITO DE SÉRIES TEMPORAIS	14
2.2 MODELOS LINEARES CLÁSSICOS: AR, MA, ARMA E ARIMA	15
2.3 MODELOS DE REFERÊNCIA: MLR, TENDÊNCIA E NAÏVE	17
2.4 MODELOS NÃO LINEARES E DE MEMÓRIA LONGA: GARCH E ARFIMA	17
2.5 MACHINE LEARNING EM SÉRIES TEMPORAIS	18
2.6 REDES NEURAIS RECORRENTES (RNN) E MEMÓRIA DE LONGO CURTO PRAZO (LSTM)	19
2.7 TÉCNICAS DE INTERPRETABILIDADE EM MODELOS PREDITIVOS: SHAP	20
3 ESTUDOS DE CASO E DISCUSSÃO – SÉRIES TEMPORAIS	23
3.1 ESTUDO DE CASO 1: PREVISÃO DE SURTOS DE DENGUE COM REDES LSTM E EXPLICABILIDADE SHAP	23
3.2 ESTUDO DE CASO 2: PREVISÃO DE DEMANDA NO VAREJO	28
3.3 ESTUDO DE CASO 3: ANÁLISE DE TRÂNSITO PARA PREVENÇÃO DE ÁREAS DE RISCO	33
3.3.1 Evolução temporal e distribuição dos acidentes	36
3.3.2 Gravidade dos acidentes e impacto sobre a população	37
3.3.3 Padrões espaciais e identificação de <i>hotspots</i>	38
3.3.4 Variações temporais e perfil dos acidentes	39
3.3.5 Fatores associados e implicações para a formulação de políticas públicas	40
4 DISCUSSÃO DOS ESTUDOS DE CASO	43
4.1 ESTUDO DE CASO 1: PREVISÃO DE SURTOS DE DENGUE COM REDES LSTM E EXPLICABILIDADE SHAP	43
4.2 ESTUDO DE CASO 2: PREVISÃO DE DEMANDA NO VAREJO	45
4.3 ESTUDO DE CASO 3: ANÁLISE DE TRÂNSITO PARA PREVENÇÃO DE ÁREAS DE RISCO	46
5 CONSIDERAÇÕES FINAIS	48
REFERÊNCIAS	50

1 INTRODUÇÃO

A análise de séries temporais é um campo fundamental da estatística e da ciência de dados, voltado à compreensão, modelagem e previsão de fenômenos que evoluem ao longo do tempo. Desde séries econômicas e financeiras até indicadores de saúde pública e padrões de mobilidade urbana, os dados temporais têm se tornado cada vez mais acessíveis e complexos, exigindo ferramentas analíticas robustas e adaptáveis. Segundo Antunes e Cardoso (2015), a interpretação adequada desses dados é crucial para a tomada de decisões estratégicas em diversos setores da sociedade.

Com o avanço das tecnologias de coleta e armazenamento de dados, tornou-se possível monitorar e analisar eventos com granularidade temporal cada vez maior, permitindo o desenvolvimento de modelos preditivos mais precisos. No entanto, essa sofisticação também impõe novos desafios metodológicos, como a escolha apropriada do modelo, a identificação de padrões sazonais, tendências e ruídos, além da avaliação do desempenho das previsões geradas. Nesse contexto, tanto os métodos estatísticos clássicos quanto os algoritmos de aprendizado de máquina têm sido explorados de forma complementar.

A metodologia adotada neste estudo é de abordagem qualitativa e exploratória, uma vez que visa compreender, descrever e comparar o uso de modelos preditivos aplicados à análise de séries temporais em contextos distintos. Trata-se de uma investigação baseada em estudo de casos múltiplos, com a análise aprofundada de três artigos científicos que abordam aplicações práticas de técnicas modernas, como redes neurais recorrentes (especialmente LSTM) e métodos de interpretabilidade como o SHAP.

A seleção dos artigos considerou critérios como relevância científica, diversidade de aplicações e uso explícito de modelos de séries temporais contemporâneos. Os estudos escolhidos contemplam diferentes áreas: saúde pública, comércio varejista e mobilidade urbana. Foram analisados os seguintes trabalhos: (1) Chen e Moraga (2025), que aplicam LSTM com variáveis climáticas e espaciais defasadas para prever surtos de dengue; (2) Carbonneau, Laframboise e Vahidov (2008), que comparam métodos tradicionais e de aprendizado de máquina na previsão

de demanda no varejo; e (3) Mohammed et al. (2023), que utilizam técnicas de GIS para realizar uma análise espaço-temporal de acidentes de trânsito urbanos.

Portanto, a análise dos trabalhos selecionados evidencia não apenas a diversidade de aplicações das séries temporais, mas também sua relevância crescente como suporte indispensável à tomada de decisão em contextos complexos e dinâmicos.

1.1 OBJETIVO GERAL

Analisar os principais conceitos e metodologias aplicadas à modelagem de séries temporais, destacando sua aplicabilidade prática por meio da comparação de estudos de caso em diferentes áreas, como saúde pública, varejo e mobilidade urbana.

1.2 OBJETIVOS ESPECÍFICOS

Compreender os fundamentos teóricos da análise de séries temporais (tendência, sazonalidade, ciclos e ruído).

Investigar abordagens estatísticas e de *machine learning*.

Comparar desempenho e desafios em três aplicações práticas (3 artigos).

1.3 ORGANIZAÇÃO DOS CAPÍTULOS

Este trabalho está estruturado em cinco capítulos. O Capítulo 2 trata da fundamentação teórica, abordando os principais conceitos de séries temporais, os modelos lineares e não lineares utilizados, bem como técnicas modernas como redes neurais recorrentes (RNN e LSTM) e métodos de interpretabilidade como o SHAP. O Capítulo 3 descreve a metodologia empregada e apresenta os três estudos de caso selecionados para análise: previsão de surtos de dengue, previsão de demanda no varejo e análise espaço-temporal de acidentes de trânsito urbanos. O Capítulo 4 apresenta a discussão crítica dos resultados obtidos em cada caso, destacando os avanços, limitações e comparações entre as abordagens. Por fim, o Capítulo 5 traz as considerações finais e propõe sugestões para futuras pesquisas na área.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 CONCEITO DE SÉRIES TEMPORAIS

O estudo de séries temporais tem se consolidado como um dos pilares da análise estatística aplicada, especialmente diante da crescente disponibilidade de dados organizados cronologicamente. Segundo Gonçalves (2020), uma série temporal é definida como uma sucessão de observações de uma variável, coletadas em intervalos regulares de tempo, cujo objetivo analítico reside na identificação de padrões, tendências e estruturas que possibilitem previsões futuras.

A importância da análise temporal reside justamente na ordenação sequencial dos dados, elemento importante para compreender a dependência entre as observações. Essa dependência, conforme destacam Gonçalves e Lopes (2013), diferencia a série temporal de outros conjuntos de dados estatísticos, pois cada valor observado pode estar correlacionado com valores passados, uma característica típica de fenômenos evolutivos como variações econômicas, oscilações ambientais e incidência de doenças.

Tradicionalmente, os modelos aplicados à análise de séries temporais eram essencialmente lineares, como os modelos autorregressivos (AR), de médias móveis (MA) e os combinados ARMA. Estes modelos assumem que o valor atual de uma série pode ser expresso como uma função linear de seus valores passados e de um termo de erro branco (Box; Jenkins, 1970). No entanto, com o surgimento de fenômenos caracterizados por comportamentos abruptos ou não estacionários, como flutuações financeiras ou dados epidemiológicos, tais modelos mostraram-se insuficientes para capturar a complexidade dos dados.

Com o intuito de lidar com essa limitação, autores como Engle (1982) e Granger e Andersen (1978) introduziram modelos não lineares, como os modelos bilineares e os modelos com heterocedasticidade condicional (ARCH/GARCH), ampliando o escopo analítico das séries temporais. Além disso, novas abordagens passaram a considerar a noção de memória longa, como no caso dos modelos ARFIMA, que, segundo Hosking (1981), permitem modelar séries em que as dependências temporais se estendem por longos períodos, o que é especialmente útil para representar fenômenos de natureza persistente.

Do ponto de vista probabilístico, uma série temporal pode ser entendida como uma realização de um processo estocástico, isto é, um conjunto de variáveis aleatórias indexadas por tempo (Gonçalves; Lopes, 2013). Em um contexto mais formal, este processo é definido sobre um espaço de probabilidade (Ω, Σ, P) , sendo caracterizado por propriedades como estacionaridade, expectativa e autocovariância, aspectos fundamentais para a escolha de modelos adequados.

Portanto, o conceito de séries temporais transcende a simples disposição sequencial de dados. Trata-se de uma ferramenta poderosa para compreender padrões de comportamento ao longo do tempo, modelar a evolução de fenômenos dinâmicos e fornecer subsídios técnicos para a tomada de decisão em áreas como saúde pública, finanças, comércio e mobilidade urbana.

2.2 MODELOS LINEARES CLÁSSICOS: AR, MA, ARMA E ARIMA

Dentre os modelos clássicos de séries temporais, destacam-se os modelos lineares, com ênfase nos modelos autorregressivos (AR), de médias móveis (MA), e suas combinações: os modelos ARMA e ARIMA. Segundo Box et al. (2015), os modelos ARIMA foram desenvolvidos como uma abordagem sistemática para identificar, estimar e validar modelos de séries temporais estacionárias e não estacionárias. A metodologia Box-Jenkins, introduzida pelos autores, tornou-se referência no campo da previsão.

No modelo $AR(p)$, o valor presente da série é uma função linear dos seus valores passados. Já o modelo $MA(q)$ considera o valor presente como uma combinação linear dos erros passados (ruído branco). A junção destes dois conceitos origina o modelo $ARMA(p,q)$, apropriado para séries estacionárias. Quando a série não é estacionária, utiliza-se a diferenciação para estabilizar a média, resultando no modelo $ARIMA(p,d,q)$, no qual “d” representa a ordem de diferenciação necessária para atingir a estacionaridade (Brockwell; Davis, 2002).

Ainda de acordo com Brockwell e Davis (2002), os modelos em espaço de estados e o filtro de Kalman expandiram significativamente a capacidade preditiva dos modelos lineares, permitindo representar componentes como tendência e sazonalidade de forma estocástica, o que contribui para uma modelagem mais flexível e realista, especialmente em aplicações econômicas e ambientais.

No contexto teórico, é fundamental compreender os conceitos probabilísticos subjacentes à modelagem de séries temporais. Como destacam Gonçalves e Lopes (2013), esses modelos são construídos com base em variáveis aleatórias e processos estocásticos, os quais exigem rigor na definição de propriedades como estacionaridade, média, variância, covariância e correlação.

2.3 MODELOS DE REFERÊNCIA: MLR, TENDÊNCIA E NAÏVE

Além dos modelos lineares clássicos já discutidos, diversas abordagens tradicionais de previsão permanecem amplamente empregadas na análise de séries temporais, sobretudo como referenciais comparativos (benchmarks) em pesquisas aplicadas. Tais métodos destacam-se pela simplicidade, baixo custo computacional e facilidade de interpretação (Hyndman; Athanasopoulos, 2018).

Entre eles, a Regressão Linear Múltipla (MLR) permite modelar a variável de interesse a partir de múltiplos preditores, incorporando variáveis exógenas que influenciam o comportamento da série. Essa característica torna o método especialmente útil em contextos onde a previsão depende de fatores externos relevantes (Montgomery; Peck; Vining, 2012).

Outras abordagens incluem o modelo de Tendência (Trend), que projeta valores futuros com base no comportamento ascendente ou descendente da série, e o método Naïve, que assume que o último valor registrado é a melhor estimativa para o próximo período. Apesar de suas limitações na captura de padrões complexos, ambos se destacam pela facilidade de implementação e clareza interpretativa (Montgomery; Jennings; Kulahci, 2015).

Embora não explorem relações não lineares ou interações complexas, esses métodos continuam relevantes em diferentes áreas de aplicação, servindo tanto de ponto de partida quanto de referência comparativa, como será demonstrado no estudo de caso referente ao trabalho de Carbonneau et al. (2008).

