

UNIVERSIDADE FEDERAL DE SÃO CARLOS– UFSCAR
CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA– CCET
DEPARTAMENTO DE COMPUTAÇÃO– DC
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO– PPGCC

Estêvão José dos Santos

**Da Estatística ao Aprendizado
Profundo: Previsão da Taxa de
Transferência Móvel**

Sorocaba
2025

Estêvão José dos Santos

**Da Estatística ao Aprendizado
Profundo: Previsão da Taxa de
Transferência Móvel**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação do Centro de Ciências Exatas e de Tecnologia da Universidade Federal de São Carlos, como parte dos requisitos para a obtenção do título de Mestre em Ciência da Computação.

Área de concentração: Metodologias e Técnicas de Computação

Orientador: Tiago Agostinho de Almeida

Sorocaba

2025



UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Ciências Exatas e de Tecnologia
Programa de Pós-Graduação em Ciência da Computação

Folha de Aprovação

Defesa de Dissertação de Mestrado do candidato Estêvão José dos Santos., realizada em 07/10/2025.

Comissão Julgadora:

Prof. Dr. Tiago Agostinho de Almeida (UFSCar)

Prof. Dr. Fabio Luciano Verdi (UFSCar)

Prof. Dr. André Luis Debiaso Rossi (UNESP)

*Dedico este trabalho a todos e todas que sonham,
aprendem e ensinam, acreditando que a educação
é a prática da liberdade e a semente
de um mundo mais justo.*

Agradecimentos

Agradeço,

À minha filha, Luiza, que cresceu junto com este trabalho e, com sua curiosidade e alegria, me ensinou diariamente sobre amor e resiliência. Mesmo nos dias mais difíceis, sua presença foi motivo de energia e inspiração.

À minha esposa, Bianca, pelo amor e companheirismo. Suas palavras francas e conselhos firmes me fizeram refletir e corrigir rotas, permitindo-me avançar mesmo quando os obstáculos pareciam intransponíveis.

Ao meu orientador, Prof. Dr. Tiago Almeida, pela orientação precisa, paciência e pragmatismo. Sua atenção e confiança foram fundamentais para a realização deste trabalho e para a publicação do meu primeiro artigo.

Ao Alessandro, meu gestor no FIT, pela compreensão e flexibilidade que me permitiram equilibrar as demandas profissionais e acadêmicas.

À minha mãe, Nylce, por todo o carinho, paciência e apoio constante, que me deram ânimo para seguir adiante mesmo nos momentos mais desafiadores, além de me apresentar ao universo das universidades federais.

Ao meu pai e mestre, Claudio, por ter trilhado esse caminho no passado e me mostrar que era possível, ensinando-me o valor da dedicação. Em conjunto com minha madrasta, Lucia, que me ajudou a discernir entre o medo de enfrentar desafios e a real incapacidade de alcançá-los.

Aos meus irmãos, Rodrigo, Nathalia, Camila e Nicole, e à minha irmã, Dra. Claudia, pelo incentivo e pelo exemplo acadêmico e profissional. O esforço e a determinação de cada um sempre me inspiraram a acreditar nos meus próprios caminhos.

A todos os familiares, amigos e colegas que, de alguma forma, contribuíram direta ou indiretamente para a realização desta dissertação, registro aqui meu sincero reconhecimento.

Por fim, agradeço profundamente o apoio fornecido pelas agências brasileiras FAPESP (SMARTNESS, processo nº 2021/00199-8) e CNPq (processo nº 311867/2023-5).

*“A ciência tem que ser um instrumento de transformação,
e não apenas de contemplação.”
(Prof. Dr. Milton Santos)*

Resumo

As limitações do *throughput* (taxa de transferência de dados) continuam sendo um dos principais desafios para o desempenho das redes móveis. Elas decorrem de fatores como restrições de capacidade, coexistência entre diferentes gerações tecnológicas, mobilidade dos usuários, integração massiva da Internet das Coisas e desigualdades de acesso em regiões como a América Latina. Essa variabilidade compromete diretamente a Qualidade de Serviço (QoS) e Qualidade da Experiência (QoE), evidenciando a necessidade de mecanismos preditivos capazes de antecipar flutuações de desempenho, otimizar a alocação de recursos de rádio e assegurar conectividade eficiente em cenários heterogêneos e complexos. Neste contexto, a previsão do *throughput* assume papel central para o gerenciamento adaptativo de recursos e para a manutenção da qualidade de serviço em redes móveis, especialmente sob condições de alta mobilidade. Este trabalho investiga, de forma sistemática, duas dimensões fundamentais da modelagem de séries temporais aplicadas a esse problema: (i) o contraste entre modelos locais e globais e (ii) o impacto da inclusão ou exclusão de covariáveis externas. Para tanto, são avaliados métodos estatísticos, de aprendizado de máquina e de aprendizado profundo com dados reais de redes móveis, em que o *throughput* é previsto a partir de métricas de qualidade de canal e da velocidade do usuário como potenciais covariáveis. Os resultados experimentais demonstram que modelos globais baseados em árvores, como o *LightGBM*, oferecem o melhor equilíbrio entre acurácia, robustez e eficiência. Além disso, o emprego de covariáveis exploradas, principalmente relacionadas à qualidade da rede, mostraram-se insuficientes para melhorar de forma consistente o desempenho preditivo diante da complexidade inerente à tarefa de previsão.

Palavras-chave: Previsão de séries temporais, Aprendizado de máquina, Redes móveis, QoS, *Throughput*.

Abstract

The limitations of *throughput* remain one of the main challenges for mobile network performance. They stem from factors such as capacity constraints, coexistence of multiple technological generations, user mobility, massive integration of the Internet of Things, and unequal access in regions such as Latin America. This variability directly compromises Quality of Service (QoS) and Quality of Experience (QoE), highlighting the need for predictive mechanisms capable of anticipating performance fluctuations, optimizing radio resource allocation, and ensuring efficient connectivity in heterogeneous and complex scenarios. In this context, accurate *throughput* forecasting plays a central role in adaptive resource management and in maintaining service quality in mobile networks, especially under high-mobility conditions. This work systematically investigates two fundamental dimensions of time series modeling applied to this problem: (i) the contrast between local and global models, and (ii) the impact of including or excluding external covariates. To this end, statistical, machine learning, and deep learning methods are evaluated using real-world mobile network data, where *throughput* is predicted from channel quality metrics and user speed as potential covariates. Experimental results show that global tree-based models, such as *LightGBM*, provide the best balance between accuracy, robustness, and efficiency. Furthermore, the use of covariates—mainly related to network quality—proved insufficient to consistently improve predictive performance given the inherent complexity of the forecasting task.

Keywords: Time series forecasting, Machine learning, Mobile networks, QoS, Throughput.

Lista de ilustrações

Figura 1 – Exemplos de evolução temporal das métricas de qualidade de canal (RSRP, RSRQ, RSSI, SNR e CQI) em diferentes cenários experimentais: <i>streaming</i> e <i>download</i> , com a Equipamento do Usuário (<i>User Equipment</i>) (UE) em condição estática ou em deslocamento. O eixo temporal está expresso em segundos, e as linhas tracejadas indicam os valores médios de cada métrica.	37
Figura 2 – Exemplo de cenário de previsão com região de treinamento, origem e horizonte de previsão.	39
Figura 3 – Exemplo de decomposição de uma série temporal em seus componentes: série original, tendência, sazonalidade e resíduos. As barras cinzas à esquerda indicam a escala relativa de cada componente em comparação à série original.	41
Figura 4 – Comparação entre as metodologias de avaliação por origem fixa (à esquerda) e por origem rolante (à direita), considerando horizonte de previsão igual a $H = 1$	44
Figura 5 – Exemplo de regressão linear simples.	48
Figura 6 – Diferença geométrica entre as regularizações ℓ_1 (Lasso) e ℓ_2 (Ridge). O centro representa a solução por Mínimos Quadrados Ordinários (MQO) (do inglês, $\hat{\beta}_{OLS}$), sem penalização.	49
Figura 7 – Estrutura de uma árvore de decisão.	51
Figura 8 – Arquitetura de um <i>Perceptron</i> de Múltiplas Camadas (MLP) com quatro entradas e uma camada oculta contendo três neurônios.	56
Figura 9 – Arquitetura de uma rede neural recorrente simples, com destaque para a propagação das entradas \mathbf{x}_ℓ , dos estados ocultos \mathbf{h}_ℓ e das saídas \mathbf{o}_ℓ	56
Figura 10 – Diagrama de uma célula LSTM, destacando os portões de esquecimento, entrada e saída, o estado da célula e as operações de produto de Hadamard (ponto a ponto).	58

Figura 11 – Arquitetura N-BEATS	60
Figura 12 – Bloco básico genérico.	61
Figura 13 – Rota do ônibus de transporte público utilizada para a coleta do conjunto de dados da rede 4G LTE em Kingston, Ontário.	64
Figura 14 – Correlação de Spearman entre características atributos da base dados celular 5G.	75
Figura 15 – Resultados de regressão com diferentes modelos de ML em redes móveis: (a) MAE, (b) RMSE, (c) Correlação e (d) R^2	79
Figura 16 – Estrutura do modelo híbrido LSTM-GRU proposto.	81
Figura 17 – Análise de erro do modelo híbrido na previsão de <i>throughput</i>	82
Figura 18 – Localização geográfica de todas as medições do conjunto de dados de Raca et al. (2020), com distinção entre pontos coletados em cenários estáticos e em deslocamento.	88
Figura 19 – Correlação de Pearson calculada para todo o conjunto de dados (condições estáticas e em condução), abrangendo métricas de canal.	90
Figura 20 – Matriz de correlação de Pearson calculada apenas para as medições em condição estática.	91
Figura 21 – Matriz de correlação de Pearson calculada apenas para as medições em condição de deslocamento.	92
Figura 22 – Proporção de séries temporais estacionárias e não estacionárias, considerando os resultados do teste de ADF aplicado ao <i>throughput</i>	93
Figura 23 – Proporção de valores ausentes por métrica do conjunto de dados.	94
Figura 24 – Protocolo experimental.	95
Figura 25 – Validação <i>walk-forward</i> : divisão treino/validação, avanço das janelas e diferença de fluxo entre modelos locais (re-treinamento) e globais (treino único).	96
Figura 26 – Análise de desempenho dos modelos de previsão, comparando o RMSE médio com o tempo de treinamento.	100
Figura 27 – Distribuição do erro de previsão dos modelos, avaliada pela métrica Raiz do Erro Quadrático Médio (<i>Root Mean Squared Error</i>) (RMSE). O boxplot apresenta os resultados consolidados para todos os cenários analisados, englobando séries estáticas e em mobilidade, permitindo a comparação global do desempenho entre diferentes tipos de modelos.	101
Figura 28 – Previsão da série 75 no cenário estático com transmissão de vídeo, utilizando o modelo <i>LightGBM</i> com covariáveis.	102
Figura 29 – Previsão da série 75 no cenário estático com transmissão de vídeo, utilizando o modelo <i>LightGBM</i> sem covariáveis.	102
Figura 30 – Previsão da série 75 no cenário estático com transmissão de vídeo, utilizando o modelo <i>Naive Mean</i> local.	103

Figura 31 – Previsão da série 50 no cenário de condução com transmissão de vídeo, utilizando o modelo <i>LightGBM</i> com covariáveis.	103
Figura 32 – Previsão da série 50 no cenário de condução com transmissão de vídeo, utilizando o modelo <i>LightGBM</i> sem covariáveis.	104
Figura 33 – Previsão da série 50 no cenário de condução com transmissão de vídeo, utilizando o modelo <i>Naive Mean</i> local.	104
Figura 34 – Comparação percentual do desempenho de todos os modelos em relação ao <i>Naive Mean</i> , utilizando como referência a métrica RMSE. Valores negativos indicam que o modelo obteve erro médio inferior (melhor desempenho) em comparação ao <i>baseline</i> , enquanto valores positivos indicam desempenho pior.	105
Figura 35 – Comparação percentual do desempenho dos modelos em relação ao <i>Naive Mean</i> , considerando exclusivamente o cenário de mobilidade em condução. A métrica de referência é o RMSE, com valores negativos representando melhora em relação ao <i>baseline</i> e positivos indicando pior desempenho.	107
Figura 36 – Comparação percentual do desempenho dos modelos em relação ao <i>Naive Mean</i> , no cenário de mobilidade estática. A referência é a métrica RMSE; valores negativos indicam modelos mais precisos que o <i>baseline</i> , enquanto valores positivos indicam maior erro em relação à referência.	108
Figura 37 – Comparação percentual do desempenho dos modelos em relação ao <i>Naive Mean</i> , considerando exclusivamente o cenário de transmissão de vídeo. A métrica utilizada é o RMSE, onde valores negativos indicam melhora em relação ao <i>baseline</i> e positivos representam pior desempenho.	109
Figura 38 – Comparação percentual do desempenho dos modelos em relação ao <i>Naive Mean</i> , considerando exclusivamente o cenário de download de arquivos. A métrica utilizada é o RMSE, com valores negativos representando ganho em relação ao <i>baseline</i> e positivos indicando perda de desempenho.	110

Lista de tabelas

Tabela 1 – Casos Especiais do Modelo ARIMA	47
Tabela 2 – Resultados obtidos com os modelos de regressão e séries temporais . .	66
Tabela 3 – Campos brutos ou obtidos a partir das APIs do Android no conjunto de dados Lumos5G. Com exceção do campo <i>timestamp</i> , os demais possuem acurácia estimada pelo Android e podem não ser totalmente precisos.	67
Tabela 4 – Campos derivados ou obtidos após processamento no conjunto de dados Lumos5G.	68
Tabela 5 – Grupos de atributos utilizado por Lumos5G.	70
Tabela 6 – Parâmetros de treinamento dos modelos avaliados no Lumos5G.	71
Tabela 7 – Resultados de regressão (MAE) para diferentes modelos e grupos de variáveis	73
Tabela 8 – Resultados de regressão (RMSE) para diferentes modelos e grupos de variáveis	73
Tabela 9 – Melhores resultados obtidos com os modelos LSTM, BiLSTM, ANN e <i>Random Forest</i>	76
Tabela 10 – Hiperparâmetros definidos para os modelos comparados pelo modelo ESDT.	78
Tabela 11 – Hiperparâmetros definidos para o modelo híbrido LSTM-GRU.	81
Tabela 12 – Síntese dos trabalhos relacionados sobre previsão de <i>throughput</i> em redes móveis.	83
Tabela 13 – Estatísticas descritivas do <i>throughput</i> (em Mbps) por mobilidade e atividade do usuário. Nota: D.P. = desvio-padrão; <i>N</i> = número de amostras.	89
Tabela 14 – Modelos de previsão indicando suporte a covariadas passadas e escopo (local ou global).	95
Tabela 15 – Parâmetros configurados para os modelos de previsão.	97

Tabela 16 – Resumo estatístico dos erros de previsão (RMSE e MAE) para múltiplos modelos considerando todos os cenários incorporados.	122
Tabela 17 – Resumo estatístico dos erros de previsão (RMSE e MAE) para múltiplos modelos no cenário estático.	123
Tabela 18 – Resumo estatístico dos erros de previsão (RMSE e MAE) para múltiplos modelos no cenário de condução.	124
Tabela 19 – Resumo estatístico dos erros de previsão (RMSE e MAE) para múltiplos modelos no cenário de transmissão de vídeo.	125
Tabela 20 – Resumo estatístico dos erros de previsão (RMSE e MAE) para múltiplos modelos no cenário de download de arquivo.	126
Tabela 21 – Resumo estatístico dos erros de previsão (RMSE e MAE) para múltiplos modelos no cenário de transmissão de vídeo em condição estática.	127
Tabela 22 – Resumo estatístico dos erros de previsão (RMSE e MAE) para múltiplos modelos no cenário de download de arquivo em condição estática.	128
Tabela 23 – Resumo estatístico dos erros de previsão (RMSE e MAE) para múltiplos modelos no cenário de transmissão de vídeo em condição de condução.	129
Tabela 24 – Resumo estatístico dos erros de previsão (RMSE e MAE) para múltiplos modelos no cenário de download de arquivo em condição de condução.	130

Lista de siglas

4G 4º geração de redes móveis

5G 5º geração de redes móveis

ACF Funções de Autocorrelação (*Autocorrelation Function*)

CQI Indicador de Qualidade do Canal (*Channel Quality Indicator*)

HO *Handover*

IoT Internet das Coisas (*Internet of Things*)

LTE 4G *Long-Term Evolution*

MAE Erro Médio Absoluto (*Mean Absolute Error*)

mmWave ondas milimétricas (*millimeter wave*)

MQO Mínimos Quadrados Ordinários

MSE Erro Médio Quadrático (*Mean Squared Error*)

ML Aprendizado de máquina (*Machine Learning*)

PACF Autocorrelação Parcial (*Partial Autocorrelation Function*)

QoE Qualidade da Experiência (*Quality of Experience*)

QoS Qualidade de Serviço (*Quality of Service*)

RMSE Raiz do Erro Quadrático Médio (*Root Mean Squared Error*)

RSRP Potência de Sinal de Referência Recebida (*Reference Signal Received Power*)

RSRQ Qualidade do Sinal de Referência Recebido (*Reference Signal Received Quality*)

RSSI Indicador de Intensidade do Sinal Recebido (*Received Signal Strength Indicator*)

SQR Soma dos Quadrados dos Resíduos

SNR Relação Sinal-Ruído (*Signal-to-Noise Ratio*)

STL Decomposição Sazonal e de Tendência por Loess (*Seasonal-Trend decomposition using Loess*)

UE Equipamento do Usuário (*User Equipment*)

Lista de símbolos

t	índice temporal (passo de tempo)
T	origem/último instante observado
N	tamanho da amostra
H	horizonte de previsão (número de passos à frente)
h	passo à frente (usado no método com <i>drift</i>)
y_t	valor observado da série no instante t
\hat{y}_t	valor ajustado/predito no instante t
$\{y_t\}_{t=-\infty}^{\infty}$	sequência temporal bilateral
$f(\cdot)$	função/mapeamento do modelo de previsão
$X_t^{(i)}$	i -ésima covariável (preditor) no instante t
\mathbf{x}_t	vetor de covariáveis no instante t
$\hat{\mathbf{y}}$	vetor de previsões futuras
\mathcal{L}	função de perda (treinamento)
θ	parâmetros do modelo (pesos em geral)
α	taxa de aprendizado (otimização)
e_t	resíduo (usa-se e_t no tempo t , em regressão, e_i); $e = y - \hat{y}$
\bar{y}	média amostral
Y_t	série original (notação da decomposição)