2.4 MODELOS NÃO LINEARES E DE MEMÓRIA LONGA: GARCH E ARFIMA

A análise de séries temporais frequentemente exige modelos capazes de capturar comportamentos complexos que não seguem estruturas lineares simples, como mudanças abruptas de regime, heterocedasticidade condicional e persistência de longo prazo. Diante dessas características, os modelos GARCH (*Generalized Autoregressive Conditional Heteroskedasticity*) e ARFIMA (*Autoregressive Fractionally Integrated Moving Average*) têm se destacado por sua capacidade de representar fenômenos econômicos, financeiros e sociais com maior realismo e precisão.

O modelo ARFIMA, introduzido por Granger e Joyeux (1980) e Hosking (1981), é uma generalização dos modelos ARIMA que permite a integração fracionária da série, ou seja, a presença de memória longa. Diferentemente dos modelos tradicionais de diferença inteira, o ARFIMA admite um parâmetro de diferenciação fracionária $d \in (0, 0.5)$, o que permite modelar séries com autocorrelação decaindo de forma hiperbólica, refletindo uma dependência persistente ao longo do tempo. Assim, tal tipo de modelagem é particularmente útil para fenômenos em que os choques passados têm influência prolongada, como taxas de inflação, indicadores ambientais ou dados de tráfego urbano (Brockwell; Davis, 2002).

Complementarmente, os modelos GARCH, propostos por Bollerslev (1986) a partir do ARCH de Engle (1982), são utilizados para modelar a heterocedasticidade condicional – a variância dos erros não é constante ao longo do tempo, mas sim dependente das observações passadas. Assim, esses modelos são importantes na análise de séries financeiras, onde choques de volatilidade tendem a se agrupar em clusters, e sua extensão FIGARCH (*Fractionally Integrated GARCH*), introduzida por Baillie et al. (1996), incorpora o conceito de memória longa à variância condicional.

No estudo de Figueiredo e Marques (2009), a inflação brasileira no período pós-Plano Real é descrita por meio de um modelo ARFIMA-FIGARCH, revelando características de memória longa tanto na média quanto na variância da série. O modelo estimado indicou que a inflação apresenta um comportamento estacionário, com lenta dissipação de choques e persistência estatisticamente significativa. Além disso, o trabalho evidenciou uma relação recíproca entre a taxa média de inflação e

sua volatilidade, demonstrando a importância da modelagem conjunta da média e da variância com estrutura de memória longa.

Ainda que não sejam empregados diretamente nos estudos de caso deste trabalho, estes achados reforçam a relevância dos modelos ARFIMA e GARCH como ferramentas robustas para a compreensão de séries temporais com comportamentos persistentes e variáveis ao longo do tempo. A combinação dessas abordagens possibilita capturar nuances importantes na dinâmica dos dados, ampliando a capacidade preditiva e a compreensão de fenômenos complexos em diferentes contextos, além de servir como contraponto aos modelos lineares clássicos.

2.5 MACHINE LEARNING EM SÉRIES TEMPORAIS

O avanço da ciência de dados e da capacidade computacional nas últimas décadas proporcionou o desenvolvimento de abordagens baseadas em Machine Learning (ML) para a análise e previsão de séries temporais. Diferentemente dos modelos estatísticos clássicos, que pressupõem uma estrutura probabilística explícita e dependem da suposição de estacionaridade ou linearidade, os modelos de aprendizado de máquina são essencialmente orientados por dados (*data-driven*) e têm a capacidade de capturar padrões complexos, não lineares e de alta dimensionalidade sem, necessariamente, especificar uma estrutura paramétrica rígida (Carbonneau; Laframboise; Vahidov, 2008).

Entre os métodos mais utilizados, destacam-se as redes neurais artificiais (ANN), máquinas de vetor de suporte (SVM), random forests, e, mais recentemente, as redes neurais recorrentes (RNN) e suas variações, como as *Long Short-Term Memory* (LSTM) e *Gated Recurrent Units* (GRU), que foram projetadas especificamente para lidar com dependências temporais e sequenciais de longo prazo (Hewamalage; Bergmeir; Bandara, 2021).

Segundo Fawaz et al. (2019), modelos de *deep learning* vêm apresentando desempenho competitivo em várias tarefas de previsão, especialmente em contextos com grandes volumes de dados e presença de múltiplas variáveis explicativas. As RNNs, por exemplo, são capazes de manter uma “memória” das observações passadas por meio de estados ocultos, o que as torna adequadas para detectar padrões sazonais, tendências e ciclos complexos ao longo do tempo.

No entanto, apesar de seu desempenho, esses modelos também apresentam desafios. Um deles é a dificuldade de interpretabilidade, ou seja, entender como e por que determinada previsão foi feita. Para contornar essa limitação, técnicas como o SHAP (*SHapley Additive exPlanations*) vêm sendo aplicadas para fornecer interpretabilidade aos modelos de ML, especialmente aqueles mais complexos. SHAP permite identificar quais variáveis e em quais momentos influenciaram mais a previsão feita, promovendo maior transparência (Lundberg; Lee, 2017).

É importante salientar que os modelos de aprendizado de máquina não substituem os métodos estatísticos clássicos, mas sim os complementam. A escolha entre abordagens tradicionais e algoritmos de ML deve considerar fatores como a quantidade de dados disponíveis, a necessidade de interpretabilidade, o grau de complexidade do fenômeno modelado e os recursos computacionais acessíveis.

Assim, a integração de métodos estatísticos e de machine learning representa uma fronteira promissora na modelagem de séries temporais, especialmente em aplicações de alta variabilidade como saúde pública, comércio varejista e mobilidade urbana, como demonstrado nos estudos de caso abordados neste trabalho.

2.6 REDES NEURAIS RECORRENTES (RNN) E MEMÓRIA DE LONGO CURTO PRAZO (LSTM)

As Redes Neurais Recorrentes (RNN) surgiram como uma evolução das redes neurais tradicionais para lidar com dados sequenciais, nos quais a ordem temporal das observações exerce papel fundamental na modelagem. As RNNs são projetadas para reconhecer padrões temporais e espaciais em dados sequenciais, utilizando conexões recorrentes que permitem a persistência de informações de estados anteriores no processamento atual.

Entretanto, a estrutura básica das RNNs enfrenta limitações relevantes, como o problema do desvanecimento ou explosão do gradiente, dificultando o aprendizado de dependências de longo prazo. Essa limitação motivou o desenvolvimento das *Long Short-Term Memory* (LSTM), uma arquitetura proposta por Hochreiter e Schmidhuber (1997) que introduz unidades de memória com mecanismos de portas (gate mechanisms), responsáveis por controlar o fluxo de informações e preservar gradientes relevantes por períodos prolongados. As LSTM conseguem assimilar

relações entre dados com defasagens superiores a mil passos de tempo discreto, superando largamente as limitações das RNNs convencionais.

As redes LSTM têm se destacado em tarefas de previsão em séries temporais complexas, como demonstrado por Chen e Moraga (2025) em estudo aplicado à previsão de casos de dengue no Brasil. Os autores propuseram um modelo LSTM integrado a variáveis climáticas e espaciais defasadas, utilizando SHAP (*SHapley Additive exPlanations*) para interpretar a importância dos preditores. O modelo demonstrou capacidade robusta de capturar as dinâmicas temporais e geoespaciais da doença, superando abordagens tradicionais e destacando o potencial da LSTM em aplicações em saúde pública e políticas preditivas.

Dessa forma, as redes LSTM representam uma alternativa promissora no contexto da previsão de séries temporais com alta complexidade, como aquelas afetadas por sazonalidades, dependências de longo prazo e influências espaciais.

2.7 TÉCNICAS DE INTERPRETABILIDADE EM MODELOS PREDITIVOS: SHAP

A crescente complexidade dos modelos de aprendizado de máquina, especialmente os baseados em redes neurais profundas e algoritmos de ensemble, tem levantado desafios importantes no que diz respeito à interpretabilidade das previsões. Em muitos contextos críticos, como saúde pública, meio ambiente ou políticas públicas, compreender por que um modelo tomou determinada decisão é tão importante quanto a sua acurácia. A transparência se torna uma demanda ética, técnica e social.

Nesse cenário, destaca-se o método SHAP (*SHapley Additive exPlanations*), proposto por Lundberg e Lee (2017), como uma das abordagens mais robustas e teoricamente fundamentadas para atribuição de importância a variáveis em modelos complexos. Inspirado na teoria dos valores de Shapley, da teoria dos jogos cooperativos, o SHAP quantifica a contribuição de cada característica individual para uma determinada predição, assegurando propriedades desejáveis como localidade, coerência e aditividade.

O diferencial do SHAP está em sua capacidade de oferecer explicações locais, ou seja, interpretações sobre previsões individuais, além de possibilitar a análise global do comportamento do modelo. A estrutura teórica proposta unifica e generaliza

outros métodos de interpretabilidade anteriormente existentes, como LIME, DeepLIFT e métodos baseados em gradientes, tornando-se uma das soluções mais utilizadas na explicação de modelos preditivos de última geração.

A aplicabilidade do SHAP tem sido demonstrada em diversas áreas. No estudo de De Souza et al. (2024), por exemplo, o SHAP foi empregado para aprimorar a interpretabilidade de modelos preditivos utilizados na análise da qualidade óptica da água em reservatórios por meio de sensoriamento remoto. Os autores evidenciam como o uso do SHAP permite a identificação precisa das variáveis ambientais mais relevantes para as previsões, promovendo maior confiança nos resultados obtidos e favorecendo a tomada de decisões fundamentadas.

Assim, a utilização de métodos como o SHAP revela-se indispensável quando se busca aliar a eficiência preditiva dos modelos complexos à responsabilidade interpretativa, sobretudo em áreas sensíveis à compreensão do processo de decisão algorítmica.

A fundamentação teórica sobre modelos lineares e redes neurais forneceu o embasamento necessário para entender como métodos como RNN e SVM superam limitações de abordagens clássicas. No Capítulo 3, essa análise ganha concretude a partir da aplicação dos modelos a uma base de dados real do setor varejista, revelando os desafios práticos e os ganhos de desempenho na previsão da demanda.

A abordagem LSTM, aliada à interpretabilidade fornecida pelo SHAP, demonstrou ser especialmente adequada para o contexto epidemiológico analisado. A partir da fundamentação teórica apresentada anteriormente, foi possível observar como as redes recorrentes de memória longa conseguem capturar padrões complexos e defasagens temporais típicas dos surtos de dengue. Além disso, o uso de técnicas de interpretabilidade permitiu compreender o papel de variáveis como temperatura e umidade nas previsões, oferecendo maior transparência aos resultados. Essa combinação teórico-prática será explorada com mais profundidade no Capítulo 3, onde o modelo LSTM-Climate-Spatial é aplicado em um estudo real de previsão de surtos de dengue em diferentes estados do Brasil.

Ademais, a integração de dados temporais e espaciais se mostrou estratégica para o entendimento de padrões de risco em áreas urbanas, como discutido na fundamentação teórica sobre séries espaço-temporais e análise com GIS. A aplicação dessas técnicas permitiu identificar regiões com maior propensão a acidentes,

considerando tanto variações horárias quanto características geográficas. No Capítulo 3, tais modelos são aplicados à realidade urbana com o objetivo de subsidiar políticas públicas voltadas à prevenção de acidentes, evidenciando o potencial preditivo e estratégico da abordagem adotada.

3 ESTUDOS DE CASO E DISCUSSÃO – SÉRIES TEMPORAIS

3.1 ESTUDO DE CASO 1: PREVISÃO DE SURTOS DE DENGUE COM REDES LSTM E EXPLICABILIDADE SHAP

A dengue representa um dos maiores desafios de saúde pública no Brasil, com surtos recorrentes e impactos significativos nas regiões urbanas e periurbanas (Bhatt et al., 2013). Diante deste cenário, Chen e Moraga (2025) propuseram um modelo preditivo com redes LSTM (*Long Short-Term Memory*), enriquecido com variáveis climáticas defasadas e efeitos espaciais de estados vizinhos, além do uso do método SHAP (*SHapley Additive exPlanations*) para a seleção e interpretação das variáveis mais relevantes.

O objetivo do estudo foi construir um modelo escalável e adaptável, capaz de prever casos de dengue com antecedência de 4 e 12 semanas em todos os 27 estados do Brasil. O modelo considera informações climáticas com defasagens temporais, padrões sazonais anuais e dados históricos de dengue de estados vizinhos, capturando assim a complexidade das dinâmicas de transmissão da doença. A proposta se diferencia de modelos tradicionais, como ARIMA e SARIMA, ao lidar de forma mais eficaz com não linearidades e com a integração de múltiplas fontes de dados em um modelo unificado.

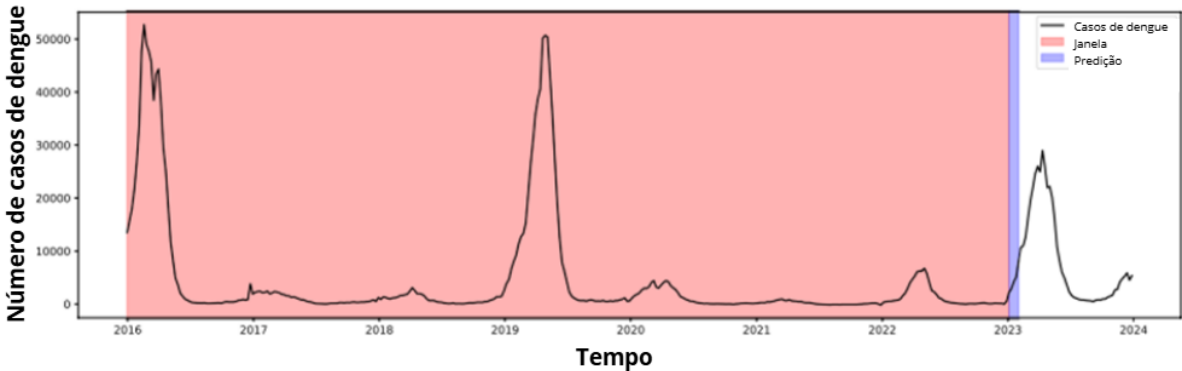
A base de dados utilizada no estudo inclui os casos semanais de dengue obtidos da plataforma InfoDengue (Codeco et al., 2018), combinados com dados climáticos do banco Copernicus ERA5 (C3S, 2017) (Tabela 1), que oferece estimativas semanais de temperatura, umidade, precipitação, entre outros fatores ambientais. Os dados foram processados por meio de uma estratégia de janela deslizante com sete anos de histórico (Figura 1), permitindo que o modelo fosse atualizado dinamicamente a cada nova semana de previsão. A arquitetura do modelo foi composta por uma camada LSTM com mil unidades, camadas densas de saída, função de perda MSE, otimizador Adam e regularização com *dropout*.

Tabela 1 – Variáveis climáticas semanais derivadas da base Copernicus ERA5 utilizadas no modelo LSTM

Variável	Unidade	Descrição
Minimum Temperature	°C	Temperatura mínima registrada na semana.
Median Temperature	°C	Temperatura mediana da semana.
Maximum Temperature	°C	Temperatura máxima registrada na semana.
Minimum Precipitation Rate	mm/h	Menor taxa horária de precipitação registrada na semana.
Median Precipitation Rate	mm/h	Mediana da taxa horária de precipitação registrada na semana.
Maximum Precipitation Rate	mm/h	Maior taxa horária de precipitação registrada na semana.
Total Precipitation	mm	Precipitação acumulada ao longo da semana.
Minimum Atmospheric Pressure	atm	Pressão atmosférica mínima registrada ao nível do mar durante a semana.
Median Atmospheric Pressure	atm	Mediana da pressão atmosférica ao nível do mar durante a semana.
Maximum Atmospheric Pressure	atm	Pressão atmosférica máxima registrada ao nível do mar durante a semana.
Minimum Relative Humidity	%	Umidade relativa mínima registrada na semana.
Median Relative Humidity	%	Mediana da umidade relativa registrada na semana.
Maximum Relative Humidity	%	Umidade relativa máxima registrada na semana.
Thermal Range	°C	Diferença entre as temperaturas máxima e mínima diárias.
Rainy Days	Dias	Número de dias com precipitação acima de 0,03 mm durante a semana.

Fonte: Adaptado de Chen e Moraga (2025).

Figura 1 – Estratégia de janela móvel utilizada para o treinamento e previsão dos casos de dengue entre 2016 e 2023



Fonte: Chen e Moraga (2025).

O SHAP foi aplicado para selecionar, de forma automática e específica para cada estado, as cinco variáveis climáticas mais relevantes, eliminando redundâncias

e aprimorando a capacidade explicativa do modelo. Ademais, entre as variáveis mais frequentes destacaram-se a umidade relativa máxima, a precipitação mínima e a temperatura mínima, o que permitiu adaptar o modelo às especificidades epidemiológicas e climáticas regionais, como demonstrado no exemplo de Minas Gerais, onde tais variáveis foram determinantes para a dinâmica da transmissão.

Quatro modelos foram comparados: LSTM-Cases (apenas dados de casos), LSTM-Climate (com dados climáticos e sazonais), LSTM-*Climate-Spatial* (modelo proposto, incluindo variáveis espaciais) e um modelo bayesiano com efeitos aleatórios. Os resultados demonstraram que o modelo LSTM-*Climate-Spatial* apresentou desempenho superior em grande parte dos estados, com menor MAE, MAPE e CRPS. Por exemplo, em Minas Gerais, o MAE caiu de 7730,47 para 5088,71 com a inclusão dos efeitos espaciais, e o CRPS de 1648,53 para 1035,86, indicando ganho significativo em acurácia.

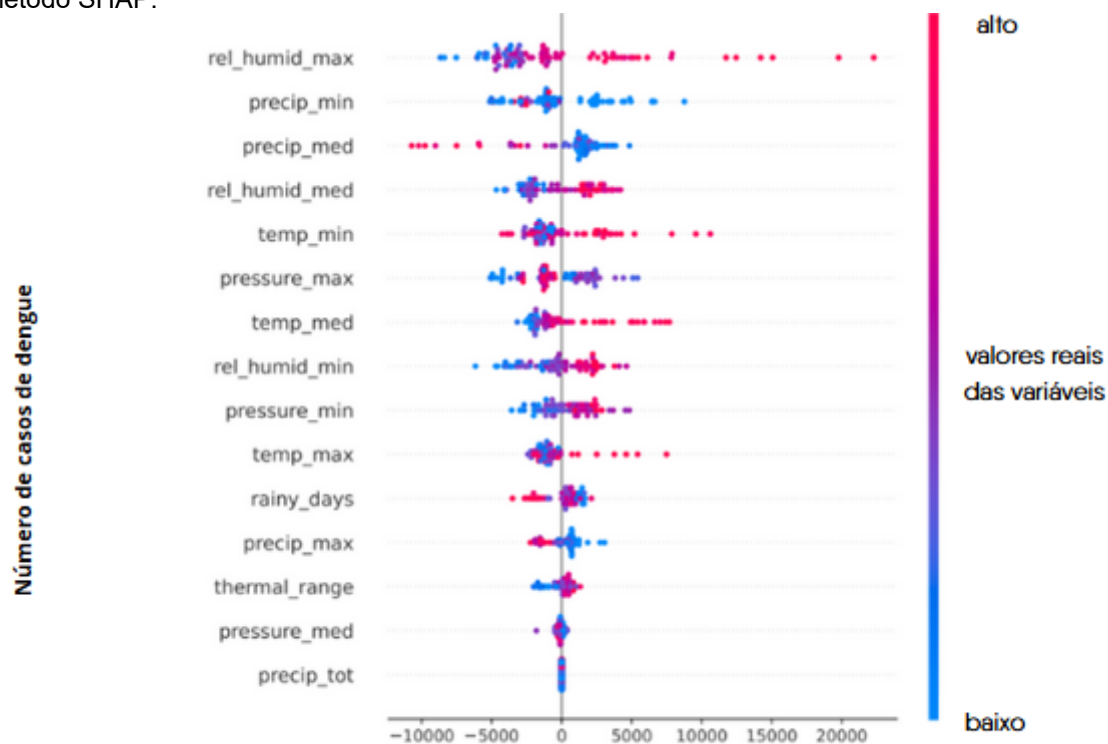
A inclusão de dados de estados vizinhos mostrou-se relevante em regiões com alta conectividade, como Minas Gerais, Paraná, Ceará e Goiás. Nessas regiões, a mobilidade populacional e as semelhanças climáticas aumentam a interdependência na propagação do vírus, favorecendo a integração espacial como elemento relevante do modelo. Por outro lado, em estados mais isolados da região Norte, como Acre, Roraima e Amapá, a inclusão dos efeitos espaciais gerou pequeno aumento de erro, indicando que, nesses contextos, a baixa densidade populacional e as barreiras geográficas limitam a utilidade desse tipo de informação.

Mesmo nas regiões onde os efeitos espaciais não trouxeram ganhos significativos, o uso de variáveis climáticas defasadas resultou em melhorias na previsão. O modelo LSTM-*Climate*, por exemplo, superou o LSTM-Cases em todos os estados, demonstrando que as condições ambientais são determinantes para a proliferação do mosquito e, portanto, importantes para a previsão da doença. Nesse contexto, tais resultados reforçam a importância da integração de múltiplas dimensões dos dados (climática, temporal e espacial) no desenvolvimento de modelos epidemiológicos robustos.

A discussão do estudo enfatiza que o sucesso do modelo LSTM-*Climate-Spatial* decorre de sua capacidade de capturar relações complexas e multifatoriais, principalmente quanto à forma como múltiplos fatores climáticos e espaciais interagem entre si para influenciar a transmissão da dengue, de maneira não linear, dinâmica e

regionalmente variável. Dessa forma, o uso do SHAP possibilitou uma melhor interpretabilidade, identificando os principais *drivers* climáticos regionais (Figura 2).

Figura 2 – Importância das variáveis climáticas para previsão de dengue em Minas Gerais, segundo o método SHAP.