T_t	tendência
S_t	sazonalidade
C_t	componente cíclica
$E[Y_t]$	valor esperado (média) da série
$\text{Var}(Y_t)$	variância da série
$\gamma(k)$	autocovariância na defasagem k
$\text{Cov}(Y_t, Y_{t-k})$	covariância na defasagem k
B	operador de defasagem ($By_t = y_{t-1}$)
$\phi_p(B)$	polinômio autorregressivo (ordem p)
$\theta_q(B)$	polinômio de médias móveis (ordem q)
d	ordem de diferenciação
p, q	ordens AR e MA do ARIMA
c	constante (termo do ARIMA)
ϵ_t	ruído branco
\hat{y}	valor predito (regressão linear)
β_0	intercepto
β_j	coeficiente associado à variável X_j
$\boldsymbol{\beta}$	vetor de coeficientes
m	número de preditores (características)
X	matriz de projeto (design matrix), dimensões $N \times m$
X_j	j -ésimo preditor (coluna de X)
\mathbf{x}_i	vetor de atributos da observação i
\mathcal{D}	conjunto de dados de treinamento $\{(\mathbf{x}_i, y_i)\}_{i=1}^N$
ℓ_1	norma ℓ_1 (penalidade Lasso), $\ \boldsymbol{\beta}\ _1 = \sum_j \beta_j $
ℓ_2	norma ℓ_2 (penalidade Ridge), $\ \boldsymbol{\beta}\ _2^2 = \sum_j \beta_j^2$
P_m	m -ésima partição do espaço de atributos

$\mathbb{I}(\mathbf{x} \in P_m)$	função indicadora da região P_m
c_m	valor de predição constante na região P_m
J	número de regiões/folhas de uma árvore
$T_k(\mathbf{x})$	predição da k -ésima árvore no ensemble
$F(\mathbf{x})$	preditor aditivo em boosting ($F(\mathbf{x}) = \sum_{k=1}^M T_k(\mathbf{x})$)
η	taxa de aprendizado (<i>shrinkage</i>) em boosting
ρ_k	fator de escala da k -ésima árvore
M	número de árvores no ensemble
n_b	número de <i>feature bundles</i> após o EFB
$\mathcal{O}(\cdot)$	notação de ordem de complexidade
\mathbf{x}	vetor de entrada
\mathbf{y}	vetor de saída
z_j	combinação linear no neurônio j
h_j	ativação (saída) do neurônio oculto j
$w_{i,j}$	peso entre a entrada i e o neurônio j
b_j	viés do neurônio j
$\sigma(\cdot)$	função sigmoide
$\text{ReLU}(\cdot)$	função ReLU
$\tanh(\cdot)$	tangente hiperbólica
d_x	número de entradas do MLP
\mathbf{x}_ℓ	entrada no passo ℓ
\mathbf{h}_ℓ	estado oculto no passo ℓ
\mathbf{o}_ℓ	saída no passo ℓ
W, U, b	matrizes de pesos (W, U) e vetor de viés (b) na RNN
o_ℓ	saída escalar no passo ℓ
$o_{iL_{\text{seq}}}$	saída final para a i -ésima sequência

N_{seq}	número de sequências (amostras) no treinamento
L_{seq}	comprimento da sequência
\mathbf{h}_τ	estado oculto no tempo τ
C_τ	estado da célula no tempo τ
F_τ, I_τ, O_τ	portões de esquecimento, entrada e saída
\tilde{C}_τ	candidato a novo estado da célula
W_{x*}, W_{h*}	matrizes de pesos (entrada/estado) nos portões
b_*	vieses dos portões
\odot	produto de Hadamard (elemento a elemento)
w	janela de observação (comprimento da entrada)
nH	janela como múltiplo do horizonte H
L	largura das camadas totalmente conectadas (N-BEATS)
$\theta_\ell^b, \theta_\ell^f$	coeficientes de <i>backcast</i> e <i>forecast</i>
\hat{x}	componente de <i>backcast</i> estimada
\hat{y}	componente de <i>forecast</i> estimada
$g_\ell^b(\cdot), g_\ell^f(\cdot)$	funções de base (backcast/forecast)

Sumário

1	INTRODUÇÃO	29
1.1	Questões de pesquisa	31
1.2	Objetivos	31
1.3	Contribuições	32
2	FUNDAMENTAÇÃO TEÓRICA	33
2.1	Limitações do <i>throughput</i> e motivação para predição	33
2.2	Métricas de qualidade de canal	35
2.2.1	Potência de sinal de referência recebida (RSRP)	35
2.2.2	Indicador de intensidade de sinal recebido (RSSI)	35
2.2.3	Qualidade de sinal de referência recebida (RSRQ)	35
2.2.4	Relação sinal-ruído (SNR)	36
2.2.5	Indicador de qualidade de canal (CQI)	36
2.2.6	Considerações sobre as métricas de qualidade de canal	37
2.3	Previsão em séries temporais	38
2.3.1	Componentes e decomposição de séries temporais	40
2.4	Estacionariedade e teste de raiz Unitária	42
2.5	Métricas de avaliação	43
2.6	Validação cruzada para séries temporais	44
2.7	Métodos de predição de séries temporais	44
2.7.1	Métodos de referência (<i>Baseline</i>)	45
2.8	Modelos estatísticos	46
2.8.1	ARIMA	46
2.8.2	Regressão linear	47
2.9	Aprendizado de máquina	50
2.9.1	Árvores de decisão	50
2.9.2	Árvores de decisão com <i>gradient boosting</i>	52

2.9.3	LightGBM	53
2.10	Redes neurais e aprendizado profundo	54
2.10.1	<i>Perceptron</i> de múltiplas camadas	54
2.10.2	Redes neurais recorrentes	55
2.10.3	Arquitetura de Memória de Longo e Curto Prazo	57
2.10.4	A arquitetura N-BEATS	59
2.11	Considerações finais	62
3	TRABALHOS RELACIONADOS	63
3.1	Avaliação de modelos clássicos e séries temporais na previsão de <i>throughput</i> em redes 4G	63
3.1.1	Pré-processamento dos dados	64
3.1.2	Modelos e métricas de avaliação	65
3.1.3	Resultados	65
3.1.4	Análise crítica	65
3.2	Um <i>framework</i> para previsão de <i>throughput</i> em redes 5G	66
3.2.1	Coleta de dados em campo	68
3.2.2	Organização dos dados em grupos de atributos	69
3.2.3	Modelos e métricas de avaliação	70
3.2.4	Resultados	72
3.2.5	Análise crítica	72
3.3	Uso de aprendizado profundo na previsão de <i>throughput</i> em redes 5G	74
3.3.1	Pré-processamento dos dados	74
3.3.2	Modelos e métricas de avaliação	74
3.3.3	Resultados	75
3.3.4	Análise crítica	76
3.4	Modelos de aprendizado de máquina com pré-processamento e seleção de atributos	76
3.4.1	Pré-processamento dos dados	77
3.4.2	Modelos e métricas de avaliação	77
3.4.3	Resultados	78
3.4.4	Análise crítica	79
3.5	Previsão de <i>throughput</i> em comunicações 5G NR V2X com aprendizado profundo híbrido	80
3.5.1	Pré-processamento dos dados	80
3.5.2	Modelos e métricas de avaliação	80
3.5.3	Resultados	82
3.5.4	Análise crítica	82
3.6	Síntese da revisão da literatura	83

3.7	Síntese crítica	84
4	METODOLOGIA	87
4.1	Explorando o conjunto de dados	87
4.2	Pré-processamento	90
4.3	Protocolo experimental	94
4.4	Avaliação e registro de resultados	97
5	RESULTADOS	99
5.1	Análise global	99
5.1.1	Equilíbrio entre acurácia e custo computacional	99
5.1.2	Distribuição global dos erros	100
5.1.3	Comparação entre o <i>LighGBM</i> e o <i>Naive Mean</i>	101
5.1.4	Comparando todos os modelos ao <i>baseline</i>	105
5.2	Análises por cenários	106
5.2.1	Cenário de condução	106
5.2.2	Cenário estático	106
5.2.3	Cenário transmissão de vídeo	107
5.2.4	Cenário <i>download</i> de arquivo	107
5.3	Síntese dos resultados	108
6	CONCLUSÃO	111
	REFERÊNCIAS	115
	 APÊNDICES	 119
	APÊNDICE A – ESTATÍSTICAS DE DESEMPENHO DOS MO-	
	DELOS POR CENÁRIO	121

Capítulo 1

Introdução

Com a implantação da 5ª geração de redes móveis (5G) e a desativação gradual dos sistemas da 4ª geração de redes móveis (4G), observa-se um cenário em que o uso dessa tecnologia aumenta significativamente, impulsionado pela expansão da base de usuários e pela ampliação da cobertura. Enquanto a 4G opera na faixa de frequência entre 700 MHz e 2,6 GHz, a 5G permite a utilização de bandas inferiores a 6 GHz, conhecidas como *sub-6 GHz*, bem como das faixas de ondas milimétricas (*millimeter wave*) (mmWave), resultando em um aumento substancial da eficiência espectral e na promessa de alcançar taxas de transmissão de dados muito superiores às das gerações anteriores (GHOSH et al., 2019).

Apesar desses avanços, a implantação do 5G continua desigual, especialmente em áreas suburbanas e rurais. Além disso, as bandas de frequência mais altas sofrem maior atenuação de sinal e são mais sensíveis a obstruções, o que limita significativamente a cobertura em comparação com as tecnologias anteriores (BOUTIBA; BAGAA; KSENTINI, 2021). Nesse cenário, garantir um comportamento de rede eficiente e adaptativo torna-se cada vez mais desafiador, principalmente sob condições variáveis de tráfego, topologia e mobilidade.

Para enfrentar esses desafios e garantir um desempenho robusto, a previsão acurada do *throughput* (taxa de transferência de dados) surge como um componente crucial para o gerenciamento adaptativo de recursos e o fornecimento de Qualidade de Serviço (*Quality of Service*) (QoS) nessas redes. Previsões de carga de tráfego em horizontes de curto prazo podem impulsionar otimizações como escalonamento oportunista, controle com eficiência energética e gerenciamento dinâmico de *Handover* (HO)¹. Tais capacidades são essenciais

¹ *Handover* (HO) é o processo de transferência da conexão de rádio de um usuário entre células distintas para garantir a continuidade do serviço durante a mobilidade.

para viabilizar serviços sensíveis à latência, como a transmissão de vídeo em tempo real e veículos autônomos (SANTOS et al., 2020).

O *throughput* é inerentemente influenciado pelas condições do canal de rádio, tipicamente quantificadas por métricas como:

- ❑ Potência de Sinal de Referência Recebida (*Reference Signal Received Power*) (RSRP);
- ❑ Qualidade do Sinal de Referência Recebido (*Reference Signal Received Quality*) (RSRQ);
- ❑ Indicador de Intensidade do Sinal Recebido (*Received Signal Strength Indicator*) (RSSI);
- ❑ Relação Sinal-Ruído (*Signal-to-Noise Ratio*) (SNR);
- ❑ Indicador de Qualidade do Canal (*Channel Quality Indicator*) (CQI);

No entanto, essa relação é volátil. A mobilidade do Equipamento do Usuário (*User Equipment*) (UE) introduz uma dimensão temporal dinâmica que afeta significativamente a estimativa do *throughput*. Em cenários de alta velocidade, HO frequentes e flutuações abruptas na SNR prejudicam a estabilidade (NARAYANAN et al., 2020). Mesmo em velocidades de caminhada, variações sutis na orientação ou na direção do movimento podem obstruir o caminho de propagação — especialmente nas bandas mmWave — degradando a qualidade do serviço. Essas flutuações induzidas pela mobilidade tornam os modelos de desempenho estáticos insuficientes e ressaltam a necessidade de abordagens preditivas, sensíveis ao tempo e orientadas por dados.

A previsão de séries temporais tem como objetivo estimar valores futuros com base em observações passadas, sendo particularmente adequada para enfrentar esse desafio. Ao contrário das tarefas tradicionais de aprendizado de máquina, a previsão requer a preservação das dependências temporais, a expressão explícita da incerteza e o tratamento de fenômenos como não estacionaridade, sazonalidade e quebras estruturais (HYNDMAN; ATHANASOPOULOS, 2021). Para garantir uma avaliação confiável nesses cenários, a validação incremental (*walk-forward validation*) é comumente adotada (HEWAMALAGE; ACKERMANN; BERGMEIR, 2023), pois respeita a ordem temporal e mitiga o risco de vazamento de dados de observações futuras.

A evolução dos métodos de previsão foi fortemente impulsionada por grandes *benchmarks* empíricos, como as Competições Makridakis (também conhecidas como Competições M) (HYNDMAN, 2020), que demonstraram que a acurácia da previsão depende não apenas da escolha do paradigma de modelagem, seja ele estatístico ou baseado em aprendizado de máquina, mas também das capacidades funcionais do modelo, em especial sua habilidade de incorporar covariáveis externas (MAKRIDAKIS; SPILLOTIS; ASSIMAKOPOULOS, 2022).

No contexto de redes móveis, tais covariáveis podem incluir a velocidade do usuário, métricas de qualidade de sinal, hora do dia e identificadores específicos da célula (ELSHERBINY et al., 2020; NARAYANAN et al., 2020; BATOOL; FOUDA; FADLULLAH, 2024; SHARMA; PANDIT; TALLURI, 2025; YEASER; HASSAN, 2025). Além disso, conforme destacado por Januschowski et al. (2020), a distinção entre abordagens estatísticas e de aprendizado de máquina tornou-se cada vez mais ambígua, à medida que ambos os campos convergem por meio de técnicas compartilhadas e do surgimento de estruturas de modelagem híbridas ou unificadas.

Apesar dos avanços recentes, a literatura sobre previsão de *throughput* em redes móveis ainda carece de maior maturidade metodológica. Observa-se, em particular, a ausência de *baselines* ingênuos e estatísticos que sirvam como referência mínima de desempenho. Essas lacunas não apenas dificultam a comparação direta entre estudos, mas também comprometem a reprodutibilidade e a avaliação da real capacidade de generalização dos modelos propostos. Tais limitações evidenciam a necessidade de protocolos experimentais mais sistemáticos e transparentes, capazes de sustentar análises comparativas consistentes.

Diante disso, um número crescente de trabalhos tem explorado duas decisões de projeto complementares e independentes em arquiteturas de previsão: (i) o uso de modelos locais, treinados separadamente para cada série temporal, *versus* modelos globais, que aprendem padrões comuns a partir de múltiplas séries (YINGJIE; ABOLGHASEMI, 2024); e (ii) a inclusão ou exclusão de covariáveis externas, que fornecem informações adicionais ao modelo além da série-alvo. Este trabalho analisa sistematicamente ambas as dimensões, com especial atenção à segunda, ao comparar o desempenho de diferentes modelos tanto com quanto sem o uso de covariáveis.

1.1 Questões de pesquisa

As questões de pesquisa investigadas foram:

QP1 Quais métodos de previsão são mais eficazes sob diferentes condições de rede e mobilidade?

QP2 Quais são as vantagens e as limitações comparativas entre as abordagens de modelagem local e global?

1.2 Objetivos

Com base nas questões de pesquisa, este estudo estabelece os principais objetivos:

- O1 Apresentar e avaliar métodos de previsão de séries temporais para a predição do *throughput* em redes móveis;
- O2 Comparar modelos locais e globais para avaliar suas capacidades de generalização, escalabilidade e sensibilidade a variações de mobilidade e sinal;
- O3 Apoiar o gerenciamento proativo de recursos de rádio, alavancando previsões de *throughput* para aprimorar a Qualidade de Serviço (QoS).

1.3 Contribuições

Ao responder adequadamente a essas questões, este estudo apresenta as seguintes contribuições principais para o gerenciamento preditivo de recursos em redes celulares:

- ✓ **Benchmarking sistemático:** Uma avaliação abrangente de modelos estatísticos e de aprendizado de máquina, com foco nos *trade-offs* entre acurácia, robustez e custo computacional;
- ✓ **Diretrizes práticas de modelagem:** Análise dos fatores que influenciam a escolha entre abordagens locais e globais, considerando mobilidade, qualidade de sinal e restrições operacionais;

Capítulo 2

Fundamentação teórica

Este capítulo apresenta os fundamentos teóricos que sustentam o desenvolvimento desta pesquisa. Inicialmente, discute-se as limitações do *throughput* em redes móveis e a motivação para sua predição, destacando o papel das métricas de qualidade de canal como elementos essenciais para o gerenciamento eficiente de recursos de rádio e para a garantia da QoS. Em seguida, são introduzidos os principais métodos de previsão de séries temporais, agrupados em duas vertentes — estatística e aprendizado de máquina — cuja distinção, conforme salientado por Januschowski et al. (2020), tem se tornado cada vez mais tênue diante da convergência de técnicas e do surgimento de abordagens híbridas de modelagem.

2.1 Limitações do *throughput* e motivação para predição

As limitações do *throughput* constituem um dos principais desafios para o desempenho das redes móveis. Desde o contexto do 4G, estudos apontavam que, mesmo com aprimoramentos para múltiplos usuários, a taxa máxima teórica no enlace de descida (*downlink*) era de apenas 150 Mbps, o que equivale a aproximadamente 37 transmissões simultâneas de vídeo em alta definição a 4 Mbps cada. Tal capacidade mostrou-se insuficiente frente ao crescimento exponencial do tráfego de dados, impulsionado pela popularização de dispositivos inteligentes e pela predominância do vídeo. Estimativas indicavam que um usuário médio poderia consumir cerca de 1 TB de dados por ano, pressionando ainda mais a capacidade das redes (AGIWAL; ROY; SAXENA, 2016).

Com a chegada do 5G, embora avanços significativos tenham sido alcançados, a he-

terogeneidade tecnológica e a mobilidade mantêm o problema do *throughput* em aberto. Kalntis et al. (2024), em estudo de abrangência nacional realizado na Grécia com dados coletados em 2024, evidenciam que diferentes gerações de acesso (2G, 3G, 4G e 5G) ainda coexistem em uma mesma infraestrutura, o que gera complexidade na gestão de mobilidade e impacta negativamente as taxas de transmissão. Embora os autores apontem que 94% dos procedimentos de transferência de conexão HO ocorram entre setores 4G/5G, aproximadamente 6% ainda envolvem transições para redes legadas, com tempos de execução que podem ultrapassar 3,8 segundos, resultando em degradação perceptível da taxa de dados. Ademais, mais de 30% dos dispositivos ainda suportam apenas até 3G, obrigando a manutenção de camadas legadas e ampliando a variabilidade no desempenho.

Outro fator de complexidade está associado à integração massiva da Internet das Coisas (*Internet of Things*) (IoT) nas redes de quinta geração, que resulta em uma infraestrutura de grande escala e de difícil gerenciamento. Nesse cenário, Reshmi e Azath (2021) destacam que abordagens tradicionais de gestão reativa não são suficientes para lidar com falhas e degradações em ambientes densificados. Os autores propõem o uso de análise preditiva aplicada a parâmetros de rádio, como RSRP, SNR e variações de banda, de modo a antecipar anomalias que afetam diretamente o *throughput*. Essa capacidade de previsão permite acionar mecanismos de autorrecuperação de forma proativa, reduzindo tempos de indisponibilidade e assegurando maior resiliência operacional.

No contexto latino-americano, projeções recentes (GSMA, 2024) indicam que o tráfego de dados móveis quadruplicará entre 2023 e 2030, passando de 7 GB para 32 GB por conexão ao mês. Apesar desse crescimento, a taxa de adoção do 5G permanece modesta, representando apenas 5% das conexões em 2023, com previsão de superar 50% apenas no final da década. Esse cenário combina o aumento esperado do consumo de dados com uma transição tecnológica mais lenta, condicionada por fatores socioeconômicos e por limitações de espectro em diversos países. Ademais, persistem restrições significativas de acesso em muitas regiões, seja pela ausência de cobertura adequada, seja pela dificuldade de expansão da infraestrutura, o que amplia a desigualdade no uso efetivo das redes móveis. Essa conjunção de fatores reforça a pressão sobre o *throughput* e evidencia a necessidade de soluções preditivas capazes de assegurar a qualidade da conectividade e mitigar os impactos dessa heterogeneidade de acesso.

Dessa forma, observa-se que o *throughput* em redes móveis não apenas constitui um recurso limitado, mas também altamente variável, sujeito a interferências, mobilidade, limitações de espectro e falhas operacionais. Essa variabilidade compromete diretamente a QoS e a Qualidade da Experiência (*Quality of Experience*) (QoE). Isso evidencia a necessidade de mecanismos capazes de antecipar tais flutuações. Nesse contexto, a predição do *throughput* emerge como uma estratégia essencial para otimizar a alocação de recursos de rádio, apoiar decisões de mobilidade e garantir níveis adequados de desempenho em aplicações sensíveis à taxa e à latência.

2.2 Métricas de qualidade de canal

O desempenho de redes móveis depende fortemente da caracterização precisa da qualidade do enlace de rádio. Para esse fim, diferentes métricas foram desenvolvidas, cada uma refletindo aspectos específicos da intensidade, qualidade ou eficiência da transmissão do sinal. Essas medidas não apenas sustentam mecanismos fundamentais, como a seleção de célula e o HO, mas também fornecem insumos para algoritmos de previsão e para a otimização do uso de recursos de rádio (AFROZ et al., 2015). A seguir, são detalhadas as principais métricas de qualidade de canal utilizadas em redes móveis.

2.2.1 Potência de sinal de referência recebida (RSRP)

A RSRP, medida em dBm, é definida como a potência média do sinal de referência recebido de uma célula específica. Essa métrica é decisiva para avaliar a força do sinal, a cobertura da célula e as decisões de HO. Em redes 4G *Long-Term Evolution* (LTE), a RSRP é utilizada para ordenar células candidatas de acordo com a intensidade do sinal (AFROZ et al., 2015; ALRAIH et al., 2023).

No contexto do 5G, especialmente com a adoção da tecnologia mmWave, a RSRP apresenta elevada variabilidade, sendo sensível a fatores ambientais, como bloqueios por obstáculos ou pelo corpo humano, além da direção e velocidade do usuário. Assim, embora continue sendo uma métrica fundamental, a RSRP isoladamente não garante previsões confiáveis de desempenho, como o *throughput*. Por esse motivo, é frequentemente combinada com variáveis contextuais adicionais. Além disso, estudos mostram que valores mais altos de RSRP estão associados a maior eficiência energética do dispositivo, pois reduzem o consumo de energia por *bit* transmitido (NARAYANAN et al., 2020; SHARMA; PANDIT; TALLURI, 2025; YEASER; HASSAN, 2025).

2.2.2 Indicador de intensidade de sinal recebido (RSSI)

O RSSI, medido em dBm, corresponde à potência total recebida em banda, englobando tanto o sinal da célula servente quanto os das células vizinhas, além das interferências e do ruído. O RSSI fornece uma visão geral do ambiente de sinal e serve como referência auxiliar para a avaliação global do canal. Diferentemente da RSRP, que mede apenas a potência dos sinais de referência, o RSSI captura a potência agregada, o que o torna menos preciso para decisões de mobilidade, mas útil em análises de interferência e ruído (AFROZ et al., 2015).

2.2.3 Qualidade de sinal de referência recebida (RSRQ)

A RSRQ, expressa em dB, é definida pela razão entre a RSRP e o RSSI. Essa relação fornece uma avaliação integrada da qualidade do sinal ao considerar simultaneamente

potência, interferência e ruído, ponderados pelo número de blocos de recursos na largura de banda do RSSI (AFROZ et al., 2015; SHARMA; PANDIT; TALLURI, 2025).

Essa métrica desempenha papel essencial em decisões de seleção de célula e HO, sobretudo quando a potência isolada não oferece confiabilidade suficiente. Ao incluir os efeitos de interferência e ruído térmico, a RSRQ oferece uma visão mais realista da qualidade do enlace. Valores mais elevados correspondem a enlaces mais estáveis, com menor taxa de erros, menos retransmissões e maior *throughput* (SHARMA; PANDIT; TALLURI, 2025).

Em redes LTE, a RSRQ é tipicamente utilizada para ordenar células candidatas, com valores variando entre aproximadamente -40 dB (qualidade baixa) e -3 dB (qualidade alta). Tal como a RSRP, a RSRQ também apresenta maior instabilidade em cenários 5G, devido à complexidade dos ambientes de propagação (AFROZ et al., 2015; SHARMA; PANDIT; TALLURI, 2025).

2.2.4 Relação sinal-ruído (SNR)

A SNR, medida em dB, é definida como a razão entre a potência do sinal útil e a do ruído no canal. É uma métrica essencial para caracterizar a qualidade do enlace, pois impacta diretamente a escolha do esquema de modulação e codificação, influenciando taxas de erro e *throughput*. Em redes LTE, a SNR apresenta proporcionalidade média com a RSRP, reforçando seu papel como indicador da qualidade do canal. Valores mais altos de SNR possibilitam o uso de modulações mais sofisticadas, reduzindo retransmissões e aumentando a taxa de transmissão.

No 5G, a relevância da SNR é ainda maior, especialmente em cenários de alta mobilidade, como comunicações veiculares. Estudos apontam que maiores valores de SNR se associam a ganhos expressivos de desempenho tanto no *uplink* quanto no *downlink* (AFROZ et al., 2015; SHARMA; PANDIT; TALLURI, 2025; YEASER; HASSAN, 2025).

2.2.5 Indicador de qualidade de canal (CQI)

O CQI é uma métrica reportada pelo dispositivo do usuário à estação base, indicando a maior taxa de modulação e codificação que pode ser utilizada com probabilidade de erro de bloco inferior a 10%. Ele sintetiza informações sobre a relação sinal-ruído e as capacidades do receptor, desempenhando papel central na adaptação do enlace e no escalonamento de pacotes. Dessa forma, o CQI permite otimizar taxas de dados e maximizar o *throughput* (AFROZ et al., 2015; RACA et al., 2020).

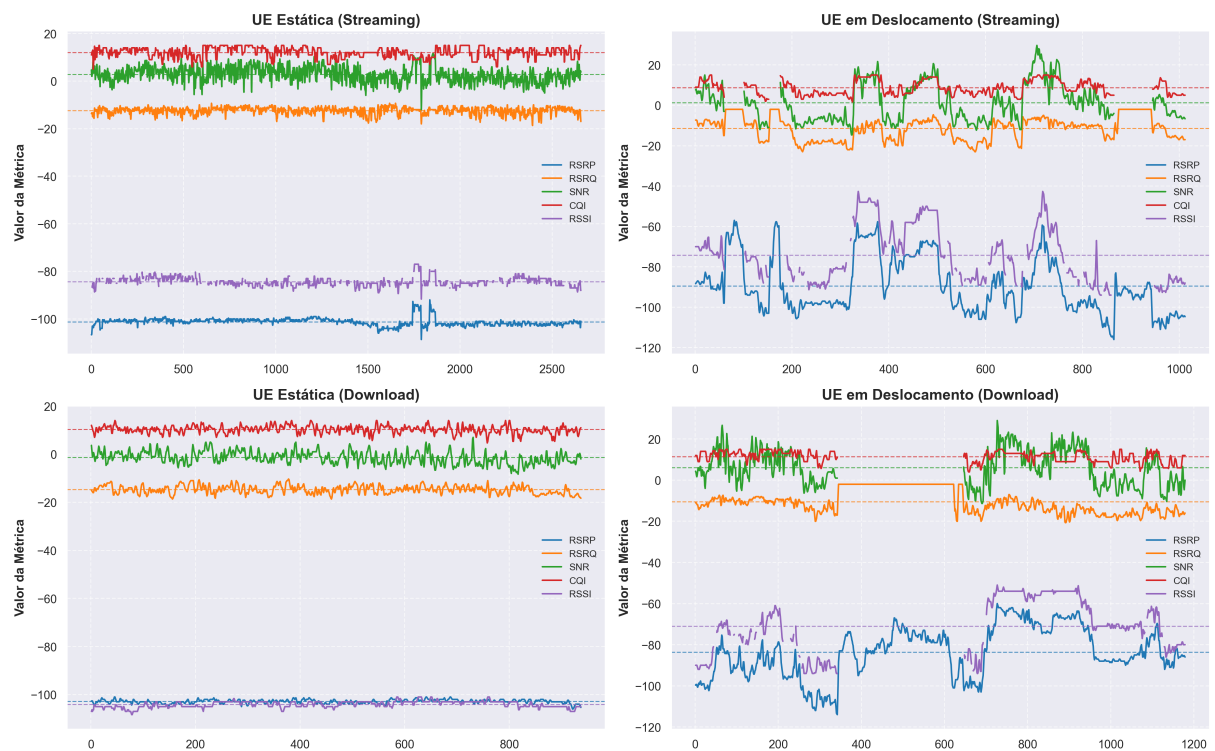
Valores elevados de CQI refletem condições de canal favoráveis, permitindo modulações mais eficientes e menor taxa de erros. Em cenários de mobilidade, como comunicações veiculares, o CQI está diretamente relacionado a fatores como velocidade, intensidade do sinal e SNR, servindo como parâmetro essencial para previsões de *throughput* e para a

gestão adaptativa de recursos de rede (AFROZ et al., 2015; RACA et al., 2020; SANTOS et al., 2020; SHARMA; PANDIT; TALLURI, 2025; YEASER; HASSAN, 2025).

2.2.6 Considerações sobre as métricas de qualidade de canal

Em conjunto, as métricas RSRP, RSRQ, RSSI, SNR e CQI oferecem perspectivas complementares sobre o estado do canal de rádio. Enquanto algumas refletem a intensidade do sinal (RSRP, RSSI), outras avaliam sua qualidade considerando interferência e ruído (RSRQ, SNR), ao passo que o CQI representa de forma adaptativa a capacidade do enlace. A Figura 1 ilustra a variação temporal dessas métricas em cenários de *streaming* e *download*, com a UE em condição estática ou em deslocamento, evidenciando diferenças significativas no comportamento do canal de acordo com o contexto. Essas medidas constituem a base para a construção de modelos de previsão em séries temporais, discutidos na próxima seção.

Figura 1 – Exemplos de evolução temporal das métricas de qualidade de canal (RSRP, RSRQ, RSSI, SNR e CQI) em diferentes cenários experimentais: *streaming* e *download*, com a UE em condição estática ou em deslocamento. O eixo temporal está expresso em segundos, e as linhas tracejadas indicam os valores médios de cada métrica.



Fonte: Autoria própria.

2.3 Previsão em séries temporais

Uma série temporal consiste em um conjunto de observações de um fenômeno de interesse, coletadas sequencialmente ao longo do tempo (HAMILTON, 2020). Em geral, a coleta de dados de uma série temporal começa em uma data específica (por exemplo, $t = 1$) e se encerra em outra (por exemplo, $t = T$), resultando na amostra observada apresentada na Equação 1.

$$(y_1, y_2, \dots, y_T) \quad (1)$$

Considera-se, frequentemente, que a amostra observada na Equação 1 é um segmento finito de uma sequência duplamente infinita, denotada por $\{y_t\}_{t=-\infty}^{\infty}$ (Equação 2). Isso ocorre sob a suposição de que poderíamos ter obtido observações anteriores ou posteriores se o tempo de observação fosse estendido.

$$\{y_t\}_{t=-\infty}^{\infty} = \{\dots, y_{-1}, y_0, y_1, y_2, \dots, y_T, y_{T+1}, y_{T+2}, \dots\} \quad (2)$$

Para a tarefa de previsão, três abordagens principais de modelagem podem ser utilizadas (HYNDMAN; ATHANASOPOULOS, 2021):

- Modelo de Série Temporal — baseia-se exclusivamente nos valores passados da própria variável, sem o uso de variáveis externas. Um exemplo genérico pode ser expresso como:

$$y_{t+1} = f(y_t, y_{t-1}, y_{t-2}, \dots, \text{erro})$$

em que a previsão de y_{t+1} depende apenas de observações anteriores da série e de um termo de erro aleatório.

- Modelo Explicativo — utiliza variáveis preditoras para explicar as causas das variações observadas. Sua formulação genérica pode ser representada por:

$$y_{t+1} = f(X_t^1, X_t^2, X_t^3, \dots, \text{erro})$$

onde X_t^1, X_t^2, X_t^3 representam variáveis externas, como temperatura, hora do dia ou fatores econômicos.

- Modelo Misto — combina os valores passados da série temporal com variáveis preditoras externas. Sua forma genérica pode ser escrita como:

$$y_{t+1} = f(y_t, y_{t-1}, X_t^1, X_t^2, \dots, \text{erro})$$

de modo que a previsão se baseia tanto na dependência temporal quanto em fatores externos.

A escolha da abordagem de modelagem mais adequada deve considerar diversos fatores relacionados ao contexto do problema e aos objetivos da pesquisa. Modelos de séries temporais podem ser particularmente recomendados quando o funcionamento do sistema em estudo não é plenamente compreendido, quando há elevada incerteza na previsão futura das variáveis preditoras, quando o objetivo central consiste em estimar valores futuros sem necessariamente explicar suas causas ou, ainda, quando esse tipo de modelo apresenta desempenho preditivo superior em cenários específicos (HYNDMAN; ATHANASOPOULOS, 2021).

No contexto deste trabalho, adota-se a denominação **Modelos Univariados** para designar os modelos de séries temporais que utilizam exclusivamente os valores passados da própria variável. Por outro lado, os modelos mistos, que combinam informações históricas da série com variáveis preditoras externas, serão denominados **Modelos Covariados**.

A modelagem preditiva para séries temporais requer, inicialmente, a divisão dos dados em um conjunto de treinamento e um horizonte de previsão, um processo que é ilustrado na Figura 2. O objetivo central, formalizado na Equação 3, consiste em estimar o vetor de observações futuras, representadas por \hat{y} , para um horizonte de previsão (H) que se inicia na origem de tempo T . Tal estimativa é produto de um modelo de previsão, descrito pela função f , que processa a sequência de dados históricos disponíveis, conforme definido na Equação 1.

Figura 2 – Exemplo de cenário de previsão com região de treinamento, origem e horizonte de previsão.



Fonte: Adaptado de Hewamalage, Ackermann e Bergmeir (2023).

$$[\hat{y}_{T+1}, \hat{y}_{T+2}, \dots, \hat{y}_{T+H}] = f(y_1, y_2, \dots, y_T) \quad (3)$$

2.3.1 Componentes e decomposição de séries temporais

A análise estrutural de uma série temporal (Y_t) fundamenta-se em sua decomposição em componentes sistemáticos, que capturam padrões previsíveis, e em um componente não sistemático, que representa a aleatoriedade. Essa decomposição é essencial para compreender a dinâmica dos dados e para a construção de modelos de previsão adequados. De modo clássico, os componentes sistemáticos são divididos em três categorias, como no exemplo ilustrado na Figura 3.

Na Figura 3, observa-se que os três componentes fundamentais — tendência, sazonalidade e resíduo — são apresentados em painéis separados e, quando combinados, reconstroem a série original exibida no painel superior. As barras cinzas posicionadas à esquerda de cada painel evidenciam as diferenças relativas de escala entre os componentes. Embora todas tenham o mesmo comprimento, seu tamanho aparente varia de acordo com a escala de cada gráfico. Isso mostra que a variabilidade do resíduo é significativamente menor que a da série original, reforçando seu caráter de ruído (HYNDMAN; ATHANASOPOULOS, 2021).

A tendência (T_t) representa a evolução de longo prazo do nível médio da série, manifestando-se como movimentos ascendentes (crescimento) ou descendentes (declínio).

A sazonalidade (S_t) reflete oscilações regulares em intervalos de tempo fixos, como meses, trimestres ou anos, sendo observada, por exemplo, no aumento da demanda por energia no inverno ou nas vendas do varejo em períodos festivos. Essa componente pode variar ao longo do tempo, de modo que anos consecutivos apresentem padrões semelhantes, enquanto períodos mais distantes revelem diferenças significativas (HYNDMAN; ATHANASOPOULOS, 2021).