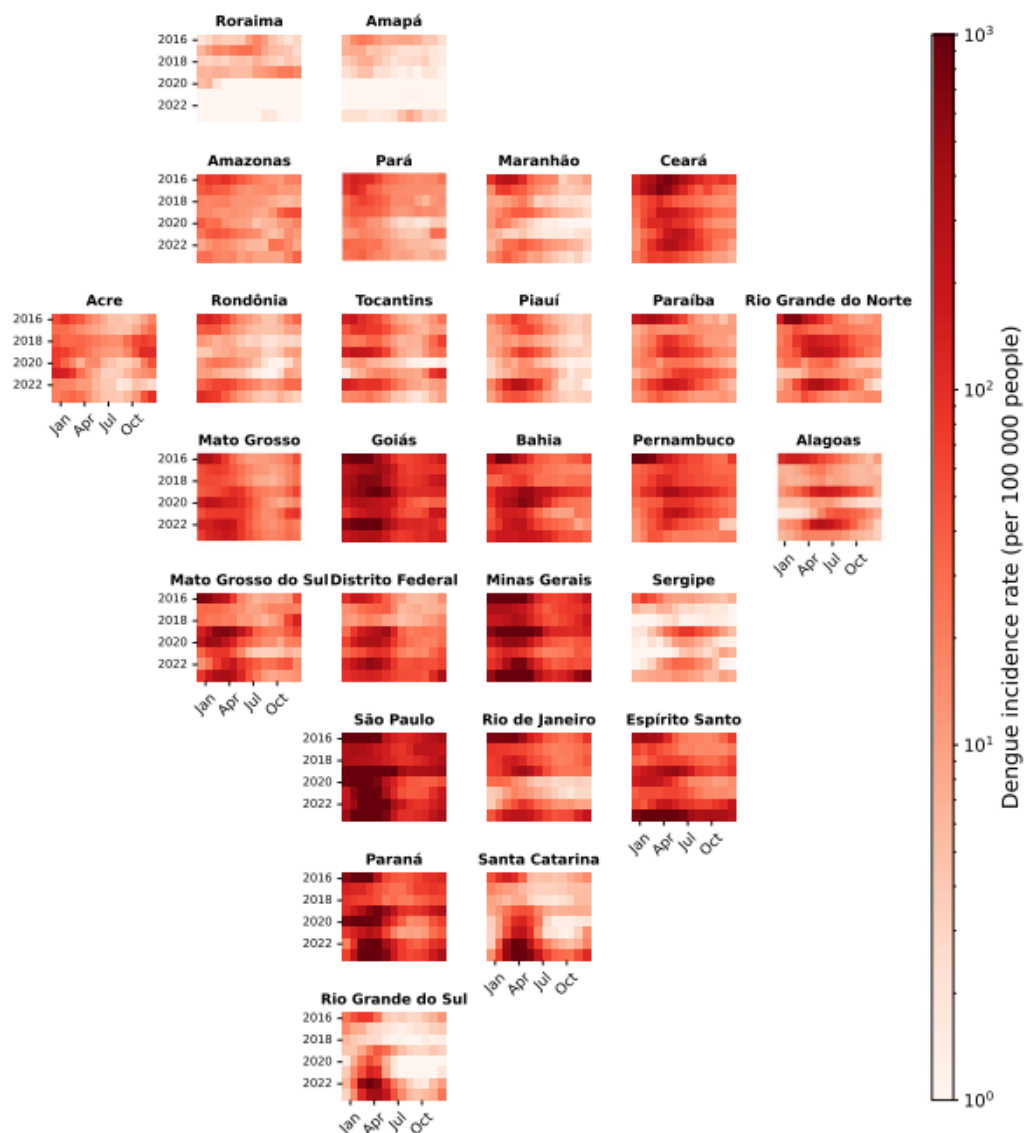


Na figura 2, cada ponto representa uma predição semanal, com a cor indicando o valor da variável (de baixo em azul a alto em rosa). O eixo horizontal mostra o impacto de cada variável na saída do modelo: quanto maior o valor absoluto do SHAP, maior a influência da variável na previsão. Observa-se que a umidade relativa máxima (*rel_humid_max*), a precipitação mínima (*precip_min*) e a precipitação mediana (*precip_med*) são as variáveis com maior impacto no modelo preditivo, sugerindo que fatores relacionados à umidade e à chuva são os principais determinantes da dinâmica da dengue em Minas Gerais. Dessa forma, compreende-se que esse tipo de análise permite melhorar a acurácia da previsão, assim como interpretar de forma transparente os fatores climáticos que contribuem para os surtos da doença.

Além disso, a inclusão de dados de estados vizinhos ampliou a capacidade preditiva do modelo em regiões com alta conectividade, enquanto a modelagem

sazonal permitiu captar padrões cíclicos mesmo não diretamente explicados pelas variáveis climáticas (Figura 3).

Figura 3 – Taxa mensal de incidência de dengue (por 100 mil habitantes) nos estados brasileiros entre 2016 e 2023



Fonte: Chen e Moraga (2025).

A Figura 3 apresenta a taxa mensal de incidência de dengue (por 100 mil habitantes) nos estados brasileiros entre os anos de 2016 e 2023, evidenciando padrões temporais e espaciais da doença. Cada quadrado representa a intensidade da incidência por mês, onde tons mais escuros indicam maior número de casos. Observa-se que estados como Goiás, Minas Gerais, São Paulo, Paraná e Ceará apresentaram picos recorrentes e intensos, especialmente em determinados períodos

do ano, indicando a presença de sazonalidade bem definida, sendo um padrão reforça a importância da modelagem sazonal no processo preditivo, já que permite capturar repetições cíclicas de surtos mesmo quando não são diretamente explicadas por variáveis climáticas. A variação entre os estados também justifica a inclusão de dados de vizinhança no modelo, pois a propagação do vírus pode seguir rotas regionais ligadas à mobilidade e à conectividade entre áreas urbanas próximas.

Ademais, ressalta-se que a escolha por defasagens climáticas de 1 a 3 meses permitiu capturar efeitos acumulativos das condições ambientais sobre os ciclos do vetor *Aedes aegypti*. Nesse sentido, é importante ressaltar que essa abordagem é útil para autoridades de saúde pública, pois permite antecipar surtos com antecedência suficiente para implementar ações preventivas. Contudo, o estudo de Chen e Moraga (2025) reconhece limitações como possíveis subnotificações nos dados de casos e a ausência de dados de mobilidade humana mais refinados, que poderiam aprimorar a modelagem espacial.

Estudos anteriores, como o de Silva (2024), concentraram-se em análises descritivas de municípios específicos, como Foz do Iguaçu, sem incorporar interações entre diferentes regiões ou variáveis espaciais mais amplas. De forma semelhante, embora o estudo de Johansen, Carmo e Alves (2015) tenha utilizado técnicas de análise espacial para mapear os casos de dengue em Campinas, sua abordagem permaneceu restrita à escala intramunicipal, sem extrapolar para articulações intermunicipais ou regionais. Tais abordagens, embora valiosas para compreender a realidade local, apresentam limitações ao desconsiderar os fluxos populacionais, os fatores climáticos regionais e as conexões entre territórios, que são importantes para a modelagem robusta da propagação da dengue em um país com alta mobilidade e diversidade climática como o Brasil.

3.2 ESTUDO DE CASO 2: PREVISÃO DE DEMANDA NO VAREJO

A previsão de demanda no varejo é uma atividade importante para a gestão eficiente da cadeia de suprimentos, especialmente em contextos de colaboração limitada entre os elos da cadeia. O estudo de Carbonneau, Laframboise e Vahidov (2008) investigou o desempenho de diferentes técnicas de aprendizado de máquina na previsão de sinais de demanda distorcidos, utilizando dados simulados e reais. O

estudo parte do reconhecimento de que, mesmo quando a demanda do consumidor final segue padrões previsíveis, a falta de integração e compartilhamento de informações entre os participantes da cadeia leva à distorção do sinal de demanda – o conhecido efeito chicote (*bullwhip effect*).

Na cadeia de suprimentos estendida, cada elo (varejista, atacadista, distribuidor, fabricante) transforma o sinal de demanda com base em seus próprios métodos de previsão, tempo de resposta e decisões internas. Na tentativa de reduzir esse problema, os autores propuseram o uso de técnicas avançadas de aprendizado de máquina, como redes neurais (NN), redes neurais recorrentes (RNN) e máquinas de vetor de suporte (SVM), comparando-as com métodos tradicionais como média móvel, regressão linear e previsão ingênua.

A simulação (Figura 4) da demanda do consumidor final foi baseada em uma função senoidal com ruído branco, criando variações que simulam sazonalidade e aleatoriedade. Com isso, o modelo simulou quatro elos da cadeia, todos sujeitos a atrasos nos pedidos e entregas, permitindo a análise dos efeitos acumulativos da distorção na demanda até o fabricante.

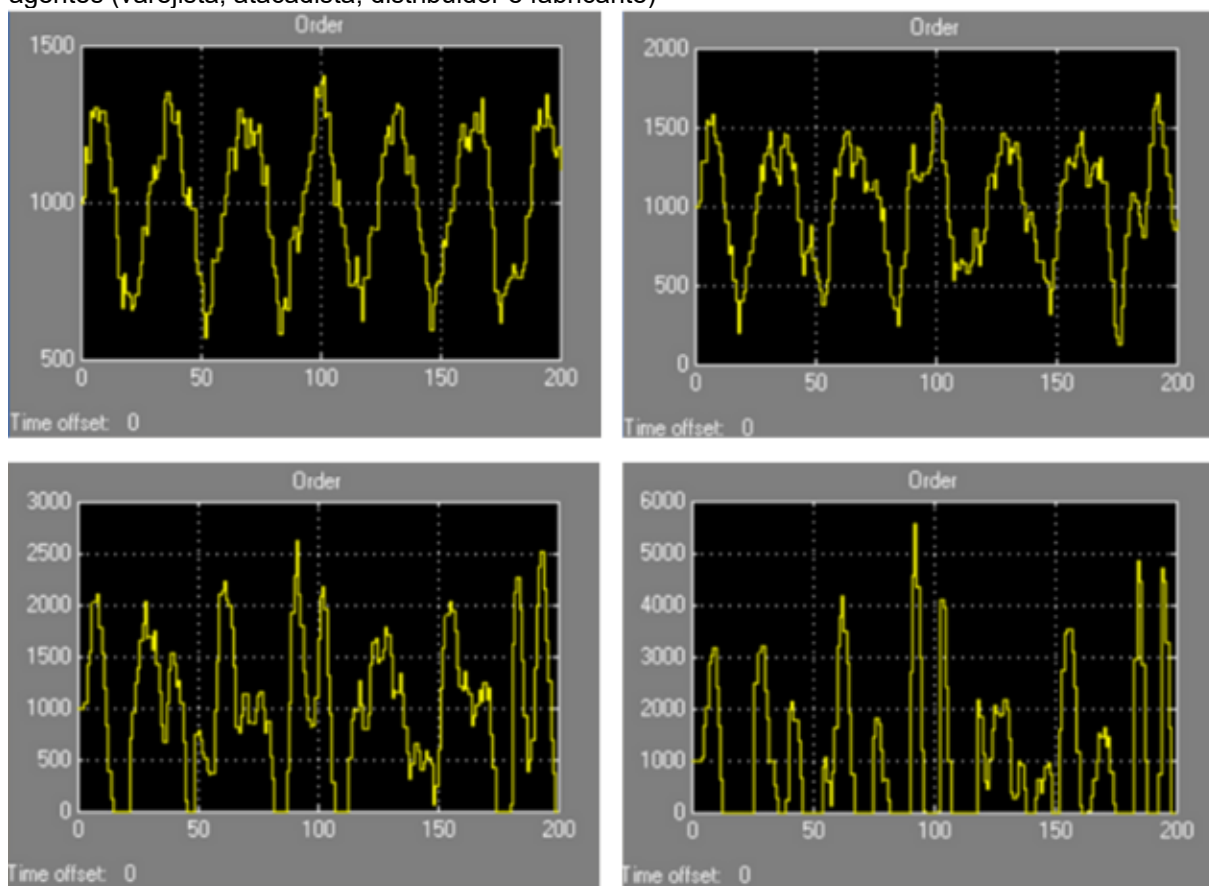
Figura 4 – Exemplo de estrutura de simulação de parceiros da cadeia

Função	Descrição
Previsão de demanda	Regressão linear sobre os últimos 10 dias
Cálculo do pedido	Considera <i>backlog</i> e recebimento de mercadorias
Processamento do pedido	Atualiza inventário, entrega e pedidos pendentes

Fonte: Carbonneau, Laframboise e Vahidov (2008).

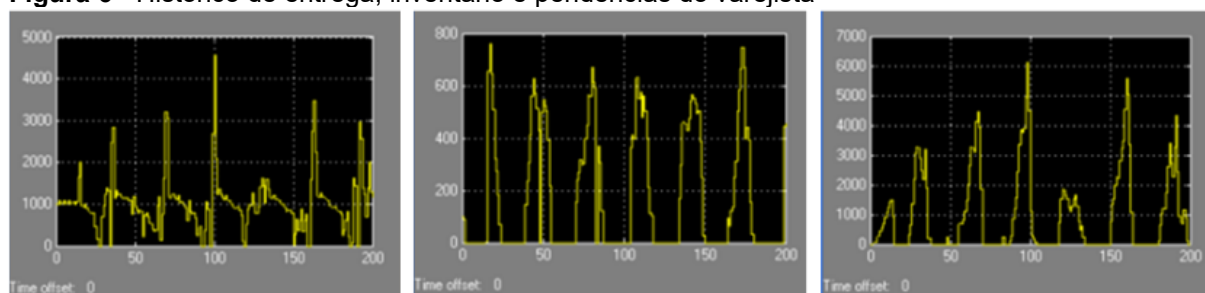
O histórico de pedidos mostra o agravamento do efeito chicote à medida que a demanda percorre a cadeia de suprimentos (Figura 5), passando de um sinal estável no varejo para um padrão caótico no fabricante. Assim, esse comportamento gera impactos operacionais diretos, como aumento de backlog, oscilações no inventário e entregas irregulares (Figura 6).

Figura 5 - Efeito da distorção da demanda ao longo da cadeia de suprimentos entre os diferentes agentes (varejista, atacadista, distribuidor e fabricante)



Fonte: Carbonneau, Laframboise e Vahidov (2008).

Figura 6 - Histórico de entrega, inventário e pendências do varejista



Fonte: Carbonneau, Laframboise e Vahidov (2008).