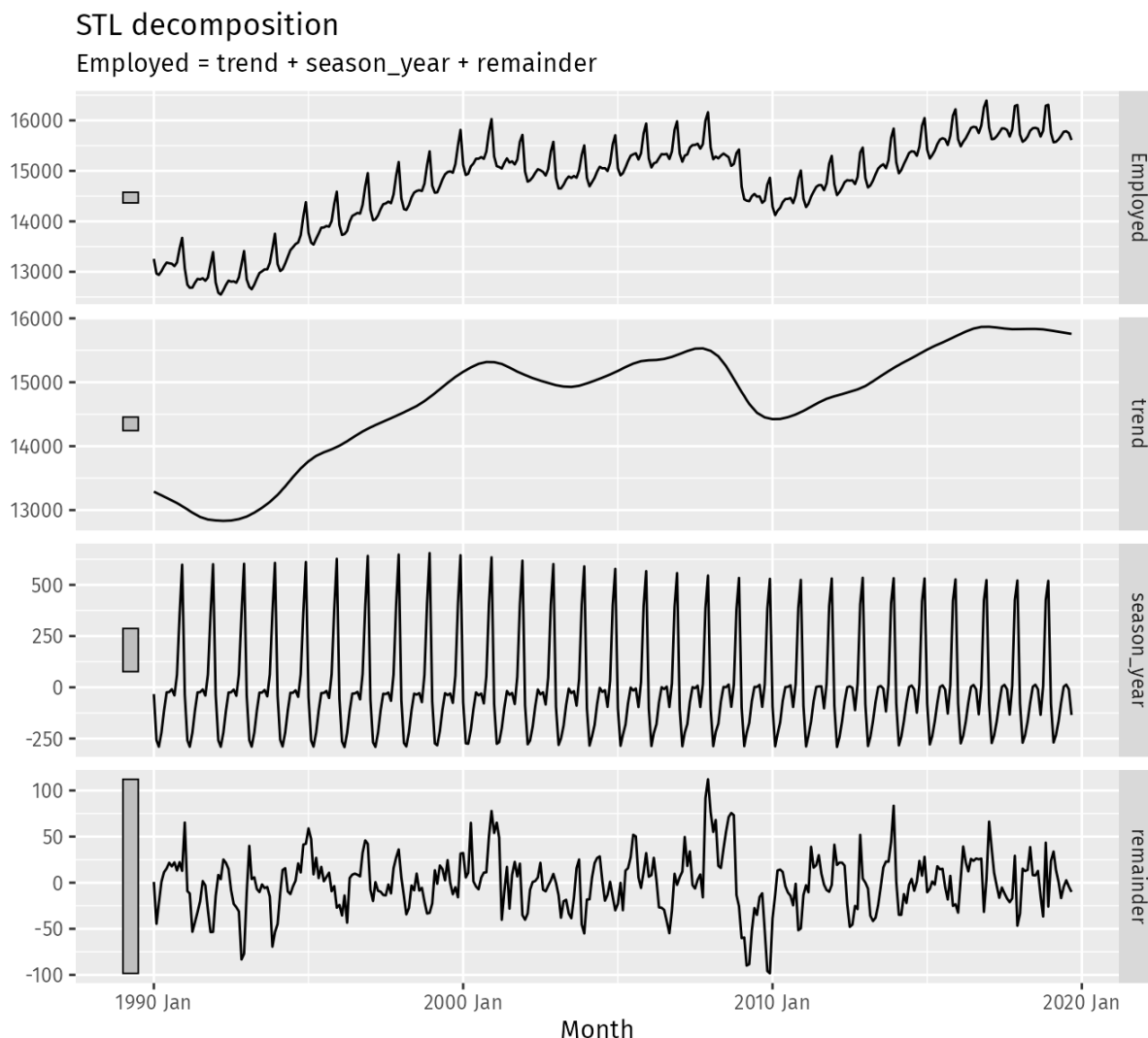
O componente cíclico (C_t) descreve flutuações de prazo mais longo que não possuem periodicidade fixa, geralmente associadas a conjunturas econômicas amplas, como ciclos de expansão e recessão. Distingue-se, portanto, da sazonalidade pela ausência de regularidade temporal definida.

O resíduo (e_t) corresponde à parcela não sistemática da série, ou seja, às variações aleatórias e imprevisíveis. Formalmente, os resíduos são obtidos como a diferença entre os valores observados (y_t) e os valores ajustados (\hat{y}_t), estimados a partir dos demais componentes:

$$e_t = y_t - \hat{y}_t.$$

A relação entre a série original e seus componentes pode ser expressa de duas formas principais. No modelo aditivo, os componentes são somados, sendo apropriado quando a magnitude das flutuações é independente do nível da tendência. Já no modelo multipli-

Figura 3 – Exemplo de decomposição de uma série temporal em seus componentes: série original, tendência, sazonalidade e resíduos. As barras cinzas à esquerda indicam a escala relativa de cada componente em comparação à série original.



Fonte: Adaptado de Hyndman e Athanasopoulos (2021).

cativo, os componentes são combinados por multiplicação, o que é mais adequado quando a amplitude das variações cresce proporcionalmente ao nível da série. As formulações são dadas por:

$$\text{Aditivo: } Y_t = T_t + S_t + C_t + e_t$$

$$\text{Multiplicativo: } Y_t = T_t \times S_t \times C_t \times e_t$$

Uma das abordagens mais difundidas para a decomposição prática de séries temporais é a Decomposição Sazonal e de Tendência por Loess (*Seasonal-Trend decomposition using Loess*) (STL), que utiliza regressões locais (*Loess*)¹ para extrair tendência e sazonalidade

¹ *Loess* (*Locally Estimated Scatterplot Smoothing*) é um método de suavização não paramétrico que

de forma adaptativa. Essa técnica é robusta a valores atípicos, apropriada para séries longas e com sazonalidades complexas, e amplamente aplicada em análises exploratórias e em métodos modernos de previsão (HYNDMAN; ATHANASOPOULOS, 2021).

A análise dos resíduos constitui uma etapa indispensável na validação de modelos de séries temporais. A ausência de padrões ou de autocorrelação residual indica que o modelo conseguiu capturar adequadamente a estrutura sistemática dos dados, servindo como critério fundamental para avaliar sua adequação e confiabilidade (HYNDMAN; ATHANASOPOULOS, 2021; JOSEPH, 2022).

2.4 Estacionariedade e teste de raiz Unitária

A estacionariedade é uma propriedade fundamental de uma série temporal, caracterizada pela invariância de suas propriedades estatísticas ao longo do tempo. No contexto da estacionariedade fraca, a mais relevante para análises aplicadas, essa condição exige que a média ($E[Y_t] = \mu$) e a variância ($\text{Var}(Y_t) = \sigma^2$) sejam constantes e que a autocovariância entre duas observações ($\gamma(k) = \text{Cov}(Y_t, Y_{t-k})$) dependa exclusivamente da defasagem (k) entre elas, e não de sua posição no tempo. Sob essa premissa, a série não exibe tendências ou sazonalidades previsíveis, flutuando com dispersão estável em torno de um nível médio constante. Tal estabilidade é um pré-requisito para a aplicação de muitos modelos de previsão, pois simplifica significativamente a modelagem e a inferência (HAMILTON, 2020; JOSEPH, 2022).

A violação da estacionariedade é frequentemente causada pela existência de uma raiz unitária no processo gerador dos dados. Esse fenômeno pode ser ilustrado pelo processo autorregressivo de primeira ordem (AR(1)), descrito pela equação:

$$y_t = \phi y_{t-1} + \epsilon_t, \quad (4)$$

onde o coeficiente ϕ determina a dinâmica da série. Se $|\phi| < 1$, o processo é estacionário, pois os choques aleatórios (ϵ_t) têm um impacto transitório. Contudo, quando $\phi = 1$, o processo contém uma raiz unitária e os choques se acumulam permanentemente, caracterizando um passeio aleatório (*random walk*). Neste cenário, a variância da série cresce indefinidamente com o tempo, tornando-a não estacionária (HYNDMAN, 2020; JOSEPH, 2022).

Para a detecção formal de uma raiz unitária, o Teste de Dickey-Fuller Aumentado (ADF) é um procedimento estatístico amplamente empregado, derivado dos trabalhos seminais de Dickey e Fuller (DICKKEY; FULLER, 1979; FULLER, 1996). O teste avalia a hipótese nula (H_0) de que a série possui uma raiz unitária (sendo, portanto, não estacionária) contra a hipótese alternativa (H_1) de estacionariedade. A decisão é baseada

ajusta regressões locais em subconjuntos da série, ponderando as observações pela proximidade. Essa técnica permite capturar tendências e sazonalidades de forma flexível, mesmo em séries ruidosas ou não lineares.

no valor-p resultante: um valor-p inferior a 5% (nível de significância) fornece evidências para rejeitar H_0 , indicando que a série é estacionária. Caso contrário, não se rejeita H_0 , e a série é considerada não estacionária. A técnica mais comum para tornar séries estacionárias é a diferenciação, que consiste em substituir a série original pela série de suas diferenças entre observações consecutivas:

$$(1 - B)y_t = y_t - y_{t-1},$$

onde B é o operador de defasagem.

2.5 Métricas de avaliação

A avaliação do desempenho dos modelos preditivos em séries temporais frequentemente utiliza a RMSE e o Erro Médio Absoluto (*Mean Absolute Error*) (MAE). Ambas mantêm a unidade original da variável, o que facilita sua interpretação, mas diferem na maneira de penalizar os erros: o MAE é robusto a *outliers* e favorece previsões centradas na mediana, enquanto o RMSE enfatiza grandes desvios e tende a gerar previsões voltadas à média (HYNDMAN; KOEHLER, 2006). De acordo com Hyndman e Athanasopoulos (2021), o MAE é simples de interpretar e adequado para comparações homogêneas, enquanto o RMSE, embora mais interpretativamente complexo, é amplamente adotado por penalizar de forma mais intensa os erros extremos (HYNDMAN; ATHANASOPOULOS, 2021; HYNDMAN, 2020).

O RMSE é definido como:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{t=1}^N (y_t - \hat{y}_t)^2}. \quad (5)$$

Por elevar os erros ao quadrado, o RMSE atribui maior penalidade a desvios de grande magnitude.

Já o MAE é dado por:

$$\text{MAE} = \frac{1}{N} \sum_{t=1}^N |y_t - \hat{y}_t|, \quad (6)$$

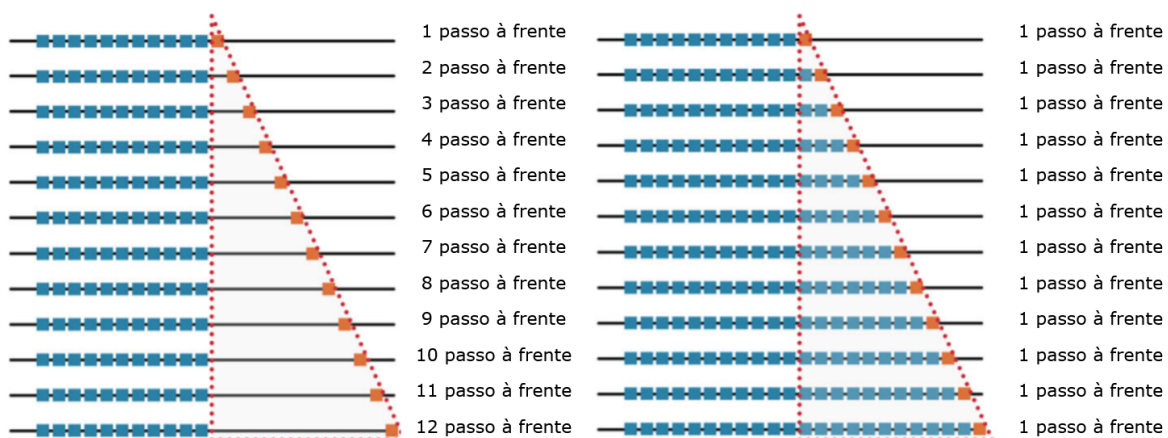
onde N é o tamanho da amostra e t indexa as observações.

É importante ressaltar que a interpretação isolada de RMSE e MAE pode ser limitada, especialmente em conjuntos compostos por séries de diferentes escalas. Nesse sentido, Hewamalage, Ackermann e Bergmeir (2023) destacam que a comparação com métodos de referência simples, como previsores ingênuos (*naïve*), fornece um ponto de base essencial para avaliar a real vantagem de modelos mais sofisticados. Assim, verificar se um modelo supera consistentemente tais métodos de referência ajuda a garantir que os ganhos observados não resultem apenas da complexidade adicional, mas sim de maior capacidade preditiva.

2.6 Validação cruzada para séries temporais

Na Figura 4, são apresentadas as duas principais metodologias para a avaliação de modelos preditivos. A figura à esquerda ilustra a avaliação com origem fixa, em que o conjunto de dados históricos utilizado para o treino permanece constante, isto é, o ponto no tempo definido como origem, denotado por T , é fixo e a previsão é gerada uma única vez para o futuro. A figura à direita, por sua vez, representa a avaliação com origem rolante, na qual a origem da previsão avança ao longo do tempo, fazendo com que T varie e permitindo a geração de múltiplos cenários de teste com dados mais recentes. Essa abordagem possibilita uma avaliação contínua e mais realista do desempenho do modelo (HEWAMALAGE; ACKERMANN; BERGMEIR, 2023).

Figura 4 – Comparação entre as metodologias de avaliação por origem fixa (à esquerda) e por origem rolante (à direita), considerando horizonte de previsão igual a $H = 1$.



Fonte: Adaptado de Hewamalage, Ackermann e Bergmeir (2023).

No contexto da abordagem com origem rolante, a forma como a região de treinamento é atualizada a cada novo passo gera duas implementações distintas. A primeira é a janela rolante, na qual o tamanho do conjunto de dados de treino é mantido constante, e, a cada avanço da origem, uma nova observação é incorporada enquanto a mais antiga é descartada. A segunda abordagem é a janela expansível, que consiste no aumento contínuo do conjunto de treino, acumulando novas observações a cada passo sem eliminar os dados anteriores (HYNDMAN; ATHANASOPOULOS, 2021).

2.7 Métodos de predição de séries temporais

Esta subseção apresenta uma visão geral das principais abordagens para a previsão de séries temporais, organizadas em três categorias: métodos de referência (*baseline*),

modelos estatísticos e modelos baseados em aprendizado de máquina. Essas abordagens são frequentemente utilizadas em estudos comparativos, servindo como referência para avaliar o desempenho de técnicas mais sofisticadas em relação a estratégias de previsão mais simples.

2.7.1 Métodos de referência (*Baseline*)

Alguns métodos de previsão possuem estrutura simples, mas apresentam desempenho surpreendentemente eficaz. Entre eles, destacam-se os modelos listados abaixo, frequentemente utilizados como referência quando comparados com modelos mais sofisticados.

- **Média simples ou Média Ingênua (*Naive mean*)**: Neste procedimento, todas as previsões futuras são estimadas como a média dos valores observados anteriormente. Considerando os dados históricos y_1, \dots, y_N , as previsões são expressas por

$$\hat{y}_{T+h|T} = \bar{y} = \frac{1}{N} \sum_{t=1}^N y_t.$$

A notação $\hat{y}_{T+h|T}$ refere-se à previsão do valor y_{T+h} condicionada aos dados disponíveis até o instante T (HYNDMAN; ATHANASOPOULOS, 2021).

- **Método Ingênuo (*Naive*)**: Neste método, as previsões futuras são assumidas iguais ao valor da última observação disponível, ou seja,

$$\hat{y}_{T+h|T} = y_T.$$

Apesar de sua simplicidade, essa abordagem apresenta desempenho surpreendentemente eficaz em muitas séries temporais, especialmente nas áreas econômica e financeira. Isso ocorre porque, quando os dados seguem um modelo de passeio aleatório — em que o valor futuro é igual ao valor presente acrescido de um termo de erro aleatório — a melhor previsão para o próximo valor é justamente o último valor observado (HYNDMAN; ATHANASOPOULOS, 2021).

- **Método Deriva Ingênua (*Naive Drift*)**: Uma variação do método ingênuo consiste em permitir que as previsões aumentem ou diminuam ao longo do tempo, de modo que a taxa de variação (*drift*) seja igual à média das variações observadas nos dados históricos. Assim, a previsão para o instante $T + h$ é dada por

$$\hat{y}_{T+h|T} = y_T + h \cdot \frac{y_T - y_1}{N - 1},$$

onde h representa o horizonte de previsão. Esse método é equivalente a traçar uma linha reta entre a primeira e a última observação disponível e projetá-la para o futuro, considerando uma tendência linear constante (HYNDMAN; ATHANASOPOULOS, 2021).

- **Suavização Exponencial:** A suavização exponencial atribui maior peso às observações mais recentes por meio de uma média ponderada com decaimento exponencial. A forma simples é dada por:

$$\hat{y}_{t+1} = \alpha_{\text{ES}} y_t + (1 - \alpha_{\text{ES}}) \hat{y}_t,$$

onde $\alpha_{\text{ES}} \in (0, 1)$ é o parâmetro de suavização. Para séries com tendência, pode-se utilizar a suavização exponencial dupla; para séries com tendência e sazonalidade, a suavização tripla (Holt–Winters) (HYNDMAN; ATHANASOPOULOS, 2021).

2.8 Modelos estatísticos

Modelos estatísticos clássicos, como ARIMA e regressão linear, constituem a base fundamental para a previsão de séries temporais, servindo como um *benchmark* confiável para a avaliação de técnicas mais complexas, como as de aprendizado de máquina. Caracterizados pela simplicidade, interpretabilidade e eficiência, esses métodos são frequentemente o ponto de partida em análises práticas e se mantêm competitivos em cenários onde os padrões são estáveis e lineares (HYNDMAN; ATHANASOPOULOS, 2021).

2.8.1 ARIMA

O modelo ARIMA (*AutoRegressive Integrated Moving Average*) representa uma classe de modelos lineares capaz de descrever séries temporais estacionárias e não estacionárias. Proposto por Box et al. (2015), ele generaliza os processos autorregressivos (AR) e de médias móveis (MA) ao incorporar um componente de integração para tratar séries não estacionárias. A abordagem se baseia em modelar a dependência estocástica nos dados, ou seja, a relação entre uma observação e os valores passados da própria série e de erros aleatórios.

A estrutura completa de um modelo ARIMA(p, d, q) é representada de forma compacta utilizando a notação do operador de defasagem (B), onde $By_t = y_{t-1}$. A forma geral do modelo é:

$$\phi_p(B)(1 - B)^d y_t = c + \theta_q(B) \epsilon_t \quad (7)$$

onde:

- $\phi_p(B) = (1 - \phi_1 B - \dots - \phi_p B^p)$ é o polinômio autorregressivo de ordem p ;
- $\theta_q(B) = (1 + \theta_1 B + \dots + \theta_q B^q)$ é o polinômio de médias móveis de ordem q ;
- $(1 - B)^d$ é o operador de diferenciação de ordem d , aplicado para induzir a estacionariedade;

- ϵ_t é um processo de ruído branco (*white noise*) com média zero e variância constante σ_ϵ^2 .

As raízes dos polinômios $\phi_p(B)$ e $\theta_q(B)$ devem estar fora do círculo unitário para garantir as condições de estacionariedade e invertibilidade do modelo, respectivamente.

A flexibilidade do ARIMA permite que ele abranja diversos outros modelos como casos específicos, o que demonstra sua generalidade, conforme ilustrado na Tabela 1.

Tabela 1 – Casos Especiais do Modelo ARIMA

Modelo	Notação ARIMA(p, d, q)
Ruído Branco	ARIMA(0,0,0) sem constante
Passeio Aleatório (<i>Random Walk</i>)	ARIMA(0,1,0) sem constante
Passeio Aleatório com Deriva (<i>Drift</i>)	ARIMA(0,1,0) com constante
Autorregressivo (AR)	ARIMA($p,0,0$)
Médias Móveis (MA)	ARIMA(0,0, q)

Fonte: Adaptado de Hyndman (2020).