Acerca dos conjuntos de dados utilizados, foram utilizados dois conjuntos: um proveniente da simulação e outro real, com dados mensais de pedidos da indústria de fundição canadense (*Foundries*), obtidos da *Statistics Canada*. Ambos os conjuntos foram preparados com defasagens temporais e variáveis de entrada ajustadas para capturar variações relativas (Tabelas 1 e 2). Os modelos foram treinados e testados com os mesmos conjuntos para garantir comparabilidade.

Tabela 1 - Amostra de observações para previsão de dados simulados

Dia	Demanda	Variação nos próximos 3 dias	Defasagem de 4 dias	Defasagem de 3 dias	Defasagem de 2 dias	Defasagem de 1 dia	Hoje
1	1000	37,8	—	—	—	—	—
2	1000	426,7	—	—	—	—	0
3	1000	537,1	—	—	—	0	0
4	1037,8	979,3	—	—	0	0	37,8
5	1426,7	595,2	—	0	0	37,8	388,9
6	1537,1	567,4	—	0	0	37,8	388,9
7	2017,1	244,4	0	37,8	388,9	110,4	480
8	2021,9	489,8	37,8	388,9	110,4	480	4,8
9	2104,5	909,8	388,9	110,4	480	4,8	82,6
10	1772,7	937,8	110,4	480	4,8	82,6	331,8
11	1532,1	1145,77	480	4,8	82,6	331,8	240,6
12	1194,7	1084,94	4,8	82,6	331,8	240,6	337,4
13	834,9	834,9	82,6	331,8	240,6	337,4	359,8
14	386,33	386,33	331,8	240,6	337,4	359,8	448,57
15	109,76	109,76	240,6	337,4	359,8	448,57	276,57

Fonte: Adaptado de Carbonneau, Laframboise e Vahidov (2008).

Tabela 2 - Amostra do conjunto de dados das fundições

Mês	Demanda	Variação relativa para o mês seguinte	Defasagem				
			4 meses	3 meses	2 meses	1 mês	Este mês
Jan-92	127.712	0,055124	—	—	—	—	—
Fev-92	134.752	0,104154	—	—	—	—	0,055124
Mar-92	148.787	0,002971	—	—	—	0,055124	0,104154
Abr-92	149.229	0,054661	—	—	0,055124	0,104154	0,002971
Mai-92	157.386	0,037024	—	0,055124	0,104154	0,002971	0,054661
Jun-92	163.213	0,188392	0,055124	0,104154	0,002971	0,054661	0,037024
Jul-92	132.465	0,108678	0,104154	0,002971	0,054661	0,037024	0,188392
Ago-92	146.861	0,094797	0,002971	0,054661	0,037024	0,188392	0,108678
Set-92	160.783	0,016289	0,054661	0,037024	0,188392	0,108678	0,094797
Out-92	158.164	0,027579	0,037024	0,188392	0,108678	0,094797	0,016289
Nov-92	153.802	0,134933	0,188392	0,108678	0,094797	0,016289	0,027579
Dez-92	133.049	0,150929	0,108678	0,094797	0,016289	0,027579	0,134933
Jan-93	153.130	0,033063	0,094797	0,016289	0,027579	0,134933	0,150929
Fev-93	158.193	0,128356	0,016289	0,027579	0,134933	0,150929	0,033063
Mar-93	178.498	0,072415	0,027579	0,134933	0,150929	0,033063	0,128356

Fonte: Adaptado de Carbonneau, Laframboise e Vahidov (2008).

As previsões geradas pelas redes neurais (NN), redes recorrentes (RNN) e SVM foram avaliadas e comparadas com as técnicas tradicionais. As tabelas 3 e 4 mostram a acurácia dos modelos nos conjuntos de teste (simulado e real), destacando que as RNNs e SVMs apresentaram menor erro médio absoluto (MAE), mas sem diferença estatística significativa em relação à regressão linear múltipla (MLR).

Tabela 3 - Comparação de desempenho das técnicas (MAE – Conjunto Simulado)

Técnica	MAE (Teste)
RNN	447,72
SVM (LS-SVM)	453,04
MLR (Reg. Linear)	453,22
NN	455,41
Naïve	520,53
Média Móvel	526,61
Tendência (Trend)	618,02

Fonte: Carbonneau, Laframboise e Vahidov (2008).

Tabela 4 - Comparação de desempenho das técnicas (MAE – Conjunto Real – Foundries)

Técnica	MAE (Teste)
RNN	20,35
SVM (LS-SVM)	20,48
MLR (Reg. Linear)	21,07
NN	25,26
Média Móvel	25,48
Tendência (Trend)	27,32
Naïve	32,59

Fonte: Carbonneau, Laframboise e Vahidov (2008).

Assim, os resultados apontam que técnicas avançadas como RNN e SVM possuem potencial para melhorar a acurácia da previsão de demanda, especialmente em contextos com dados reais e mais ruídos. Contudo, essa superioridade não foi estatisticamente significativa quando comparada à regressão linear múltipla, que se mostrou uma alternativa mais simples e competitiva. Métodos como média móvel, tendência e *naïve forecasting* foram os menos eficazes.

Portanto, a escolha do modelo deve considerar o desempenho preditivo, assim como a simplicidade e viabilidade computacional. Desta forma, a previsão de demanda utilizando aprendizado de máquina, mesmo sem colaboração plena entre os elos da cadeia, pode reduzir o efeito chicote e melhorar o planejamento logístico. Ressalta-se que estudos futuros podem integrar dados de mobilidade e estratégias colaborativas para refinar ainda mais a previsão.

3.3 ESTUDO DE CASO 3: ANÁLISE DE TRÂNSITO PARA PREVENÇÃO DE ÁREAS DE RISCO

A segurança no trânsito constitui um desafio global de grande magnitude, responsável por milhares de mortes e ferimentos graves todos os anos, além de consideráveis impactos econômicos e sociais. No contexto do Catar, o aumento da frota de veículos e a rápida urbanização intensificaram a complexidade da mobilidade urbana, exigindo políticas mais eficazes de prevenção e mitigação de acidentes. Nesse cenário, compreender os padrões espaciais e temporais dos acidentes viários se torna fundamental para o planejamento de um sistema de transporte mais seguro e sustentável (Mohammed et al. 2023).

O presente estudo de caso baseia-se na pesquisa desenvolvida por Mohammed et al. (2023), a qual emprega técnicas de análise espacial e espaço-temporal para identificar áreas críticas de ocorrência de acidentes de trânsito, conhecidas como *hotspots*. A análise foi aplicada a dados de acidentes registrados no Catar entre os anos de 2015 e 2019, considerando variáveis como localização geográfica, tipo de via, gravidade do acidente, dia da semana e características populacionais. Entre os métodos utilizados, destacam-se o Space-Time Cube, o Índice de *Moran* Local, o modelo de regressão ponderada geograficamente (*Geographically Weighted Regression – GWR*) e a estatística *Getis-Ord Gi**.

Os resultados indicaram uma forte concentração dos acidentes na região centro-leste do país, com padrões de aumento e intensificação especialmente evidentes no ano de 2019. Ressalta-se que tal concentração está fortemente associada a comportamentos de risco por parte dos condutores, bem como à infraestrutura viária e à dinâmica demográfica da região. Como implicações práticas, o estudo subsidia a formulação de políticas públicas mais precisas, voltadas à expansão de rodovias, requalificação de trechos críticos, monitoramento contínuo e alocação eficiente de recursos para resposta emergencial.

Com base no estudo de Mohammed et al. (2023), a pesquisa adotou uma abordagem metodológica estruturada para analisar acidentes de trânsito a partir de dados espaciais e temporais. A Figura 7 apresenta o diagrama metodológico da pesquisa, que sintetiza as etapas desde a identificação do problema até a definição de implicações políticas com base na identificação de áreas de risco. O processo inclui

a coleta e tratamento de dados, análise descritiva espacial, aplicação de métodos estatísticos como *Space-Time Cube* e *Geographically Weighted Regression*, e, por fim, a formulação de estratégias preventivas.

Figura 7 - Diagrama metodológico da pesquisa

Etapa	Descrição
Identificação do problema	Coleta e pré-processamento de dados.
Preparação do conjunto de dados	Inclusão de variáveis espaciais (ex.: município do acidente, coordenadas, zona) e temporais (ex.: dia da semana, época do ano), além da gravidade, causa e vítimas.
Análise descritiva espacial	Análise exploratória dos dados para entender padrões de distribuição espacial e temporal dos acidentes.
Métodos estatísticos espaciais	Aplicação de técnicas para identificação de <i>hotspots</i> : – The space-time cube – Local <i>Moran's I</i> – Getis-Ord <i>Gi*</i> – Geographically Weighted Regression (GWR)
Identificação e verificação de hotspots	Confirmação das áreas críticas por meio da sobreposição de resultados dos métodos aplicados.
Discussão de implicações políticas	Definição de medidas para tomada de decisão, como: – Expansão da malha viária – Crescimento populacional e de expatriados – Áreas de trabalho – Nova lei de trânsito (2016) – Melhoria nas respostas emergenciais.

Fonte: Mohammed et al. (2023).

Complementarmente, a Figura 8 apresenta a frequência de acidentes de trânsito no período de 2015 a 2019, detalhando a evolução anual dos registros policiais segundo a gravidade das ocorrências, o número de vítimas e indicadores de mobilidade urbana.

Figura 8 – Frequência de acidentes de trânsito (2015–2019) e variação percentual no período

Categoria	2015	2016	2017	2018	2019	% de Variação (2015–2019)
Acidentes de trânsito com veículos automotores registrados pela polícia						
Acidentes fatais	194	155	159	154	134	–31%
Acidentes graves	548	666	579	530	608	+11%
Acidentes leves	5.129	5.284	5.319	5.474	5.805	+13%
Total	5.871	6.105	6.057	6.158	6.547	+12%
Vítimas de acidentes de trânsito registradas pela polícia						
Vítimas fatais	227	178	177	168	154	–32%
Ferimentos graves	693	871	742	683	778	+12%
Ferimentos leves	7.488	8.028	7.962	8.116	8.396	+12%
Total	8.408	9.077	8.881	8.967	9.328	+11%
Outras estatísticas nacionais						
Quilometragem percorrida por veículos (em milhões de km)	30.936	33.097	34.818	36.306	43.593	+41%
População residente	2.335.068	2.520.621	2.614.626	2.662.845	2.712.959	+16%
Veículos registrados	1.352.979	1.447.478	1.522.733	1.587.815	1.655.676	+22%
Motoristas habilitados	1.234.350	1.340.271	1.441.594	1.538.407	1.625.339	+32%

Fonte: Mohammed et al. (2023).

A metodologia adotada no estudo de Mohammed et al. (2023) é baseada em uma análise espaço-temporal dos acidentes de trânsito no Estado do Qatar, com o objetivo de identificar padrões de ocorrência e auxiliar no planejamento de um sistema de transporte mais seguro e sustentável. O estudo utilizou dados oficiais da polícia local sobre acidentes reportados entre os anos de 2015 e 2019, abrangendo variáveis como a gravidade dos acidentes, causas prováveis, localização, e número de vítimas.

Inicialmente, os dados passaram por etapas de coleta, filtragem e pré-processamento, conforme ilustrado no diagrama metodológico da Figura 1. Em seguida, foi realizada uma análise descritiva espacial do conjunto de dados, com base em características espaciais (zona, município e coordenadas) e temporais (dias da semana e época do ano), além de variáveis relacionadas à gravidade, causa e vítimas do acidente.

Para identificação dos *hotspots*, áreas de maior concentração de acidentes, foram aplicadas técnicas estatísticas espaciais, como o Space-Time Cube, o índice local de *Moran*, a estatística *Getis-Ord Gi** e a Regressão Geograficamente Ponderada (GWR). O uso de tais ferramentas permitiram detectar padrões emergentes, persistentes ou decrescentes na distribuição dos acidentes ao longo do tempo e do espaço.

Após a identificação e validação das zonas críticas, foi realizada a discussão das implicações políticas e de gestão, considerando fatores como expansão da malha viária, crescimento populacional, zonas de trabalho, legislação de trânsito vigente e capacidade de resposta emergencial.

3.3.1 Evolução temporal e distribuição dos acidentes

A análise espaço-temporal dos acidentes de trânsito no Estado do Catar entre os anos de 2015 e 2019 revelou uma tendência de crescimento no número total de ocorrências, passando de 5.871 casos registrados em 2015 para 6.547 em 2019, o que representa um aumento de 12%. Embora esse acréscimo global aparente seja moderado, ao desagregar os dados por gravidade observa-se uma redução expressiva nos acidentes fatais (queda de 31%) e um crescimento nos acidentes leves (aumento de 13%), o que podem indicar avanços na infraestrutura viária e nos sistemas de resposta emergencial, resultando na mitigação da gravidade dos sinistros, mesmo diante do aumento da frota e da mobilidade urbana.

Ao mesmo tempo, as estatísticas revelam uma significativa elevação nos indicadores de exposição ao risco, como o aumento de 41% nos quilômetros rodados (VKT) e de 22% no número de veículos registrados. Assim, esses fatores sugerem que, apesar do crescimento demográfico e da motorização, medidas de segurança viária vêm sendo capazes de conter parte dos impactos negativos, especialmente os mais graves.

Do ponto de vista geográfico, a distribuição espacial dos acidentes em 2015 e 2019 evidenciaram uma concentração significativa na região centro-leste do país, especialmente na capital Doha. Em 2019, as ocorrências se apresentaram de forma ainda mais concentrada, sugerindo padrões sistemáticos e persistentes de risco viário

ao longo do tempo, possivelmente relacionados à densidade populacional, ao uso intensivo do solo e à complexidade da malha viária urbana.

3.3.2 Gravidade dos acidentes e impacto sobre a população

A análise da gravidade dos acidentes de trânsito no Catar no período de 2015 a 2019 revela importantes transformações na severidade dos eventos e em seus efeitos sobre a população. Apesar do aumento no número total de acidentes e vítimas, a proporção de acidentes fatais apresentou tendência de queda, reduzindo-se em 31%. Ressalta-se que esse dado é corroborado pela redução de 32% no número de óbitos registrados no mesmo intervalo (de 227 em 2015 para 154 em 2019), o que pode refletir uma melhoria na capacidade de resposta dos serviços de emergência, no uso de tecnologias de segurança automotiva e na eficácia de medidas educativas e punitivas.

Por outro lado, o número de acidentes com ferimentos graves aumentou 11%, passando de 548 para 608 registros, e o número de feridos graves aumentou 12%, de 693 para 778 vítimas. O crescimento das lesões graves, em contraste com a queda dos óbitos, indica uma necessidade de atenção aos fatores que agravam a severidade dos acidentes, como velocidade excessiva, ausência do uso de equipamentos de proteção (como cinto de segurança e capacete), e possíveis lacunas na fiscalização viária.

Com isso, esses dados apontam que, embora políticas públicas estejam surtindo efeito na redução da letalidade, ainda há desafios significativos na prevenção de ferimentos incapacitantes e na mitigação dos custos sociais e econômicos dos acidentes. O aumento das lesões leves (de 7.488 para 8.396 casos) também sugere a persistência de comportamentos de risco no trânsito, mesmo que com consequências menos graves.

Além disso, deve-se destacar que o aumento da população residente (16%) e do número de motoristas licenciados (32%) contribui para maior pressão sobre o sistema viário, o que exige a adoção de estratégias integradas que considerem o planejamento urbano, a educação para o trânsito e a expansão da infraestrutura rodoviária com foco na prevenção e na segurança.

3.3.3 Padrões espaciais e identificação de *hotspots*

A análise espaço-temporal revelou que a distribuição dos acidentes de trânsito no Catar não ocorre de forma aleatória, mas apresenta padrões espacialmente concentrados, sobretudo nas zonas urbanas de Doha e seus arredores. Entre 2015 e 2019, observou-se uma intensificação da frequência de acidentes nas zonas ao norte e oeste de Doha, regiões que passaram por forte crescimento populacional, urbanização acelerada e ampliação da malha viária. Por outro lado, zonas ao sul da capital apresentaram redução nos índices, possivelmente relacionada à mudança nas rotas de transporte terrestre após o bloqueio imposto pela Arábia Saudita em 2017.

As ferramentas estatísticas aplicadas, como o Índice de *Moran* e a análise *Getis-Ord Gi**, evidenciaram uma autocorrelação espacial positiva significativa, especialmente nos dias de semana, indicando a existência de áreas com padrão de aglomeração de acidentes (*hotspots*). Essa tendência aponta para uma distribuição não aleatória, mas sim agrupada, reforçando a necessidade de intervenções localizadas. Áreas mistas (residenciais e comerciais) com tráfego intenso e presença de vias arteriais, como as do centro e oeste de Doha, destacaram-se como zonas críticas.

A análise de *hotspots* emergentes mostrou também que algumas áreas passaram de padrões esporádicos para padrões persistentes e crescentes de acidentes, sugerindo um agravamento contínuo do problema. Portanto, esses dados são importantes para o planejamento urbano e para a alocação de recursos voltados à segurança viária, permitindo o direcionamento de ações preventivas, como sinalização mais eficaz, fiscalização intensificada, melhorias na iluminação pública e redução de velocidade.

Desta forma, vale destacar que a expansão do sistema metroviário, com estações inauguradas em 2019, pode ter influenciado os padrões de acidentes, especialmente nas zonas centrais. O monitoramento contínuo dessas mudanças, aliado à análise espaço-temporal, torna-se fundamental para avaliar os efeitos das intervenções e ajustar políticas públicas de mobilidade e segurança.

3.3.4 Variações temporais e perfil dos acidentes

A análise temporal revelou que os acidentes de trânsito no Catar apresentam variações marcantes ao longo do tempo, especialmente entre os dias da semana e os fins de semana. Em 2015, os acidentes estavam majoritariamente concentrados na região sudoeste de Doha durante os dias úteis, refletindo padrões de deslocamento casa-trabalho e a maior densidade de tráfego nesse período. Em 2019, essas ocorrências se intensificaram e se espalharam para áreas anteriormente menos impactadas, como as zonas norte e oeste, indicando não apenas o crescimento urbano nessas regiões, mas também a possível insuficiência de infraestrutura viária e fiscalização diante do aumento da demanda.

As diferenças entre dias úteis e fins de semana também foram significativas. Durante a semana, os acidentes mostraram-se mais clusterizados, com maior densidade e padrão consistente, enquanto nos fins de semana, o padrão tendeu à aleatoriedade. Assim, isso pode estar relacionado ao menor volume de tráfego e à mudança nos motivos das viagens, que passam de deslocamentos obrigatórios (trabalho, escola) para atividades de lazer e consumo.

Além disso, a análise sazonal mostrou que os meses de clima mais ameno, especialmente entre dezembro e fevereiro, registraram maior número de pedestres nas vias, contribuindo para o aumento de acidentes com vítimas dessa categoria. O calor extremo do verão, por sua vez, tende a inibir a circulação de pedestres, alterando o perfil dos envolvidos nos acidentes. Essa dinâmica confirma a influência direta das condições climáticas no comportamento dos usuários das vias.

No que diz respeito à natureza dos acidentes, observou-se uma redução significativa nos acidentes fatais entre 2015 e 2019, especialmente nas regiões em processo de urbanização. Tal diminuição pode estar associada tanto ao aumento da densidade populacional, que naturalmente reduz a velocidade média dos veículos, quanto à melhoria nos serviços de resposta emergencial. Entretanto, o número total de acidentes aumentou no período, indicando que, embora mais acidentes sejam não fatais, a frequência geral continua a representar um desafio.

Assim, tais dados temporais reforçam a importância de políticas públicas sensíveis ao tempo e ao espaço, que considerem os padrões de deslocamento da

população, as variações climáticas e os perfis dos usuários da via no planejamento de intervenções mais eficazes.

3.3.5 Fatores associados e implicações para a formulação de políticas públicas

A identificação dos fatores associados à distribuição espacial e temporal dos acidentes de trânsito em Doha revelou padrões relevantes para o desenvolvimento de políticas públicas mais eficazes e direcionadas. Entre os principais fatores, destaca-se a falha humana, como julgamentos inadequados, direção imprudente e desrespeito às normas de trânsito, responsável por mais da metade dos acidentes registrados tanto em 2015 quanto em 2019. A persistência desse fator sugere que medidas educativas isoladas podem não ser suficientes e que é necessário reforçar a fiscalização e o uso de tecnologias de monitoramento viário.

A análise espacial evidenciou zonas com alta concentração de acidentes envolvendo pedestres, especialmente em áreas residenciais de baixa renda no sudoeste de Doha. Tais locais abrigam, em sua maioria, trabalhadores migrantes que se deslocam a pé, o que os torna mais vulneráveis. Estudos como o de Loukaitou-Sideris et al. (2007) corroboram a relação entre baixos níveis socioeconômicos e maior risco de acidentes com pedestres, apontando para a urgência de políticas inclusivas de mobilidade urbana, como a ampliação de calçadas, travessias seguras e sinalizações adequadas nesses territórios.

Outro ponto relevante identificado foi o impacto da expansão urbana e da infraestrutura viária no aumento da frequência dos acidentes em zonas anteriormente menos afetadas. O crescimento rápido e desordenado, sem o devido planejamento em segurança viária, gerou novos pontos críticos, como observado nas regiões norte e oeste da cidade. A análise de *hotspots* emergentes indica que áreas de uso misto (residencial e comercial), com vias arteriais e alta densidade de tráfego, são mais propensas à intensificação dos acidentes. Tais achados sustentam a importância da integração entre planejamento urbano e políticas de trânsito.

As ferramentas estatísticas utilizadas, como o índice de *Moran* e a análise *Getis-Ord Gi**, permitiram não apenas localizar os agrupamentos espaciais dos acidentes, mas também acompanhar sua evolução no tempo. Essa abordagem reforça a ideia de que os acidentes não ocorrem de maneira aleatória, mas seguem padrões que

podem ser previstos e, portanto, prevenidos. A incorporação dessas análises no processo decisório permite o monitoramento contínuo das intervenções e a avaliação de sua efetividade.

Nesse contexto, a adoção do marco do *Safe Systems*, que propõe a eliminação de fatalidades e lesões graves por meio de ações coordenadas em cinco frentes (gestão da segurança viária, vias seguras, veículos seguros, usuários seguros e resposta pós-acidente), apresenta-se como um caminho viável para o Catar. Intervenções adaptadas a cada zona, considerando os padrões locais de risco, podem gerar resultados mais efetivos com menor custo social.

Portanto, o presente estudo de caso, baseado na pesquisa de Mohammed et al. (2023), demonstrou a eficácia de técnicas de análise espaço-temporal na identificação de padrões de acidentes de trânsito em Doha, Catar, entre os anos de 2015 e 2019. A utilização combinada das ferramentas Time-Space Cube, Índice de *Moran* e *Getis-Ord Gi** permitiu compreender a dinâmica evolutiva dos acidentes, revelando áreas com alta concentração de ocorrências, conhecidas como *hotspots*, bem como as transformações ocorridas ao longo do tempo.

Os resultados evidenciaram que a região central e leste do país concentra a maior parte dos acidentes, sendo essa área também a mais populosa e com maior densidade de tráfego. O crescimento urbano, sobretudo nas zonas norte e oeste, associado à expansão da malha viária, contribuiu para o surgimento de novos pontos críticos, refletindo a importância de um planejamento urbano alinhado à segurança viária.

Ademais, fatores comportamentais, como a imprudência ao volante e o não cumprimento das normas de trânsito, se mostraram como as principais causas dos acidentes, reforçando a necessidade de medidas integradas que combinem educação, fiscalização e intervenção urbana. As áreas de baixa renda, onde se concentram trabalhadores migrantes, apresentaram vulnerabilidades acentuadas, especialmente em relação aos acidentes com pedestres, evidenciando desigualdades na infraestrutura urbana e na mobilidade segura.

Diante disso, o estudo reforça a relevância de políticas públicas embasadas em dados espaciais e temporais, que permitam intervenções localizadas e eficazes. A adoção do modelo *Safe Systems*, com ações voltadas para vias seguras, veículos

seguros, usuários conscientes e resposta pós-acidente, surge como uma alternativa promissora para a redução das mortes e lesões graves no trânsito.