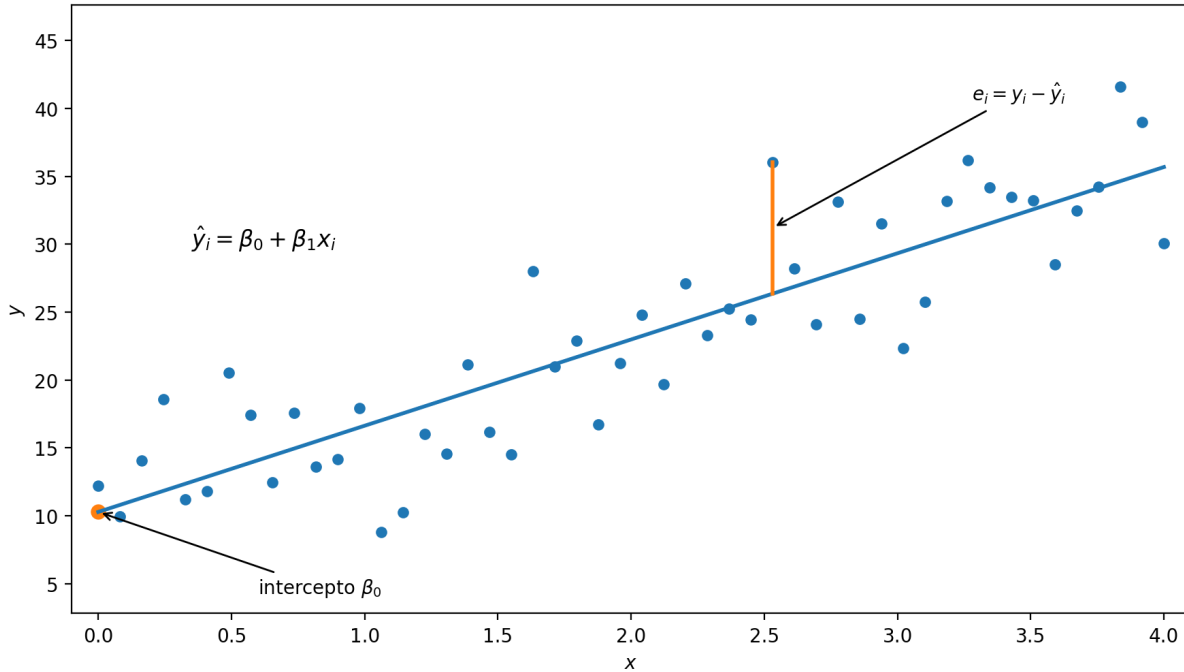
Mais do que apenas um modelo, a abordagem ARIMA é um processo iterativo de modelagem que segue três estágios, guiado pelo princípio da parcimônia — a preferência pelo modelo mais simples que ofereça um ajuste adequado.

1. Identificação: Nesta fase, determina-se a ordem de diferenciação (d) necessária para tornar a série estacionária. Em seguida, analisam-se as Funções de Autocorrelação (*Autocorrelation Function*) (ACF) e a Autocorrelação Parcial (*Partial Autocorrelation Function*) (PACF) da série estacionária para hipotetizar as ordens p e q do modelo.
2. Estimação: Uma vez identificada uma estrutura provisória para o modelo ARIMA(p, d, q), seus parâmetros (ϕ_1, \dots, ϕ_p e $\theta_1, \dots, \theta_q$) são estimados, geralmente por meio do método de máxima verossimilhança.
3. Verificação de diagnóstico: Esta é uma etapa crucial. O modelo estimado é avaliado pela análise de seus resíduos ($\hat{\epsilon}_t = y_t - \hat{y}_t$). Se o modelo estiver bem especificado, os resíduos devem se comportar como ruído branco, isto é, sem autocorrelação. A análise da ACF dos resíduos e testes estatísticos, como o teste de Ljung-Box, são utilizados para essa verificação (BOX et al., 2015; HYNDMAN; ATHANASOPOULOS, 2021). Se houver estrutura remanescente nos resíduos, o ciclo é repetido, retornando à fase de identificação para refinar o modelo.

2.8.2 Regressão linear

A Figura 5 ilustra o caso de uma regressão linear simples, em que a relação entre uma única variável preditora e a resposta é modelada por uma reta de ajuste.

Figura 5 – Exemplo de regressão linear simples.



Fonte: Adaptado de Hyndman e Athanasopoulos (2021).

A regressão linear modela a relação entre uma variável dependente e uma ou mais variáveis independentes por meio de uma função linear (MONTGOMERY; PECK; VINING, 2012; HASTIE; TIBSHIRANI; FRIEDMAN, 2009). A forma geral é:

$$\hat{y} = \beta_0 + \sum_{j=1}^m X_j \beta_j,$$

onde \hat{y} é o valor predito, m é o número de preditores, X_j são as variáveis independentes, β_j os respectivos coeficientes, e β_0 o termo constante (intercepto).

Os parâmetros β são estimados a partir do conjunto de treinamento

$$\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^N, \quad \text{com } \mathbf{x}_i \in \mathbb{R}^m,$$

pelo método dos Mínimos Quadrados Ordinários (MQO) (SEBER; LEE, 2003), que minimiza a Soma dos Quadrados dos Resíduos (SQR):

$$\text{SQR} = \sum_{i=1}^N (y_i - \hat{y}_i)^2,$$

em que $e_i = y_i - \hat{y}_i$ são os resíduos. Minimizar as distâncias euclidianas ao quadrado evita o cancelamento de desvios positivos e negativos. Uma vantagem do MQO é a solução analítica em forma fechada (quando $X^T X$ é inversível), $\hat{\beta}_{\text{MQO}} = (X^T X)^{-1} X^T \mathbf{y}$.

A validade das inferências do modelo requer, em geral (KUTNER et al., 2005):

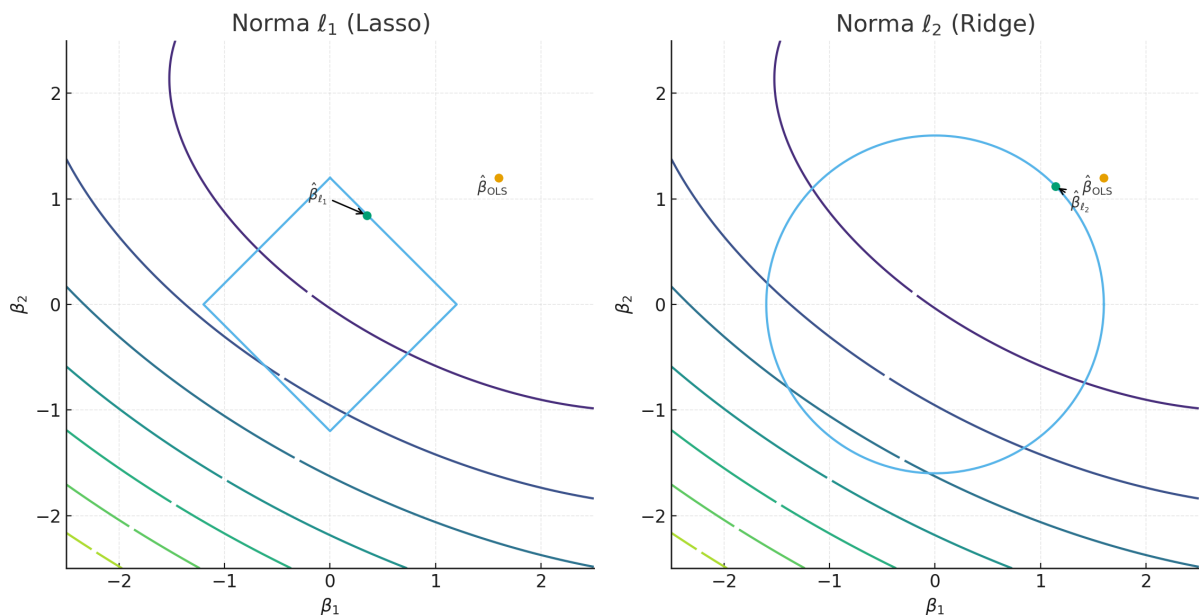
- ❑ **Linearidade:** relação linear entre preditores e resposta;
- ❑ **Normalidade dos erros:** resíduos aproximadamente normais;
- ❑ **Homocedasticidade:** variância constante dos erros;
- ❑ **Ausência de autocorrelação:** erros independentes;
- ❑ **Ausência de multicolinearidade:** preditores não altamente correlacionados.

Para fins puramente preditivos, alguns desses pressupostos podem ser relaxados; a linearidade, porém, permanece crucial.

Entretanto, em cenários de multicolinearidade, a estimação por MQO pode se tornar instável. A regularização mitiga esse problema ao adicionar uma penalização à função de custo SQR, favorecendo soluções com coeficientes de menor magnitude. As formas mais comuns são ℓ_1 (Lasso) (TIBSHIRANI, 1996) e ℓ_2 (Ridge) (HOERL; KENNARD, 1970): a primeira tende a anular alguns coeficientes (seleção interna de características), enquanto a segunda os encolhe em direção a zero, aumentando a robustez.

Na Figura 6, os contornos representam níveis de erro no espaço dos coeficientes (β_1, β_2) , cujo centro é $\hat{\beta}_{\text{MQO}}$ (do inglês, $\hat{\beta}_{\text{OLS}}$), a solução sem penalização. As regiões de restrição são um diamante (ℓ_1 , Lasso) e um círculo (ℓ_2 , Ridge). A geometria com cantos da bola ℓ_1 favorece soluções esparsas (coeficientes exatamente nulos), ao passo que a bola ℓ_2 , suave, promove encolhimento sem necessariamente zerá-los.

Figura 6 – Diferença geométrica entre as regularizações ℓ_1 (Lasso) e ℓ_2 (Ridge). O centro representa a solução por MQO (do inglês, $\hat{\beta}_{\text{OLS}}$), sem penalização.



Fonte: Adaptado de Joseph (2022).

2.9 Aprendizado de máquina

O aprendizado de máquina reúne técnicas capazes de identificar padrões complexos em séries temporais a partir dos dados, sem depender de hipóteses rígidas sobre a estrutura do processo gerador. Em contraste com os modelos estatísticos tradicionais, que geralmente se apoiam em pressupostos lineares e estacionários, os métodos de Aprendizado de máquina (*Machine Learning*) (ML) oferecem maior flexibilidade para lidar com múltiplas covariáveis, não linearidades e dinâmicas de longo prazo (HYNDMAN; ATHANASOPOULOS, 2021).

Entretanto, tais modelos também impõem desafios, como a maior demanda computacional, a necessidade de grandes volumes de dados para treinamento e a dificuldade de interpretação. Nas subseções seguintes, são descritas algumas das abordagens mais representativas, desde algoritmos clássicos baseados em árvores até arquiteturas de aprendizado profundo desenvolvidas especificamente para dados sequenciais.

2.9.1 Árvores de decisão

As árvores de decisão (*Decision Trees*) constituem uma das técnicas mais consolidadas de aprendizado supervisionado, com formulação clássica apresentada por Breiman et al. (1984). A presente exposição baseia-se em Joseph (2022), que descrevem de forma didática seus principais conceitos.

Diferentemente de funções lineares tradicionais, as árvores de decisão possuem maior capacidade de expressão, pois seu funcionamento se apoia no particionamento sucessivo do espaço de atributos em subespaços distintos, atribuindo a cada partição um modelo simples — geralmente a média das observações. Essa característica as torna adequadas em cenários nos quais modelos lineares podem incorrer em subajuste (*underfitting*).

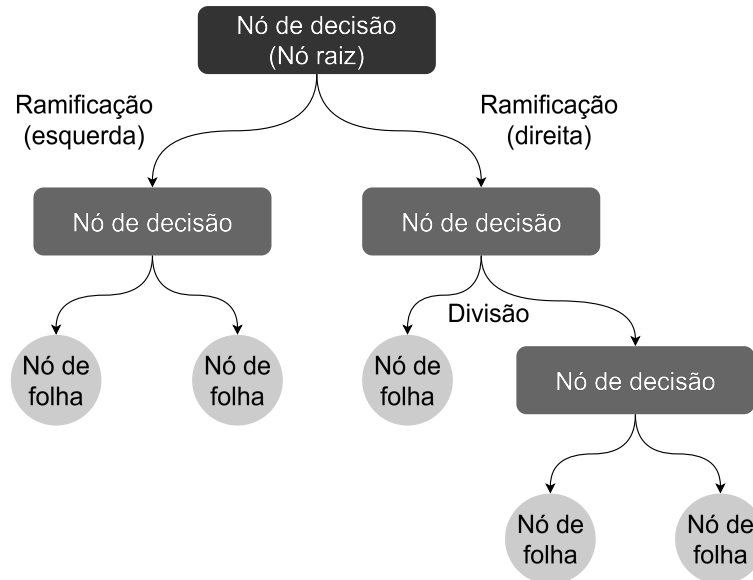
Na Figura 7, observa-se que a estrutura da árvore é composta por dois tipos principais de nós: os de decisão (*decision nodes*) e os nós folha (*leaf nodes*). Os primeiros implementam regras condicionais do tipo “se–então–senão”, avaliando atributos específicos e encaminhando os exemplos para diferentes ramificações em um processo denominado *splitting*. O nó raiz (*root node*) corresponde ao primeiro nó da árvore, enquanto os nós folha representam as partições finais do espaço de atributos, armazenando os valores de predição associados a cada região (JOSEPH, 2022).

Formalmente, a predição \hat{y} gerada por uma árvore de decisão que divide o espaço de atributos em J partições (P_1, P_2, \dots, P_J) pode ser expressa por:

$$\hat{y} = \sum_{m=1}^J c_m \mathbb{I}(\mathbf{x} \in P_m),$$

em que \mathbf{x} é o vetor de entrada, c_m corresponde à resposta constante atribuída à região P_m , e $\mathbb{I}(\cdot)$ é a função indicadora que assume valor 1 se $\mathbf{x} \in P_m$ e 0 caso contrário. Em árvores

Figura 7 – Estrutura de uma árvore de decisão.



Fonte: Adaptado de Joseph (2022).

de regressão baseadas no erro quadrático, c_m é definido como a média das respostas y_i cujas entradas \mathbf{x}_i pertencem à partição P_m (JOSEPH, 2022).

A determinação do conjunto ótimo de partições que minimiza a função de perda global é um problema intratável. Por isso, utiliza-se um algoritmo guloso (*greedy*), que constrói a árvore de forma incremental ao selecionar, em cada nó, a divisão localmente ótima. Em cada etapa, busca-se pelo atributo f e pelo ponto de divisão s que originam duas novas partições, \tilde{P}_1 e \tilde{P}_2 , minimizando a soma dos erros quadráticos:

$$\min_{f,s} \left[\sum_{\mathbf{x}_i \in \tilde{P}_1(f,s)} (y_i - c_1)^2 + \sum_{\mathbf{x}_i \in \tilde{P}_2(f,s)} (y_i - c_2)^2 \right],$$

em que c_1 e c_2 representam as médias das respostas y nas partições \tilde{P}_1 e \tilde{P}_2 , respectivamente (JOSEPH, 2022).

Apesar de sua flexibilidade, as árvores de decisão apresentam duas limitações relevantes: (i) alta propensão ao sobreajuste (*overfitting*), pois podem gerar partições excessivamente específicas ao conjunto de treinamento, resultando em elevada variância; e (ii) ausência de capacidade de extrapolação, já que as previsões permanecem constantes dentro de cada partição e restritas ao intervalo dos dados observados.

Para mitigar o sobreajuste, aplicam-se técnicas de regularização, como restrição da profundidade máxima da árvore, definição de um número mínimo de amostras por divisão

e exigência de redução mínima de impureza para efetuar a partição (JOSEPH, 2022). Esses fundamentos são essenciais para compreender algoritmos mais sofisticados baseados em árvores, como os modelos de *Gradient Boosting* (GBDT), sobre os quais o *LightGBM* está estruturado.

2.9.2 Árvores de decisão com *gradient boosting*

O *Boosting* é uma técnica de *ensemble* que combina múltiplos modelos fracos (*weak learners*) para formar um modelo robusto e preciso. Diferentemente de abordagens como o *bagging*, que constroem modelos em paralelo, o *boosting* adota uma estratégia sequencial: cada novo modelo é treinado para corrigir os erros cometidos pelo anterior (JOSEPH, 2022).

Embora, em teoria, qualquer modelo possa atuar como *weak learner*, na prática as árvores de decisão são a escolha predominante. O algoritmo *Gradient Boosting Decision Trees* (GBDT) constrói o modelo final como uma soma aditiva de várias árvores de decisão:

$$F(\mathbf{x}) = \sum_{k=1}^M T_k(\mathbf{x}),$$

em que $T_k(\mathbf{x})$ representa a k -ésima árvore no conjunto de M árvores. Como a otimização conjunta de todas as árvores é inviável do ponto de vista computacional, o GBDT utiliza uma abordagem sequencial e *stagewise*, adicionando uma árvore por vez com base no gradiente da função de perda (JOSEPH, 2022).

O processo de treinamento pode ser descrito em três etapas principais:

1. **Inicialização do modelo:** define-se uma predição constante $F_0(\mathbf{x})$ que minimiza a função de perda global. Para erro quadrático, esse valor corresponde à média dos alvos; para erro absoluto, à mediana.
2. **Construção iterativa das árvores:** a cada iteração k , calculam-se os *pseudo-resíduos*, definidos como o gradiente negativo da função de perda em relação às predições anteriores. Em seguida, treina-se uma árvore de regressão $T_k(\mathbf{x})$ para aproximar esses resíduos.
3. **Atualização do modelo:** o modelo é atualizado somando a nova árvore ponderada pela taxa de aprendizado η e um fator de escala ρ_k :

$$F_k(\mathbf{x}) = F_{k-1}(\mathbf{x}) + \eta \rho_k T_k(\mathbf{x}).$$

Esse processo se repete até que o número máximo de árvores seja alcançado ou que um critério de parada antecipada seja atendido.

Assim como árvores de decisão individuais, o GBDT é altamente suscetível ao *overfitting*, o que exige estratégias de regularização. Entre elas destacam-se: limitar a

profundidade máxima das árvores, definir o número mínimo de amostras por nó e estabelecer critérios de divisão mais restritivos. Além disso, a complexidade do modelo é fortemente influenciada pelo número de árvores (M), enquanto a taxa de aprendizado (η), também chamada de *shrinkage*, controla a contribuição de cada árvore individual. Taxas menores reduzem o risco de sobreajuste, mas requerem maior número de iterações para convergência.

Outras técnicas relevantes incluem o *early stopping*, que interrompe o treinamento quando não há mais ganho em um conjunto de validação, e a subamostragem de linhas e/ou colunas, que reduz a variância de forma análoga ao *Random Forest*. Implementações modernas ainda incorporam regularização ℓ_1 e ℓ_2 diretamente na função objetivo, penalizando árvores excessivamente complexas e aumentando a estabilidade do modelo.

Apesar de seu frequente elevado desempenho, o GBDT possui limitações importantes. O treinamento é custoso devido à natureza sequencial do algoritmo, que reduz o potencial de paralelização, tornando-o mais lento que alternativas como o *Random Forest*. Além disso, assim como outros métodos baseados em árvores, o GBDT não realiza extrapolação: suas previsões ficam restritas ao intervalo de valores observados no conjunto de treinamento, o que limita sua capacidade de capturar tendências de longo prazo em séries temporais (JOSEPH, 2022).

2.9.3 LightGBM

Conforme descrito por Ke et al. (2017), o *Light Gradient Boosting Machine* (LightGBM) constitui uma implementação do GBDT, desenvolvida com ênfase em eficiência e escalabilidade. Essa abordagem é especialmente adequada para cenários que envolvem grandes volumes de dados e elevado número de atributos. Os autores identificaram que implementações anteriores de GBDT apresentavam limitações de desempenho, pois exigiam a varredura de todas as instâncias de dados para cada atributo a fim de determinar os pontos de divisão ótimos, o que resultava em um custo computacional elevado. Para superar esse gargalo, o LightGBM introduziu duas técnicas de otimização centrais.

A primeira delas é a *Gradient-based One-Side Sampling* (GOSS), ou Amostragem Unilateral Baseada em Gradiente. Essa técnica parte do princípio de que as instâncias de dados não contribuem de forma uniforme para o ganho de informação. Instâncias associadas a gradientes de maior magnitude — ou seja, aquelas ainda pouco ajustadas pelo modelo — são consideradas mais relevantes para o processo de convergência e, por isso, são sempre preservadas. Já as instâncias com gradientes menores são submetidas a uma amostragem aleatória. Para evitar vieses na distribuição resultante, aplica-se um fator de correção às instâncias com gradientes baixos durante o cálculo do ganho de informação. Dessa forma, o GOSS permite estimativas precisas do ganho utilizando apenas um subconjunto reduzido de dados, o que acelera significativamente o treinamento.

A segunda técnica é o *Exclusive Feature Bundling* (EFB), ou Agrupamento Exclusivo de Atributos. Essa otimização explora a esparsidade comumente observada em espaços de alta dimensionalidade, nos quais muitos atributos são mutuamente exclusivos e raramente assumem valores não nulos simultaneamente. O EFB agrupa tais atributos em um único conjunto, denominado *feature bundle*. Embora a identificação do agrupamento ótimo seja um problema classificado como NP-difícil, Ke et al. (2017) propuseram um algoritmo guloso eficiente como solução aproximada. Essa técnica reduz drasticamente a dimensionalidade efetiva, modificando a complexidade da construção de histogramas de $\mathcal{O}(Nm)$ para $\mathcal{O}(Nn_b)$, em que N é o número de instâncias, m o número de atributos e n_b o número de *bundles* (tipicamente $n_b \ll m$).

2.10 Redes neurais e aprendizado profundo

As redes neurais podem ser formalmente descritas como uma função $f_\theta : \mathcal{X} \mapsto \mathcal{Y}$, em que \mathcal{X} corresponde ao espaço de entrada e \mathcal{Y} ao espaço da variável de saída (GOODFELLOW; BENGIO; COURVILLE, 2016). No contexto da previsão de séries temporais, a entrada costuma ser um histórico de w observações, isto é, $\mathbf{x} = \{y_t, y_{t-1}, \dots, y_{t-w+1}\}$, enquanto a saída é definida pelo horizonte de previsão de H passos à frente, $\mathbf{y}_{t+1:t+H} = (y_{t+1}, \dots, y_{t+H})$ (HYNDMAN, 2020).

Além das observações passadas da série, a entrada pode ser enriquecida por variáveis exógenas, representadas por $\mathbf{x}_t = (x_t^{(1)}, x_t^{(2)}, \dots, x_t^{(d_x)})$. Assim, a formulação pode ser expandida para $\mathcal{X} = \{y_t, \dots, y_{t-w+1}, \mathbf{x}_t, \dots, \mathbf{x}_{t-w+1}\}$, permitindo à rede capturar dependências tanto no histórico da série quanto em covariáveis que influenciam sua dinâmica (HAMILTON, 2020; GOODFELLOW; BENGIO; COURVILLE, 2016).

O treinamento consiste em minimizar uma função de perda \mathcal{L} que quantifica a discrepância entre as previsões $\hat{\mathbf{y}}_\theta$ e os valores observados \mathbf{y} . Um otimizador amplamente usado é o gradiente descendente (GD), cuja regra de atualização é:

$$\theta_{i+1} \leftarrow \theta_i - \alpha \frac{\partial \mathcal{L}}{\partial \theta}, \quad (8)$$

em que α é a taxa de aprendizado. A derivada $\partial \mathcal{L} / \partial \theta$ é obtida por retropropagação (regra da cadeia), ajustando eficazmente os pesos da rede (RUMELHART; HINTON; WILLIAMS, 1986; GOODFELLOW; BENGIO; COURVILLE, 2016).

2.10.1 *Perceptron* de múltiplas camadas

As redes neurais artificiais são modelos computacionais formados por unidades elementares de processamento, denominadas neurônios, organizadas em camadas interconectadas (HYNDMAN; ATHANASOPOULOS, 2021). A primeira camada corresponde ao vetor de entrada \mathbf{x} , enquanto a última camada representa as saídas previstas \mathbf{y} . Entre

ambas, podem existir camadas intermediárias compostas por neurônios ocultos (*hidden neurons*), responsáveis por ampliar a capacidade de modelagem da rede. Na ausência dessas camadas, a arquitetura se reduz a uma *Feedforward Neural Network* (FFN) linear, funcionalmente equivalente à regressão linear, em que os coeficientes das variáveis correspondem aos pesos (*weights*). Nesse caso, as previsões resultam de combinações lineares das entradas, com os pesos ajustados por algoritmos de otimização, como a descida do gradiente, que minimizam funções de perda, por exemplo, MAE ou Erro Médio Quadrático (*Mean Squared Error*) (MSE).

A introdução de camadas intermediárias é o que confere não linearidade às redes neurais. Esse conceito está ilustrado na Figura 8. Nessa estrutura, cada camada recebe como entrada as saídas da camada anterior. Em cada neurônio, essas entradas são combinadas linearmente, conforme a Equação 9, e transformadas por uma função de ativação não linear — por exemplo, a sigmoide σ (Equação 10), a ReLU (Equação 11) ou a tangente hiperbólica \tanh (Equação 12). A entrada \mathbf{x} pode ainda incluir covariáveis externas ou variáveis exógenas (HAMILTON, 2020).

$$z_j = b_j + \sum_{i=1}^{d_x} w_{i,j} x_i, \quad (9)$$

$$\sigma(z) = \frac{1}{1 + e^{-z}}, \quad (10)$$

$$\text{ReLU}(z) = \max(0, z), \quad (11)$$

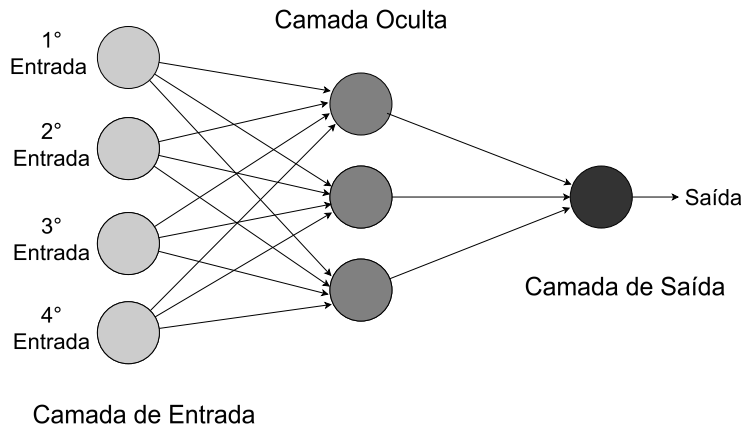
$$\tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}. \quad (12)$$

Os parâmetros $w_{i,j}$ e b_j são inicializados de forma aleatória e ajustados iterativamente durante o treinamento, por meio do algoritmo de retropropagação (*Backpropagation*). Esse procedimento atualiza os pesos de modo a minimizar a função de perda, sendo conceitualmente análogo ao algoritmo de retropropagação ao longo do tempo (BPTT) empregado em RNNs e LSTMs, mas restrito ao fluxo unidirecional de informações em arquiteturas do tipo *feedforward*.

2.10.2 Redes neurais recorrentes

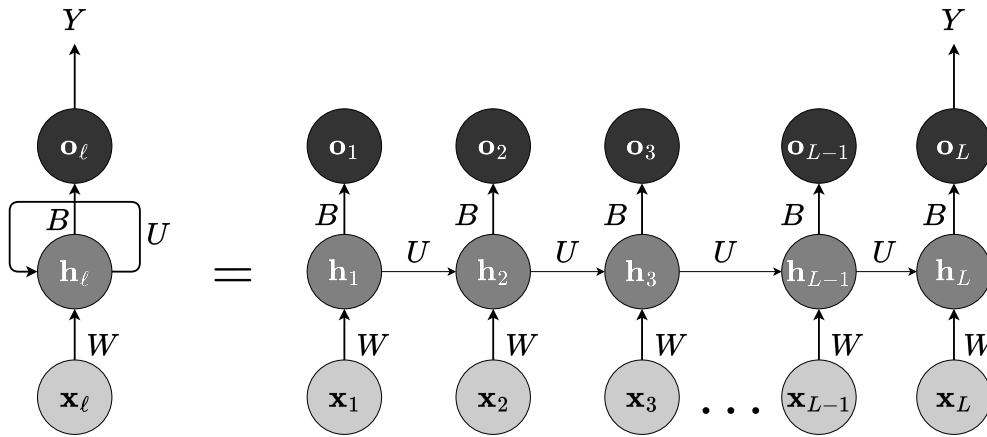
As Redes Neurais Recorrentes (*Recurrent Neural Networks* – RNN) são modelos de aprendizado profundo projetados para dados sequenciais. A entrada \mathbf{X} é representada como uma sequência de vetores $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{L_{\text{seq}}}\}$, podendo cada \mathbf{x}_ℓ incluir a série principal e covariáveis externas (HAMILTON, 2020; JAMES et al., 2023). Propostas por Rumelhart, Hinton e Williams (1986), as RNNs superam limitações das FFNs ao capturar

Figura 8 – Arquitetura de um *Perceptron* de Múltiplas Camadas (MLP) com quatro entradas e uma camada oculta contendo três neurônios.



Fonte: Adaptado de Hyndman e Athanasopoulos (2021).

Figura 9 – Arquitetura de uma rede neural recorrente simples, com destaque para a propagação das entradas \mathbf{x}_ℓ , dos estados ocultos \mathbf{h}_ℓ e das saídas \mathbf{o}_ℓ .



Fonte: Adaptado e modificado de James et al. (2023).

dependências temporais em sequências. A saída \mathbf{Y} pode ser uma sequência ou um valor escalar, dependendo da tarefa (JOSEPH, 2022; JAMES et al., 2023).

O estado oculto \mathbf{h}_ℓ é calculado a partir da entrada atual \mathbf{x}_ℓ e do estado anterior $\mathbf{h}_{\ell-1}$, inicializado em \mathbf{h}_0 . Dependendo da arquitetura, é calculada uma saída intermediária \mathbf{o}_ℓ a cada passo, ou apenas a saída final $\mathbf{o}_{L_{\text{seq}}}$:

$$h_{\ell k} = g \left(b_k + \sum_{j=1}^{d_x} w_{kj} x_{\ell j} + \sum_{s=1}^K u_{ks} h_{\ell-1, s} \right), \quad (13)$$

$$o_\ell = \beta_0 + \sum_{k=1}^K \beta_k h_{\ell k}, \quad (14)$$

em que $g(\cdot)$ corresponde a uma função de ativação não linear, como a tangente hiperbólica \tanh (Equação 12), a sigmoide σ (Equação 10) ou a ReLU (Equação 11). Na Figura 9, W , U e b representam, respectivamente, as matrizes de pesos das conexões entre a entrada e o estado oculto, entre estados ocultos consecutivos e o vetor de vies, sendo diretamente relacionados aos parâmetros w_{kj} , u_{ks} e b_k utilizados nas Equação 13 e Equação 14, sendo K o número de neurônios na camada oculta.

Considerando $o_{iL_{\text{seq}}}$ como a saída final da RNN para a i -ésima sequência, a função de perda baseada no erro quadrático total é definida como:

$$\sum_{i=1}^{N_{\text{seq}}} (y_i - o_{iL_{\text{seq}}})^2. \quad (15)$$

O treinamento das redes neurais recorrentes é realizado por meio do algoritmo de retropropagação ao longo do tempo (*Backpropagation Through Time*) (BPTT). Uma extensão desse modelo, denominada Rede Neural Recorrente Bidirecional (*Bidirectional Recurrent Neural Network*) (BRNN), proposta por Schuster e Paliwal (1997), processa a sequência de entrada em ambas as direções, concatenando os estados ocultos correspondentes, o que permite gerar previsões mais informadas (JOSEPH, 2022).

Apesar de sua eficácia na modelagem de sequências, as RNNs apresentam uma limitação significativa. O uso do algoritmo de retropropagação ao longo do tempo (BPTT) implica que o número de unidades pelas quais o gradiente precisa ser propagado aumenta drasticamente conforme o comprimento da sequência utilizada no treinamento. Ao retropropagar através de um grafo computacional muito longo, surgem os problemas de dissipação ou explosão do gradiente. No primeiro caso, os valores do gradiente tendem a zero, o que interrompe a capacidade de aprendizado da rede. No segundo, os valores do gradiente aumentam excessivamente, resultando em instabilidade no processo de treinamento (JOSEPH, 2022).

2.10.3 Arquitetura de Memória de Longo e Curto Prazo

Apesar de eficazes, as RNNs apresentam dificuldades em capturar dependências de longo prazo devido ao desaparecimento ou explosão de gradientes. Para superar essa limitação, Hochreiter e Schmidhuber (1997) propuseram a arquitetura de Memória de Longo e Curto Prazo (*Long Short-Term Memory* – LSTM), que introduziu autociclos para manter o fluxo do gradiente estável. Posteriormente, Gers, Schmidhuber e Cummins (2000) aprimoraram a proposta, tornando o peso desses autociclos adaptativo ao contexto.

Cada célula LSTM é composta pelo *estado da célula* (C_τ), que armazena informações de longo prazo, e pelo *estado oculto* (\mathbf{h}_τ), que contém as informações usadas nas previsões.

O fluxo de dados é controlado por três portões: o *portão de esquecimento* (F_τ), o *portão de entrada* (I_τ) e o *portão de saída* (O_τ). Esses portões utilizam combinações lineares da entrada atual (\mathbf{x}_τ) e do estado oculto anterior ($\mathbf{h}_{\tau-1}$), seguidas pela função sigmoide $\sigma(\cdot)$, que produz valores entre 0 e 1 (GOODFELLOW; BENGIO; COURVILLE, 2016).

As equações que descrevem o funcionamento da célula são:

$$F_\tau = \sigma(W_{xf}\mathbf{x}_\tau + W_{hf}\mathbf{h}_{\tau-1} + b_f) \quad (\text{Portão de esquecimento})$$

$$I_\tau = \sigma(W_{xi}\mathbf{x}_\tau + W_{hi}\mathbf{h}_{\tau-1} + b_i) \quad (\text{Portão de entrada})$$

$$O_\tau = \sigma(W_{xo}\mathbf{x}_\tau + W_{ho}\mathbf{h}_{\tau-1} + b_o) \quad (\text{Portão de saída})$$

$$\tilde{C}_\tau = \tanh(W_{xc}\mathbf{x}_\tau + W_{hc}\mathbf{h}_{\tau-1} + b_c) \quad (\text{Candidato a novo estado})$$

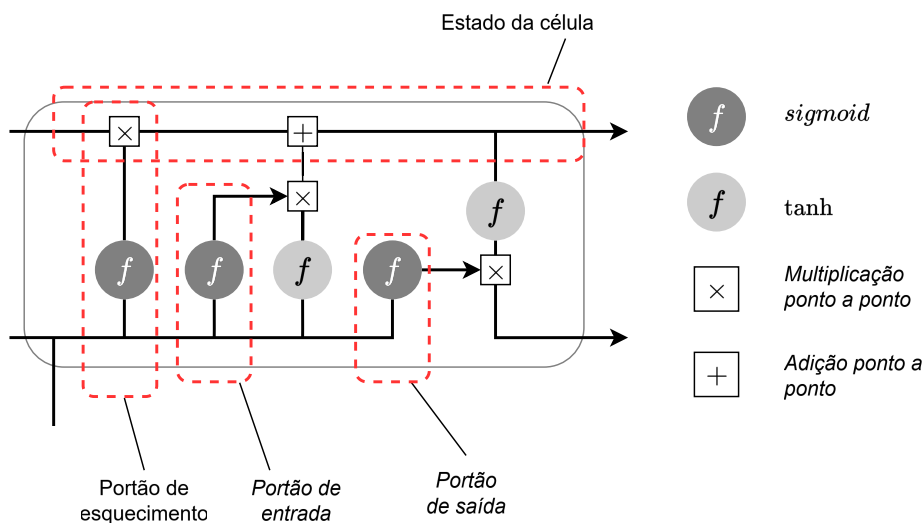
$$C_\tau = F_\tau \odot C_{\tau-1} + I_\tau \odot \tilde{C}_\tau \quad (\text{Atualização da célula})$$

$$\mathbf{h}_\tau = O_\tau \odot \tanh(C_\tau) \quad (\text{Novo estado oculto})$$

em que: W_{x*}, W_{h*} representam as matrizes de pesos treináveis (entrada/estado); b_* são vetores de viés; $\sigma(\cdot)$ é a função sigmoide; $\tanh(\cdot)$ a função tangente hiperbólica; e \odot o produto de Hadamard (ponto a ponto) (HOCHREITER; SCHMIDHUBER, 1997; GOODFELLOW; BENGIO; COURVILLE, 2016).

A LSTM amplia a capacidade das RNNs ao incorporar memória de longo prazo e atualização seletiva, o que a torna especialmente eficaz para previsão de séries temporais. A Figura 10 apresenta o diagrama de uma célula LSTM.

Figura 10 – Diagrama de uma célula LSTM, destacando os portões de esquecimento, entrada e saída, o estado da célula e as operações de produto de Hadamard (ponto a ponto).



Fonte: Adaptado e modificado de Joseph (2022).

Além disso, arquiteturas recorrentes inspiradas no uso de LSTMs em modelos do tipo *Sequence-to-Sequence* (Seq2Seq) (SUTSKEVER; VINYALS; LE, 2014) deram origem a variações aplicadas em previsão de séries temporais. Um exemplo é o Block RNN, no qual a entrada é dividida em blocos de tamanho fixo, processados por um codificador recorrente, enquanto camadas totalmente conectadas subsequentes produzem saídas também em blocos. Essa abordagem é particularmente útil em cenários em que se desejam previsões em janelas contínuas de horizonte fixo, oferecendo uma alternativa eficiente para a modelagem de dependências temporais de médio prazo (JOSEPH, 2022).