Destaca-se que a aplicação contínua dessas metodologias pode auxiliar gestores e formuladores de políticas públicas na avaliação da efetividade das intervenções, promovendo melhorias constantes na segurança viária e contribuindo para um sistema de transporte mais sustentável e equitativo.

4 DISCUSSÃO DOS ESTUDOS DE CASO

4.1 ESTUDO DE CASO 1: PREVISÃO DE SURTOS DE DENGUE COM REDES LSTM E EXPLICABILIDADE SHAP

Diferentemente do modelo proposto por Chen e Moraga (2025), que integra variáveis climáticas defasadas, dados de estados vizinhos e modelagem sazonal com o auxílio de redes LSTM, o estudo de Silva et al. (2018) optou por uma abordagem regionalizada com redes neurais recorrentes (RNN), restringindo-se ao estado do Pará. O modelo implementado por Silva et al. se destaca por utilizar dados semanais de municípios específicos, como Belém, Altamira e Santarém, para antecipar surtos de dengue e disparar alertas automáticos, contribuindo para o planejamento de ações pelos gestores de saúde pública. Apesar de alcançar bons níveis de acurácia, a proposta apresenta limitações metodológicas em termos de generalização, por não incorporar fatores espaciais nem variáveis ambientais externas, como temperatura, umidade e precipitação, que são amplamente reconhecidas na literatura como determinantes para a proliferação do vetor *Aedes aegypti*.

Nesse sentido, o trabalho de Chen e Moraga (2025) representa um avanço considerável ao construir um modelo escalável, ajustável a todos os 27 estados brasileiros e capaz de considerar, por meio do uso de janelas temporais móveis, os efeitos acumulativos e sazonais das variáveis climáticas. Outro diferencial é o uso da técnica SHAP, que permite não apenas identificar as variáveis com maior impacto nas previsões, mas também interpretar a influência de cada uma em termos quantitativos e específicos para cada localidade. Essa camada de explicabilidade é fundamental no contexto da saúde pública, pois transforma modelos antes considerados "caixas-pretas" em ferramentas auditáveis e transparentes. Como reforça Cesaro (2021), a interpretabilidade é um requisito ético e técnico crescente no campo do aprendizado de máquina, sobretudo quando as previsões orientam decisões que impactam vidas humanas, como intervenções em epidemias.

Ademais, a integração entre modelagem climática, análise temporal e estrutura espacial evidencia uma compreensão sistêmica da dinâmica da dengue, o que falta em abordagens exclusivamente univariadas como a de Silva et al. (2018). Enquanto a RNN utilizada nesse trabalho regional capta bem a complexidade da série temporal,

ela não considera a interdependência entre municípios ou as rotas de mobilidade populacional, elementos que Chen e Moraga demonstram ser cruciais em regiões com alta conectividade.

Assim, embora ambos os estudos evidenciem a aplicabilidade de redes neurais na previsão epidemiológica, o modelo LSTM-Climate-Spatial se mostra metodologicamente mais robusto e alinhado às práticas atuais de ciência de dados em saúde, sobretudo por conciliar precisão, transparência e escalabilidade.

Para aprofundar a análise comparativa entre abordagens preditivas baseadas em redes neurais aplicadas à dengue, apresenta-se a seguir um quadro sinóptico (Figura 9) entre o estudo nacional de Chen e Moraga (2025) e a proposta regional de Silva et al. (2018).

Figura 9 – Comparação entre os estudos de Chen e Moraga (2025) e Silva et al. (2018)

Critério	Chen e Moraga (2025)	Silva et al. (2018)
Abrangência geográfica	Nacional – abrange todos os 27 estados do Brasil	Regional – focado no estado do Pará (Belém, Altamira, Parauapebas e Santarém)
Tipo de rede neural	LSTM (Long Short-Term Memory)	Redes Neurais Recorrentes (RNN)
Variáveis utilizadas	Casos históricos + variáveis climáticas (defasadas) + dados de estados vizinhos + modelagem sazonal	Apenas séries temporais dos casos de dengue
Fontes de dados	InfoDengue + ERA5 (dados climáticos de temperatura, umidade, precipitação, etc.)	Dados públicos semanais dos casos de dengue
Técnicas adicionais aplicadas	SHAP (para interpretabilidade e seleção de variáveis relevantes)	Módulo de alertas com base no aumento repentino de casos
Explicabilidade do modelo	Alta – uso do SHAP permite identificar e interpretar as variáveis com maior impacto	Baixa – modelo caixa-preta sem ferramenta de interpretabilidade
Aplicações práticas	Previsão antecipada com explicações regionais; útil para políticas públicas nacionais	Disparo de alertas locais para apoio à gestão municipal
Limitações apontadas	Subnotificação e ausência de dados refinados de mobilidade	Falta de integração com variáveis ambientais e ausência de contextualização espacial
Diferenciais	Integração climática, temporal e espacial em um modelo escalável e interpretável	Boa acurácia local com RNNs, mas sem articulação intermunicipal ou regional

A referida comparação evidencia avanços metodológicos recentes, como a incorporação de variáveis climáticas, espaciais e técnicas de interpretabilidade, os quais superam as limitações das abordagens locais mais tradicionais.

4.2 ESTUDO DE CASO 2: PREVISÃO DE DEMANDA NO VAREJO

Os achados de Carbonneau et al. (2008) demonstraram que, embora técnicas avançadas como RNN e SVM tenham apresentado desempenho superior em termos de erro médio absoluto (MAE), a regressão linear múltipla (MLR) mostrou-se estatisticamente comparável, destacando-se pela simplicidade e viabilidade computacional. Essa constatação permanece relevante, especialmente quando comparada a estudos mais recentes que adotam métodos mais robustos de machine learning.

Viana et al. (2024), por exemplo, aplicaram algoritmos como *XGBoost*, *LightGBM* e *Random Forest* no contexto de marketplaces, e os compararam com métodos tradicionais como SARIMA e Holt-Winters. Os resultados evidenciaram que os modelos de *boosting* (em especial o *XGBoost*) superaram os métodos clássicos nas métricas MAE, RMSE e MAPE, principalmente ao lidar com cenários complexos e com múltiplas variáveis. No entanto, os autores também destacaram o maior custo computacional e a necessidade de ajustes finos nos hiperparâmetros, fatores que reabrem a discussão sobre o equilíbrio entre desempenho e aplicabilidade prática.

De maneira semelhante, a pesquisa de Bessa (2021), focada no setor de varejo de moda, confirmou que o *XGBoost* com tratamento sazonal e seleção de variáveis obteve os melhores resultados em acurácia. No entanto, a análise AHP (*Analytic Hierarchy Process*), que considerou não apenas a precisão, mas também critérios como complexidade, custo e tempo de implementação, indicou o modelo *Naïve* ajustado como mais adequado para o contexto analisado. Este achado reforça a ideia de que a melhor técnica preditiva nem sempre é a mais complexa, alinhando-se à conclusão de Carbonneau et al. (2008), que ressaltaram a competitividade da regressão linear em determinadas condições.

Portanto, ao comparar os três estudos, é possível observar uma tensão constante entre sofisticação algorítmica e aplicabilidade operacional. Enquanto os métodos de aprendizado de máquina oferecem ganhos em acurácia, especialmente em ambientes com grande volume de dados e alta variabilidade, métodos estatísticos simples ainda se mostram eficientes em cenários com menos recursos computacionais ou integração limitada de dados. A escolha do modelo, portanto, deve

considerar múltiplos critérios, indo além da métrica preditiva, para atender às demandas reais do negócio e às restrições técnicas envolvidas.

4.3 ESTUDO DE CASO 3: ANÁLISE DE TRÂNSITO PARA PREVENÇÃO DE ÁREAS DE RISCO

A análise espaço-temporal de acidentes de trânsito, como proposta por Mohammed et al. (2023), evidencia o potencial das ferramentas geoespaciais para prever áreas críticas e subsidiar estratégias de prevenção de acidentes rodoviários. No contexto brasileiro, essas análises se tornam ainda mais relevantes diante da complexidade urbana, desigualdade social e infraestrutura viária deficitária. A utilização de sistemas de informações geográficas (SIG) e técnicas de sensoriamento remoto permite não apenas mapear áreas de alto risco, mas também compreender como a configuração urbana influencia a ocorrência de acidentes.

Um exemplo notável dessa abordagem pode ser observado no estudo de Batista Júnior (2016), que analisou a relação entre expansão urbana e segregação socioespacial em Salvador/BA. A pesquisa demonstrou que a ocupação desordenada do espaço urbano produz ambientes com diferentes níveis de infraestrutura e acessibilidade, resultando em desigualdade no acesso à mobilidade segura. No caso da comunidade de Canabrava, observou-se um padrão de exclusão territorial em que a precariedade do sistema viário e a ausência de serviços urbanos contribuem para a vulnerabilidade da população aos riscos ambientais e de tráfego. A superposição entre áreas de exclusão social e indicadores de urbanização revelou um padrão espacial de risco que pode ser comparado a outras metrópoles brasileiras com dinâmicas semelhantes.

Nesse sentido, o uso de técnicas de machine learning para a classificação de risco de acidentes, como explorado por Amorim et al. (2019), amplia a capacidade preditiva das análises, especialmente ao lidar com grandes volumes de dados. Ao utilizar diferentes combinações de algoritmos, como MLP, Random Forest e BernoulliNB, os referidos autores demonstraram que é possível obter alta acurácia (acima de 85%) na identificação de segmentos rodoviários propensos a acidentes graves. Assim, tal abordagem é particularmente eficaz quando combinada a bases de dados balanceadas e múltiplas variáveis, como horário do acidente, tipo de via,

orientação da estrada e condições climáticas, permitindo decisões mais precisas e direcionadas.

Complementando essa perspectiva, Ferreira et al. (2024) aplicaram técnicas de inteligência artificial para prever acidentes na cidade de Campo Grande (MS), utilizando dados históricos de tráfego. A proposta evidenciou a utilidade de modelos preditivos no apoio à tomada de decisão do poder público, possibilitando a implementação de políticas pontuais em horários e regiões com maior concentração de incidentes. A inclusão de variáveis como horários de pico, meteorologia e tipo de acidente potencializou os modelos, indicando que o uso de IA pode não apenas prever ocorrências, mas também servir de base para campanhas educativas e fiscalização direcionada.

Além disso, a inovação no uso de dados colaborativos destaca-se no trabalho de Arce Junior et al. (2024), que investigaram a utilização de dados do aplicativo Waze na Região Metropolitana de Manaus. A coleta de informações em tempo real via *crowdsourcing* permitiu identificar bairros com maior incidência de acidentes, revelando uma nova possibilidade para análise de risco baseada em dados gerados por usuários. A aplicação de técnicas de análise de tráfego e telemetria possibilitou o mapeamento de padrões comportamentais dos motoristas, como excesso de velocidade e rotas de risco, permitindo uma resposta mais rápida e eficaz por parte das autoridades locais.

Dessa forma, os estudos analisados convergem para uma compreensão ampliada dos acidentes de trânsito enquanto fenômenos multidimensionais, que envolvem fatores urbanos, sociais e comportamentais. A integração entre análise geoespacial, algoritmos de aprendizado de máquina e dados colaborativos oferece um panorama promissor para a construção de políticas públicas mais eficazes, orientadas não apenas pela estatística histórica, mas também por sistemas inteligentes capazes de antecipar e mitigar riscos. No contexto brasileiro, marcado por desigualdade territorial e expansão urbana desigual, essas tecnologias têm o potencial de promover justiça espacial ao direcionar investimentos e ações de prevenção para as áreas mais vulneráveis.

5 CONSIDERAÇÕES FINAIS

O presente estudo permitiu compreender a importância da análise de séries temporais como ferramenta estratégica na interpretação e previsão de fenômenos que evoluem ao longo do tempo. Ao conciliar os conceitos teóricos com aplicações práticas, foi possível constatar que as séries temporais não apenas revelam padrões históricos, mas também oferecem subsídios valiosos para a tomada de decisões em contextos diversos e complexos.

A análise dos três estudos de caso, previsão de surtos de dengue com LSTM e SHAP, previsão de demanda no varejo com modelos de machine learning e análise espaço-temporal de acidentes de trânsito com SIG e algoritmos preditivos, evidenciou que a escolha do modelo e da abordagem metodológica deve ser orientada pelas características específicas dos dados e pelos objetivos da previsão. Enquanto o uso de redes neurais recorrentes (LSTM) mostrou-se eficaz para lidar com variáveis defasadas e dependências espaciais em dados epidemiológicos, os modelos baseados em árvores (como XGBoost e LightGBM) se destacaram pela alta performance e interpretabilidade em ambientes altamente voláteis, como o varejo. Já no campo da mobilidade urbana, a integração entre dados espaciais e algoritmos de aprendizado de máquina demonstrou grande potencial para a identificação de áreas críticas e prevenção de acidentes.

Os estudos analisados também mostraram que, embora técnicas de aprendizado profundo e machine learning apresentem alto desempenho preditivo, sua efetividade está condicionada à qualidade dos dados, ao tratamento de variáveis sazonais, e à correta escolha de hiperparâmetros. Além disso, foi evidenciado que modelos mais simples, como o Naïve ajustado, ainda podem apresentar boa performance em determinados contextos, reforçando a necessidade de avaliação criteriosa e contextualizada.

Portanto, a presente pesquisa reforça a relevância das séries temporais não apenas como técnica estatística, mas como ferramenta multidisciplinar aplicada a problemas reais, promovendo eficiência operacional, redução de riscos e suporte à formulação de políticas públicas. O avanço contínuo na coleta de dados, aliado ao desenvolvimento de modelos híbridos e mais interpretáveis, tende a ampliar ainda mais o potencial dessas análises nos próximos anos.

Ainda que os resultados evidenciem o potencial das séries temporais, é importante reconhecer algumas limitações deste trabalho, já que a análise foi conduzida a partir de um número restrito de artigos, o que naturalmente limita a diversidade de aplicações consideradas. Além disso, por se tratar de um estudo baseado em pesquisas já publicadas, não foi possível explorar de forma prática aspectos como a coleta e o tratamento de dados ou os desafios computacionais envolvidos em implementações reais. Essas restrições, contudo, não comprometem as contribuições alcançadas, mas apontam oportunidades de aprofundamento em futuras investigações.

Para estudos futuros, recomenda-se o aprofundamento em abordagens híbridas, que integrem técnicas estatísticas tradicionais com métodos de aprendizado profundo, de forma a unir interpretabilidade e robustez preditiva. Além disso, a ampliação do escopo de análise para outros setores, como energia, finanças ou mudanças climáticas, poderá enriquecer ainda mais a compreensão sobre o potencial transformador da modelagem temporal. A análise da influência de fatores externos (como políticas públicas, eventos extremos ou flutuações econômicas) sobre séries temporais também se mostra uma linha promissora de investigação.

REFERÊNCIAS

AMORIM, Brunna de Sousa Pereira et al. **Uso de aprendizado de máquina para classificação de risco de acidentes em rodovias**. 2019.

ANTUNES, José Leopoldo Ferreira; CARDOSO, Maria Regina Alves. Uso da análise de séries temporais em estudos epidemiológicos. **Epidemiologia e Serviços de Saúde**, v. 24, p. 565-576, 2015.

ARCE JUNIOR, Francisco Lourenço Duarte et al. **Análise de dados do Waze para auxílio na diminuição de acidentes de trânsito na Região Metropolitana de Manaus**. 2024. Disponível em: <https://repositorio.ufsc.br/handle/123456789/259670>. Acesso em: 21 jun. 2025.

BATISTA JUNIOR, Eudoxio Antonio. **Caracterização da relação entre a expansão urbana e a segregação socioespacial com uso de sistema de informações geográficas e sensoriamento remoto**. 2016. Dissertação de mestrado. Disponível em: <https://repositorio.ufba.br/handle/ri/19531>. Acesso em: 20 jun. 2025.

BESSA, Adriana Bezerra. **Ensaio sobre previsão de vendas no varejo de moda**. 2021. Tese de Doutorado. Disponível em: <https://repositorio.fgv.br/server/api/core/bitstreams/6c19f0b7-3224-41d0-aefa-df0f6ddc6f96/content>. Acesso em: 18 jun. 2025.

BHATT, Samir et al. Distribuição global e carga da dengue. **Nature**, v. 496, n. 7446, p. 504-507, 2013.

BOX, George EP et al. **Análise de séries temporais: previsão e controle**. John Wiley & Sons, 2015.

BROCKWELL, Peter J.; DAVIS, Richard A. (Ed.). **Introduction to time series and forecasting**. New York, NY: Springer New York, 2002.

CARBONNEAU, Real; LAFRAMBOISE, Kevin; VAHIDOV, Rustam. Application of machine learning techniques for supply chain demand forecasting. **European Journal of Operational Research**, v. 184, n. 3, p. 1140–1154, 2008. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0377221706009243>. Acesso em: 29 jun. 2025.

CESARO, Juliana. **Avaliação de Discriminação em Aprendizagem de Máquina usando Técnicas de Interpretabilidade**. 2021. Tese de Doutorado. Universidade de São Paulo. Disponível em: https://web.archive.org/web/20210608061100id_/https://teses.usp.br/teses/disponiveis/3/3141/tde-31052021-114333/publico/JulianaCesaroCorr21.pdf. Acesso em: 18 jun. 2025.

CODECO, Claudia; COELHO, Flávio; CRUZ, Oswaldo; OLIVEIRA, Sérgio; CASTRO, Thiago; BASTOS, Leonardo. Infodengue: A nowcasting system for the surveillance of

arboviruses in Brazil. **Revue d'Épidémiologie et de Santé Publique**, v. 66, supl., p. S386, 2018.

COPERNICUS CLIMATE CHANGE SERVICE (C3S). ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate, 2017. Disponível em: <https://doi.org/10.24381/cds.adbb2d47>. Acesso em: 19 mai. 2025.

CHEN, Xiang; MORAGA, Paula. Forecasting dengue across Brazil with LSTM neural networks and SHAP-driven lagged climate and spatial effects. **BMC Public Health**, v. 25, 2025. Disponível em: <https://doi.org/10.1186/s12889-025-22106-7>. Acesso em: 29 jun. 2025.

DE SOUZA, Anderson Paulino et al. **Enhancing machine learning models to evaluate optical water quality parameters of reservoirs through remote sensing data**. 2024.

FERREIRA, Thays Martines et al. Utilização da inteligência artificial na predição de acidentes de trânsito na cidade de Campo Grande—MS. **Revista Camalotes**, v. 3, n. 2, 2024.

FIGUEIREDO, Erik Alencar de; MARQUES, André M. Inflação inercial como um processo de longa memória: análise a partir de um modelo Arfima-Figarch. **Estudos Econômicos (São Paulo)**, v. 39, p. 437-458, 2009.

GONÇALVES, Esmeralda; MENDES-LOPES, Nazaré. **Probabilidades-Princípios Teóricos**. Escolar Editora, 2013.

GRANGER, C. W. J.; JOYEUX, R. An introduction to long-memory time series models and fractional differencing. **Journal of Time Series Analysis**, 1980.
GRANGER, C.; ANDERSEN, A. On the invertibility of time series models. **Stochastic Processes and their Applications**, v. 8, p. 87–92, 1978.

HEWAMALAGE, Hansika; BERGMEIR, Christoph; BANDARA, Kasun. Redes neurais recorrentes para previsão de séries temporais: Situação atual e direções futuras. **International Journal of Forecasting**, v. 37, n. 1, p. 388-427, 2021.

HOCHREITER, Sepp; SCHMIDHUBER, Jürgen. Memória de longo prazo. **Computação neural**, v. 9, n. 8, pág. 1735-1780, 1997.

HOSKING, J. R. M. Fractional differencing. **Biometrika**, v. 68, n. 1, p. 165–176, 1981.

HYNDMAN, R. J.; KOEHLER, A. B. Another look at measures of forecast accuracy. **International Journal of Forecasting**, v. 22, n. 4, p. 679–688, 2006.

HYNDMAN, R. J.; ATHANASOPOULOS, G. **Forecasting: Principles and Practice**. 2. ed. OTexts, 2018.

ENGLE, R. F. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. **Econometrica**, v. 50, n. 4, p. 987–1007, 1982.

JOHANSEN, Igor Cavallini; CARMO, Roberto Luiz do; ALVES, Luciana Correia. *O espaço como dimensão analítica central na compreensão da dinâmica da Dengue na escala intramunicipal.*: 2015. Disponível em: https://d1wqtxts1xzle7.cloudfront.net/98005034/2638-libre.pdf?1675096200=&response-content-disposition=inline%3B+filename%3DO_espaco_como_dimensao_analitica_central.pdf&Expires=1748624619&Signature=CiYlvRQ1vbRqeag90WKWgQD~A5-NCwRHbrxnO5GDdma4f4nut8P8Ko1bDwR9g3k4O444YbG41~Iz3IRn7R2bOjlzPTeZutlnZP1c55hbKUnRv-MbA11x1g7Sf-EIGNSp4LEGaYNOjVbyXz5RVwyd0MY5HbNRgb1XgslFhcp2OYv~WlhFws6CF010rMKbNCsNvllpl2E3PR1PfJHYqFDURK9rf89eehJ2vGV-hJ57v70XrxpuoEZDPI7GITHV6nCQNLiujSAMBjKH2jMWBvwJCGSCc7~dxBQRW1~A~UQCx35QO2P0nblDgnS0p0yl8xiHJYhAIDvD3~mRRb4d19Xpw__&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA. Acesso em: 29 mai. 2025.

LOUKAITOU-SIDERIS, Anastasia; LIGGETT, Robin; SUNG, Hyun-Gun. Death on the crosswalk: A study of pedestrian-automobile collisions in Los Angeles. **Journal of Planning Education and Research**, v. 26, n. 3, p. 338-351, 2007.

LUNDBERG, Scott M.; LEE, Su-In. Atribuição consistente de características para conjuntos de árvores. **arXiv pré-impressão arXiv:1706.06060**

LUNDBERG, Scott M.; LEE, Su-In. Uma abordagem unificada para interpretar previsões de modelos. **Avanços em sistemas de processamento de informações neurais**, v. 30, 2017.

MOHAMMED, Semira; ALKHEREIBI, Aya Hasan; ABULIBDEH, Ammar; JAWARNEH, Rana N.; BALAKRISHNAN, Perumal. GIS-based spatiotemporal analysis for road traffic crashes; in support of sustainable transportation planning. **Transportation Research Interdisciplinary Perspectives**, v. 20, 2023. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2590198223000891>. Acesso em: 29 jun. 2025.

MONTGOMERY, D. C.; PECK, E. A.; VINING, G. G. **Introduction to Linear Regression Analysis**. 5. ed. Wiley, 2012.

MONTGOMERY, D. C.; JENNINGS, C. L.; KULAHCI, M. **Introduction to Time Series Analysis and Forecasting**. 2. ed. Wiley, 2015.

PASCOAL, Rui. **Apontamentos de Análise de Séries Financeiras**. Faculdade de Economia, Universidade de Coimbra, 2019/2020.

BOX, George E. P. et al. Análise de séries temporais: previsão e controle. **John Wiley & Sons**, 2015.

SILVA, Natália Regina Souza da. *Padrões temporais e demográficos da incidência de dengue no município de Foz do Iguaçu-PR: um estudo descritivo*. 2024. Trabalho de Conclusão de Curso (Graduação em Medicina) – Universidade Federal da Integração Latino-Americana (UNILA), Foz do Iguaçu, 2024. Disponível em:

<https://dspace.unila.edu.br/items/b0716f6d-4884-4f07-95e7-79c7cfaa2003>. Acesso em: 30 mai. 2025.

SILVA, Wilson Rogério Soares et al. Metodologia de monitoramento de epidemias: uma abordagem baseada em redes neurais artificiais. 2018. Disponível em: <https://repositorio.ufpa.br/handle/2011/10056>. Acesso em: 18 jun. 2025.

VIANA, Leonardo de Almeida Lino et al. Aplicação de machine learning na previsão de demandas: otimização de recursos no setor de e-commerce. **Revista Contemporânea**, v. 4, n. 11, p. e6545-e6545, 2024.

WORLD HEALTH ORGANIZATION. **Dengue and Severe Dengue**. 2023. <https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue>. Acesso em: 19 mai. 2025.