2.10.4 A arquitetura N-BEATS

A arquitetura *Neural Basis Expansion Analysis for Interpretable Time Series Forecasting* (N-BEATS) é simples tanto no nível conceitual quanto de implementação. Em vez de abordar a previsão como um problema sequencial, ela trata como um problema de regressão multivariada não linear (ORESHKIN et al., 2019; ORESHKIN et al., 2021), conforme apresentado na Figura 11. Na imagem, há três colunas de blocos retangulares, cada uma representando uma visão detalhada da outra, que serão gradualmente utilizadas na explicação da arquitetura.

Na entrada do modelo, a janela de observação (w), equivalente a um múltiplo do horizonte de previsão nH , é inserida na Pilha 1. O comprimento de w varia de $2H$ a $7H$. Para cada uma dessas seis variações, é treinado um modelo que posteriormente é combinado para gerar previsões. Essa abordagem, conhecida como *ensembling*, foi amplamente utilizada pelos principais participantes da competição M4 (ORESHKIN et al., 2019; MAKRIDAKIS; SPILIOTIS; ASSIMAKOPOULOS, 2020).

Na coluna mais à esquerda da Figura 11, cada Bloco ℓ é constituído por uma pilha de quatro camadas totalmente conectadas (*Fully Connected layer*) (FC) com largura L , responsáveis por transformar uma entrada em uma representação oculta \mathbf{h}_ℓ . Este conjunto de camadas FC antecede o que Oreshkin et al. (2019) chamam de coeficientes de expansão *backcast*, θ_ℓ^b , e *forecast*, θ_ℓ^f . Esses coeficientes são essenciais para antecipar, respectivamente, os dados passados (*backcast*), revelando como o bloco teria previsto w , e qual sua previsão futura (*forecast*) para H .

As camadas FC empregam funções de ativação ReLU; por exemplo, a Camada FC1 pode ser representada por

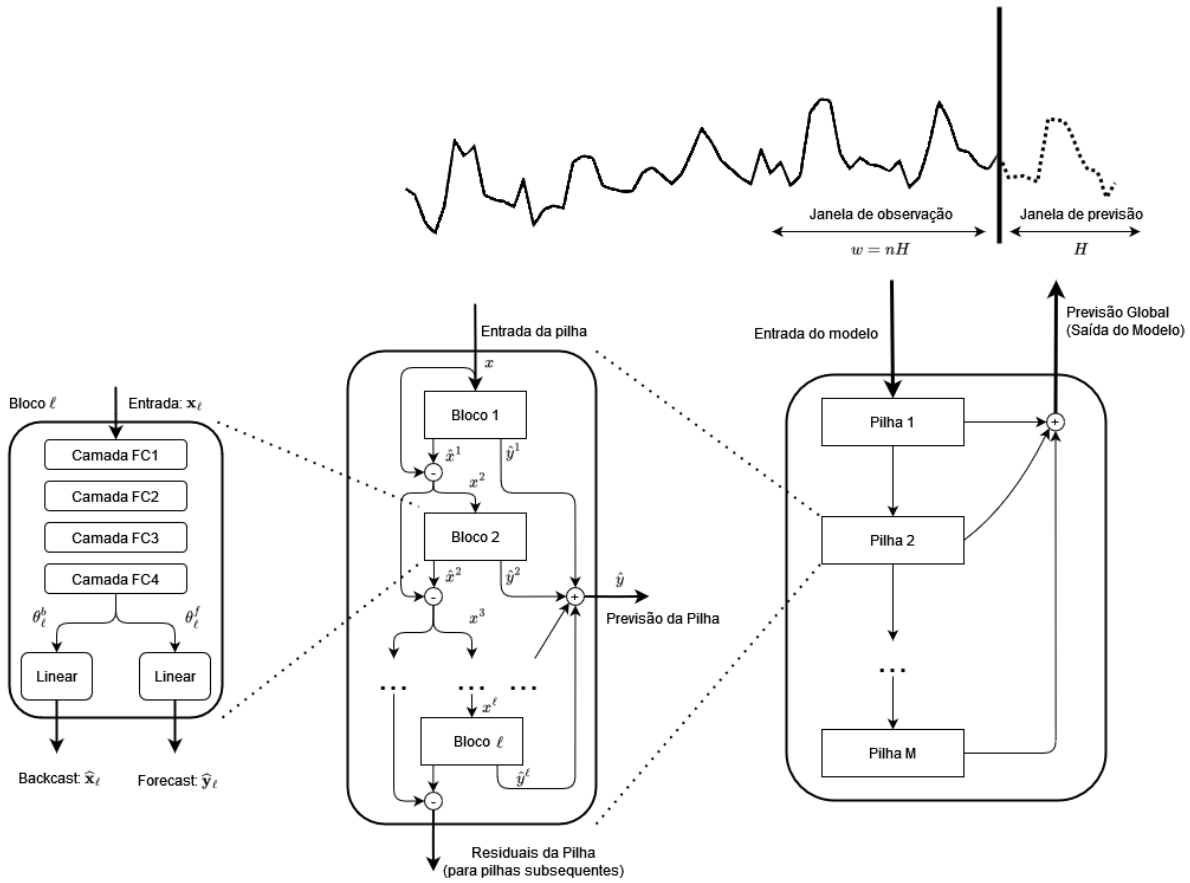
$$\mathbf{h}_{\ell,1} = \text{FC}_{\ell,1}(\mathbf{x}_\ell) \quad \text{ou} \quad \mathbf{h}_{\ell,1} = \text{ReLU}(\mathbf{W}_{\ell,1}\mathbf{x}_\ell + \mathbf{b}_{\ell,1}).$$

Já os preditores θ_ℓ^b e θ_ℓ^f são gerados por meio de uma função linear que utiliza a saída da Camada FC4 como entrada,

$$\theta_\ell^b = \text{Linear}_\ell^b(\mathbf{h}_{\ell,4}) \quad \text{e} \quad \theta_\ell^f = \text{Linear}_\ell^f(\mathbf{h}_{\ell,4}).$$

A função Linear é simplesmente uma camada de projeção linear sem viés, como $\theta_\ell^f = \mathbf{W}_\ell^f \mathbf{h}_{\ell,4}$.

Figura 11 – Arquitetura N-BEATS



Fonte: Adaptado de Oreshkin et al. (2019).

Após a obtenção dos coeficientes preditores θ_ℓ^b e θ_ℓ^f , é realizada a última etapa do bloco ℓ , onde são gerados os valores para *backcast*, \hat{x} , e o *forecast*, \hat{y} . A fase final do bloco emprega esses coeficientes de expansão e os direciona para a saída por meio de um conjunto de camadas de base g_ℓ^b e g_ℓ^f , transformando-os nos resultados desejados:

$$\hat{x} = g_\ell^b(\theta_\ell^b) \quad \text{e} \quad \hat{y} = g_\ell^f(\theta_\ell^f).$$

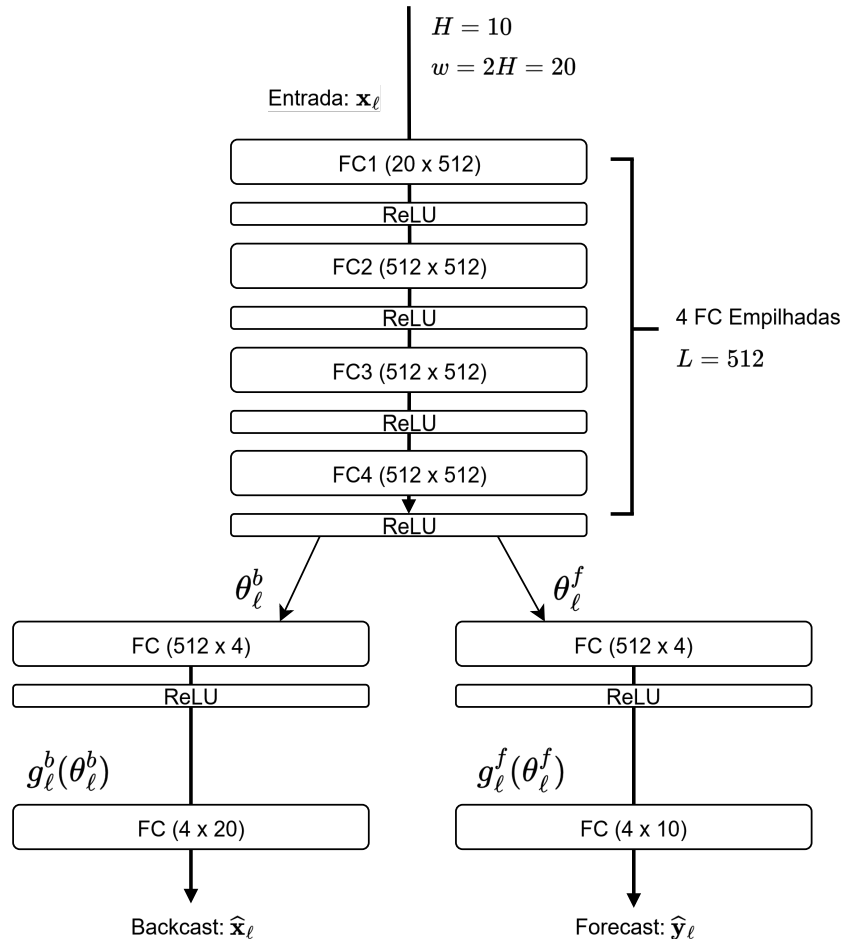
Como uma função pode ser concebida como um ponto em um espaço vetorial, e cada função dentro desse espaço pode ser expressa como uma combinação linear das funções de base, g_ℓ^b e g_ℓ^f fazem uso desse conceito. Portanto, ao aprender os coeficientes de expansão, há, fundamentalmente, uma busca pela função ideal que pode representar a saída, sendo essa busca restringida pelas funções de base (JOSEPH, 2022).

O N-BEATS opera em dois modos distintos para as funções g_ℓ^b e g_ℓ^f : um genérico e outro interpretável. De acordo com Oreshkin et al. (2019), em ambas as abordagens, o modelo demonstrou superioridade em relação ao desempenho máximo alcançado na competição M4. O modo genérico opera sem restrições de função de base, permitindo

que a função seja totalmente aprendida pelo modelo por meio de uma projeção linear dos coeficientes de base.

Na Figura 12, é representado um bloco básico genérico com parâmetros: $H = 10$, uma janela w com $n = 2$ e as quatro FC com largura $L = 512$.

Figura 12 – Bloco básico genérico.



Fonte: Adaptado de Oreshkin et al. (2019).

Essa abordagem carece de interpretabilidade humana, já que não se tem conhecimento de como as diversas funções são aprendidas e do significado de cada pilha. No modo interpretável, são utilizadas funções de base fixas que restringem o espaço da função, com o intuito de capturar tendências e sazonalidades. A inclusão desses blocos torna a previsão final mais interpretável, proporcionando uma decomposição mais detalhada do fenômeno em análise. Para isso, Oreshkin et al. (2019) empregaram uma função polinomial de grau pequeno (r) para representar a tendência. Quanto à função de base para a sazonalidade, optaram por uma base de Fourier, pois conduz a saída da previsão a ser expressa como funções de base sinusoidal.

2.11 Considerações finais

Este capítulo reuniu os fundamentos teóricos para a predição de *throughput* em redes móveis. Discutiu-se que a vazão permanece um recurso escasso e altamente variável devido à heterogeneidade tecnológica, à mobilidade e ao crescimento do tráfego. Nesse contexto, foram revisadas as principais métricas de qualidade de canal (RSRP, RSRQ, RSSI, SNR e CQI) e seus papéis na seleção de célula, no HO e na adaptação do enlace, ressaltando-se que nenhuma métrica, isoladamente, captura adequadamente o desempenho; combinações que integrem intensidade de sinal, qualidade e interferência/ruído oferecem maior poder informativo para modelos preditivos e para a garantia de QoS.

Apresentaram-se, ainda, os elementos formais do problema de previsão em séries temporais (janela de entrada e horizonte), a decomposição em componentes estruturais (tendência, sazonalidade, ciclos e resíduos) e a importância da estacionariedade, com destaque para testes de raiz unitária e transformações como a diferenciação. Na avaliação de desempenho, enfatizaram-se MAE e RMSE e a necessidade de comparação com previsores ingênuos (*baselines*) para aferir ganhos substanciais.

Por fim, foram sistematizados os principais métodos de previsão em três grupos: *baselines*, modelos estatísticos (como ARIMA e regressão com regularização) e abordagens de aprendizado de máquina, que vão de modelos baseados em árvores e *gradient boosting* (e.g., LightGBM) a arquiteturas profundas para sequências (RNN/LSTM e N-BEATS). Discutiram-se também estratégias de validação compatíveis com dados temporais (origem fixa e origem rolante). Essas diretrizes enquadram as denominações adotadas neste capítulo para **modelos univariados** (sem covariáveis) e **modelos covariados** (com preditores externos), fornecendo o arcabouço teórico necessário às análises aqui descritas.

Capítulo 3

Trabalhos relacionados

Este capítulo detalha o estado da arte na predição de *throughput* em redes móveis, com ênfase nos modelos adotados, nas estratégias de pré-processamento e nas métricas de avaliação. São apresentados resultados comparativos e limitações metodológicas dos estudos analisados, destacando-se sua relevância como base de referência para as decisões de modelagem, validação e escopo desta dissertação.

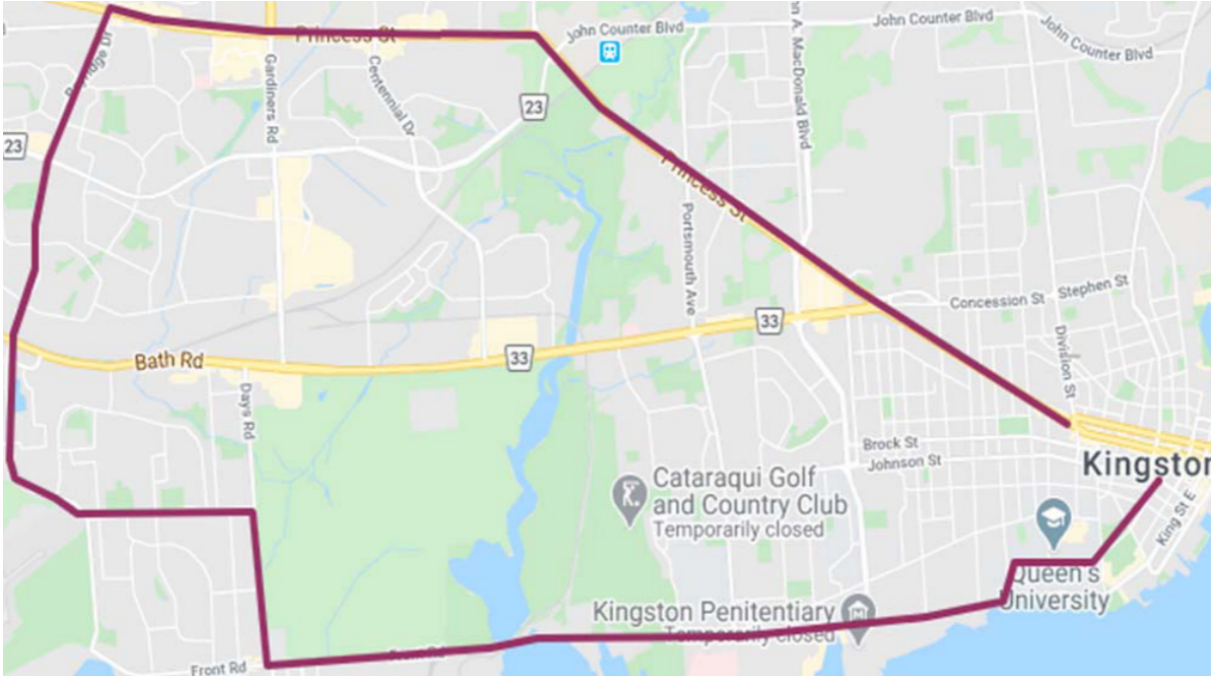
3.1 Avaliação de modelos clássicos e séries temporais na previsão de *throughput* em redes 4G

O estudo de Elsherbiny et al. (2020) teve como principal objetivo o desenvolvimento e a avaliação de diferentes métodos de modelagem e previsão de *throughput* em redes móveis, a partir de medições coletadas em uma rede 4G LTE em operação. Para isso, os autores realizaram uma análise comparativa entre algoritmos clássicos de aprendizado de máquina e técnicas de previsão de séries temporais, aplicadas especificamente à estimativa do *throughput* de *downlink*, com vistas a apoiar estratégias de alocação proativa de recursos.

O experimento foi conduzido ao longo do trajeto de um ônibus de transporte público em Kingston, Ontário, ilustrado na Figura 13. O conjunto de dados utilizado é publicamente disponível e composto por 30 viagens repetidas, cada uma com duração aproximada de uma hora, realizadas em três diferentes horários de dias úteis (9h, 12h e 18h). Esse conjunto inclui parâmetros de rede como RSRP, RSRQ, RSSI e SNR, além de medições de *throughput* de *downlink* e *uplink*, bem como informações contextuais, como coordenadas de GPS e velocidade do veículo, registradas com granularidade de um segundo. Apresenta características semelhantes ao descrito em Raca et al. (2020), porém

sem contemplar medições da tecnologia 5G.

Figura 13 – Rota do ônibus de transporte público utilizada para a coleta do conjunto de dados da rede 4G LTE em Kingston, Ontário.



Fonte: Adaptado de Elsherbiny et al. (2020).

3.1.1 Pré-processamento dos dados

O conjunto passou por quatro etapas principais de pré-processamento. Na primeira, foram detectados e removidos *outliers* por meio do método *z-score*, que mede a distância de um ponto em relação à média em unidades de desvio padrão, conforme Equação 16:

$$z = \frac{(x - \mu)}{\sigma} \quad (16)$$

em que μ representa a média e σ o desvio padrão do conjunto de dados.

Na segunda etapa, tratou-se a presença de valores ausentes mediante imputação, substituindo-os pela média ou mediana de cada atributo, de acordo com a Equação 17:

$$x'_i = \begin{cases} x_i, & \text{se } x_i \neq \text{NaN} \\ \bar{x} \text{ ou } \tilde{x}, & \text{se } x_i = \text{NaN} \end{cases} \quad (17)$$

em que \bar{x} representa a média e \tilde{x} a mediana dos valores observados.

A terceira etapa consistiu no escalonamento das características, essencial em modelos que dependem de medidas de distância. Para isso, aplicou-se novamente a normalização pelo *z-score*, já definida na Equação 16, de modo a transformar os dados para média zero e desvio padrão unitário.

Por fim, foi realizado o agrupamento de dados (discretização), com agregação em janelas de cinco segundos e suavização pela média de cada intervalo, a fim de reduzir o efeito do ruído nos dados. A validação adotou divisão de 70% para treinamento e 30% para teste/validação.

3.1.2 Modelos e métricas de avaliação

No estudo, a identificação dos atributos mais relevantes foi realizada com base na análise da importância das variáveis estimada por um modelo de *Random Forest*, o que indicou RSRP, RSSI, SNR, RSRQ, além de latitude e longitude como os principais preditores. Esses atributos foram então utilizados no treinamento dos modelos. Consideraram-se duas estratégias: (i) regressão supervisionada — SVR, KNN, *Ridge Regression* e *Random Forest* — alimentada por parâmetros de rede e contexto; e (ii) previsão univariada — ARIMA e LSTM — baseada apenas no histórico do próprio *throughput*.

O desempenho foi avaliado por R^2 e RMSE (Equação 5). Embora R^2 calcule a variância explicada, em séries com comportamento de passeio aleatório — isto é, nível não estacionário — ele pode produzir valores artificialmente altos e deve ser interpretado com cautela (HEWAMALAGE; ACKERMANN; BERGMEIR, 2023). Já o RMSE possui leitura direta nas unidades do alvo; no conjunto analisado no estudo, o *throughput* varia entre 0 e 80.000 Kbps, o que orienta a magnitude aceitável do erro absoluto.

3.1.3 Resultados

Os resultados reportados no estudo estão sintetizados na Tabela 2. Entre os modelos de regressão, o *Random Forest* apresentou o melhor desempenho ($R^2 = 0,78$; RMSE = 8.200), atribuído à sua capacidade de generalização e redução da correlação entre árvores. O *Ridge Regression* também apresentou desempenho competitivo ($R^2 = 0,71$), enquanto SVR e KNN obtiveram resultados insatisfatórios, com baixo poder explicativo e maiores erros. Já entre os modelos de séries temporais, o ARIMA superou o LSTM, alcançando $R^2 = 0,62$ e RMSE de 10.400, contra 0,59 e 10.800, respectivamente. O fraco desempenho do LSTM foi associado ao tamanho reduzido do conjunto de dados, considerado insuficiente para explorar a complexidade de suas unidades internas.

3.1.4 Análise crítica

De modo geral, Elsherbiny et al. (2020) apontam que a previsão de *throughput* pode atuar como ferramenta proativa para otimização de recursos em redes, possibilitando desde alocação antecipada até detecção de anomalias. Os resultados evidenciaram que modelos clássicos, como *Random Forest* e ARIMA, apresentaram desempenho superior a arquiteturas mais complexas, como o LSTM, o que foi atribuído às limitações do conjunto

Tabela 2 – Resultados obtidos com os modelos de regressão e séries temporais

Modelo	RMSE	R ²
Regressão		
SVR	14000 ± 300	0.36 ± 0.03
KNN	13700 ± 300	0.38 ± 0.02
<i>Ridge Regression</i>	9300 ± 200	0.71 ± 0.02
<i>Random Forest</i>	8200 ± 200	0.78 ± 0.01
Séries Temporais		
ARIMA	10400	0.62
LSTM	10800	0.59

Fonte: Adaptado de Elsherbiny et al. (2020)

de dados utilizado. Os autores ainda ressaltam que a incorporação de informações adicionais provenientes de operadoras e a utilização de maior granularidade temporal poderiam contribuir para o aprimoramento das previsões.

Entretanto, observa-se que o estudo não incluiu modelos de referência mais simples, como abordagens *naive*, que poderiam ter estabelecido uma linha de base para comparação. Além disso, tanto os modelos de aprendizado de máquina quanto os de séries temporais foram aplicados sem a incorporação de covariáveis, restringindo o potencial de captura de dependências externas. Outro ponto relevante é que os dados aparentam ter sido concatenados em uma única série, com a validação realizada apenas sobre os 30% finais, o que pode limitar a avaliação da capacidade de generalização dos modelos.

3.2 Um *framework* para previsão de *throughput* em redes 5G

O trabalho de Narayanan et al. (2020) investigou se o *throughput* do 5G mmWave é previsível, dada sua elevada variabilidade causada por fatores ambientais e de mobilidade, e se modelos de aprendizado de máquina podem realizar tal previsão. Para isso, os autores propuseram o *framework* Lumos5G, que integra diferentes grupos de atributos — localização, mobilidade, parâmetros da torre e conexão de rádio — para estimar o desempenho do enlace¹. O estudo mostra que abordagens tradicionais baseadas apenas na localização, eficazes em redes 4G, não conseguem prever com precisão o *throughput* em 5G mmWave, cuja performance é fortemente influenciada por múltiplos fatores e suas interações, evidenciando uma complexidade bem maior em relação às gerações anteriores (3G e 4G). Os autores identificaram que o *throughput* em redes 5G mmWave pode variar de forma drástica, oscilando de valores próximos a 2 Gbps até quase zero, em decorrência

¹ O termo *enlace* refere-se ao canal de comunicação entre o equipamento do usuário (UE) e a estação rádio base, cujo desempenho é avaliado por métricas como *throughput*, potência e qualidade do sinal, latência e taxa de erros.

da mobilidade do usuário e de obstruções ambientais, que frequentemente desencadeiam processos de *handover* (NARAYANAN et al., 2020).

As Tabela 3 e Tabela 4 apresentam os principais campos utilizados no *framework* Lumos5G, abrangendo tanto valores brutos obtidos diretamente das APIs do Android quanto informações derivadas de processamento adicional. Entre os atributos coletados destacam-se parâmetros de contexto, como localização geográfica, atividade do usuário, velocidade e direção de deslocamento, que permitem caracterizar a mobilidade. Complementarmente, foram incluídos indicadores relacionados à rede, como *throughput*, tipo de tecnologia de acesso, identificação da célula, intensidade de sinal e ocorrência de *handover* horizontal ou vertical².

Tabela 3 – Campos brutos ou obtidos a partir das APIs do Android no conjunto de dados Lumos5G. Com exceção do campo *timestamp*, os demais possuem acurácia estimada pelo Android e podem não ser totalmente precisos.

Campo	Descrição
timestamp	Registra a data e o horário em intervalos de um segundo.
latitude, longitude	Coordenadas geográficas detalhadas do usuário, obtidas pela API do Android, juntamente com a acurácia estimada da localização.
detected Activity	Informa se o usuário está andando, parado, dirigindo, entre outras atividades, por meio da <i>Activity Recognition API</i> do Google.
moving speed	Velocidade de deslocamento do usuário, reportada pela API do Android.
compass direction	Direção horizontal de deslocamento do usuário em relação ao Polo Norte (azimute) e respectiva acurácia.

Fonte: Adaptado de Narayanan et al. (2020).

Além desses parâmetros, os autores também incluíram medidas geométricas, como a distância entre o usuário e o painel, bem como os ângulos posicional (θ_p) e de mobilidade (θ_m), obtidos por meio de cálculos adicionais, já que não estão disponíveis diretamente nas APIs do Android. Ressalta-se, nesse contexto, que o acesso direto à camada física dos dispositivos é restrito pelo sistema Android³, o que restringe a coleta de métricas mais detalhadas de rádio e reforça a necessidade de estratégias complementares de instrumentação.

² O *handover* horizontal ocorre quando há transferência entre diferentes células da mesma tecnologia de acesso (por exemplo, de um painel 5G para outro), enquanto o *handover* vertical refere-se à transferência entre diferentes tecnologias de acesso (por exemplo, de 4G para 5G) (ANDREWS et al., 2014).

³ O sistema operacional Android limita a exposição de informações detalhadas da camada física (*PHY*), como medições avançadas de potência, qualidade do canal ou indicadores internos do modem, permitindo apenas um subconjunto de métricas acessíveis por APIs de alto nível. Essa limitação é uma prática de segurança e padronização em dispositivos comerciais, mas representa um desafio para estudos experimentais (NARAYANAN et al., 2020).

Tabela 4 – Campos derivados ou obtidos após processamento no conjunto de dados Lumos5G.

Campo	Descrição
throughput	<i>Throughput</i> de <i>downlink</i> , mensurado por meio da ferramenta iPerf (versão 3.7).
radio type	Tipo de tecnologia de acesso utilizada (4G ou 5G), identificada a partir do objeto <i>ServiceState</i> .
cell ID	Identificador da célula (mCid) à qual o usuário está conectado, extraído do objeto <i>ServiceState</i> .
signal strength	Intensidade de sinal de LTE (RSRP, RSRQ, RSSI) e 5G (SS-RSRP, SS-RSRQ, SS-RSSI), extraída do objeto <i>SignalStrength</i> .
horizontal handover	Processo de transferência entre diferentes painéis 5G (células adjacentes).
vertical handover	Processo de transferência entre diferentes tecnologias de acesso (por exemplo, de 4G para 5G).
UE-Panel Dist.	Distância entre o equipamento do usuário e o painel ao qual está conectado.
Positional Angle (θ_p)	Ângulo entre a posição do usuário e a linha normal à frente do painel 5G.
Mobility Angle (θ_m)	Ângulo entre a trajetória do usuário e a linha normal à frente do painel 5G.

Fonte: Adaptado de Narayanan et al. (2020).

3.2.1 Coleta de dados em campo

Segundo Narayanan et al. (2020), o conjunto de dados utilizado nos experimentos não é um corpus público pré-existente, mas foi inteiramente construído pelos autores a partir de uma campanha de coleta conduzida ao longo de seis meses com a ferramenta de monitoramento *Lumos5G*. Para tal, foram empregados quatro smartphones, utilizados como UEs, conectados à rede comercial da Verizon, em Minneapolis (EUA). A campanha contemplou três cenários urbanos com cobertura mmWave: (i) um cruzamento no centro da cidade, composto por três torres 5G com múltiplos painéis; (ii) a área interna do aeroporto internacional de Minneapolis–Saint Paul (MSP), equipada com duas torres de painel único posicionadas frente a frente; e (iii) um trajeto de aproximadamente 1300 metros no entorno do estádio U.S. Bank, abrangendo ruas, restaurantes, cafés e parques.

O conjunto resultante compreendeu cerca de 563.840 amostras de *throughput*, registradas em resolução de um segundo, totalizando 38 GB de tráfego transferido. As medições abrangeram diferentes condições de mobilidade, incluindo 331 km percorridos a pé, 132 km em deslocamento veicular e cenários estacionários. Cada trajeto foi repetido diversas vezes, como no caso do cruzamento urbano, que contou com 12 trajetórias distintas, cada uma repetida mais de 30 vezes. Nos experimentos móveis, a velocidade de deslocamento variou entre 0 e 45 km/h.

A partir desse conjunto de dados original, Narayanan et al. (2020) analisaram o impacto de diferentes fatores sobre a variabilidade do *throughput* em mmWave. No que se refere à geolocalização, os autores apontam que, embora relevante, sua utilização isolada não é suficiente para prever com precisão o desempenho do enlace, em contraste com redes 4G, nas quais a posição geográfica fornecia indícios consistentes da QoS.

A mobilidade demonstrou exercer papel determinante, sendo que a inclusão de informações como distância até o painel, ângulo posicional, ângulo de mobilidade e velocidade de deslocamento reduziu significativamente a variabilidade do *throughput* e o erro dos modelos em comparação ao uso exclusivo da geolocalização. Já a distância entre o usuário e o painel apresentou impacto não linear: em determinados cenários, o *throughput* decaiu rapidamente com o aumento da distância, mas foi recuperado quando a linha de visão voltou a ser estabelecida após a presença de obstáculos.

Os ângulos relativos ao painel também se mostraram críticos. O ângulo de mobilidade influenciou diretamente o desempenho: quando o usuário se desloca paralelamente à frente do painel ($\theta_m = 0$), o corpo pode bloquear o sinal, reduzindo a qualidade do enlace. O ângulo posicional, por sua vez, condicionou variações no desempenho em função da orientação relativa entre terminal e painel, reforçando a importância de considerar a geometria do enlace.

De forma geral, os resultados indicam que o desempenho do *throughput* em 5G mmWave decorre da interação entre múltiplos fatores — geolocalização, mobilidade, distância e ângulos relativos ao painel. Essa complexidade, substancialmente maior que nas gerações anteriores, como o 4G, evidencia a necessidade de modelos preditivos capazes de incorporar variáveis contextuais adicionais.

3.2.2 Organização dos dados em grupos de atributos

Diferentemente de outros trabalhos da literatura, Narayanan et al. (2020) não descrevem um processo detalhado de pré-processamento envolvendo técnicas de remoção de *outliers*, imputação de valores ausentes ou normalização de atributos. O foco central do estudo recai sobre a coleta massiva de dados em cenários reais e sobre a posterior organização das variáveis por meio de agrupamentos.

Conforme sintetizado na Tabela 5, os autores estruturaram as variáveis em quatro grupos primários: L (Localização), composto por coordenadas geográficas pixelizadas, de forma a mitigar o ruído de medições brutas de GPS; M (Mobilidade), incluindo velocidade e direção da bússola (azimute), obtidas diretamente por sensores do UE; T (Torre), englobando distância ao painel 5G, ângulo posicional (θ_p) e ângulo de mobilidade (θ_m), atributos que não são coletados nativamente pelo terminal, mas calculados a partir de informações externas sobre a localização e a orientação das torres, medidas pelos autores ou fornecidas pela operadora; e C (Conexão), reunindo histórico de *throughput* e métricas da camada física, como tipo de rádio, intensidade de sinal (RSRP, RSRQ, RSSI, SS-RSRP)

e eventos de *handover* horizontal e vertical. Além desses grupos primários, o estudo também avaliou combinações entre eles, como L+M, T+M, L+M+C e T+M+C, com vistas a investigar ganhos preditivos decorrentes da integração de múltiplos fatores contextuais.

A análise conduzida pelos autores evidenciou que o uso isolado de geolocalização (L) não é suficiente para prever o desempenho do *throughput*, pois as coordenadas absolutas são específicas de uma região, o que restringe a capacidade de generalização dos modelos. Ainda assim, a localização mantém relevância, embora com menor peso do que em redes 4G, nas quais antenas tipicamente omnidirecionais tornavam a posição geográfica um bom indicador da qualidade de enlace. No caso do 5G mmWave, em que os painéis são direcionais, o desempenho passa a depender fortemente da orientação e da geometria do enlace. Nesse contexto, as variáveis de torre (T) e mobilidade (M) complementam as informações de L, permitindo capturar de forma mais precisa as condições reais do canal, enquanto atributos de conexão (C) adicionam uma camada adicional de contexto, tornando os modelos mais robustos.

Tabela 5 – Grupos de atributos utilizado por Lumos5G.

Grupo de variáveis	Lista de atributos
Primários	
L	Coordenadas geográficas pixelizadas (latitude e longitude)
M	Velocidade do dispositivo (UE) + Direção da bússola do dispositivo (azimute)
T	Distância UE–painel + Ângulo posicional UE–painel (θ_p) + Ângulo de mobilidade UE–painel (θ_m)
C	Histórico de <i>throughput</i> + (Métricas da camada física: tipo de rádio + intensidade de sinal LTE + intensidade de sinal 5G + <i>handover</i> horizontal + <i>handover</i> vertical)
Combinações	
L+M	(L) + Velocidade do dispositivo (UE) + Direção da bússola do dispositivo (azimute)
T+M	Velocidade do dispositivo (UE) + Distância UE–painel + Ângulo posicional UE–painel + Ângulo de mobilidade UE–painel
L+M+C	(L+M) + Tipo de rádio + Intensidade de sinal LTE + Intensidade de sinal 5G + <i>handover</i> horizontal + <i>handover</i> vertical
T+M+C	(T+M) + Tipo de rádio + Intensidade de sinal LTE + Intensidade de sinal 5G + <i>handover</i> horizontal + <i>handover</i> vertical

Fonte: Adaptado de Narayanan et al. (2020).

3.2.3 Modelos e métricas de avaliação

No estudo de Narayanan et al. (2020), a tarefa de previsão em redes 5G mmWave foi formulada sob duas perspectivas: (i) como um problema de classificação, em que o objetivo é estimar a qual faixa de desempenho do enlace pertence — baixa (menor que

300 Mbps), média (entre 300 e 700 Mbps) ou alta (acima de 700 Mbps); e (ii) como um problema de regressão, voltado à estimação de valores contínuos de *throughput*. Ambas as formulações foram avaliadas em horizontes de curto e longo prazo, de acordo com a natureza da aplicação considerada.

Segundo os autores, a previsão de curto prazo é particularmente relevante para cenários que demandam respostas rápidas, como a adaptação de taxa em serviços de *streaming* de vídeo, nos quais a simples identificação do nível de desempenho já é suficiente para orientar decisões. A previsão de longo prazo, por sua vez, está associada a trajetórias de mobilidade e ao planejamento de rede, exigindo a estimação contínua da capacidade do enlace ao longo do tempo. Esse horizonte demanda modelos capazes de capturar dependências temporais mais complexas, o que justifica a adoção de arquiteturas avançadas, como o modelo Seq2Seq.

Para abordar esses cenários, Narayanan et al. (2020) utilizaram dois modelos já discutidos no Capítulo 2: o GDBT, adotado por sua leveza, interpretabilidade, compatibilidade com diferentes grupos de atributos e desempenho superior a métodos clássicos em redes celulares anteriores, e o modelo Seq2Seq baseado em LSTM, escolhido pela capacidade de prever sequências em horizontes mais longos.

Tabela 6 – Parâmetros de treinamento dos modelos avaliados no Lumos5G.

Modelo	Configuração
GDBT	8000 estimadores profundidade máxima = 8 taxa de aprendizado = 0,01 função de perda = MSE
Seq2Seq (LSTM)	2 camadas LSTM 128 unidades ocultas 2000 épocas <i>batch size</i> = 256 comprimento da sequência de entrada e saída = 20

Fonte: Adaptado de Narayanan et al. (2020).

Os experimentos foram conduzidos a partir de uma divisão de dados em 70% para treinamento e 30% para teste, sendo as sequências temporais de medições escolhidas de forma aleatória, ou seja, o número de sequências foi particionado aleatoriamente entre treino e teste. Após um processo de calibração inicial, os hiperparâmetros foram fixados para todos os experimentos. O GDBT foi configurado com um grande número de estimadores, profundidade máxima limitada e taxa de aprendizado reduzida, equilibrando custo computacional e desempenho preditivo. Já o Seq2Seq adotou uma arquitetura profunda com múltiplas camadas, exigindo maior tempo de treinamento em GPU. A Tabela 6 resume os principais parâmetros empregados em cada modelo. Enquanto o GDBT foi treinado em questão de minutos, o Seq2Seq demandou entre 6 e 44 horas por modelo

gerado, evidenciando a diferença de complexidade entre as abordagens.

No processo de avaliação, para os experimentos de regressão, foram utilizadas as métricas MAE (Equação 6) e RMSE (Equação 5), por serem amplamente empregadas em tarefas de previsão de séries temporais. Já nos experimentos de classificação, foi utilizado o *Weighted F1-score* como métrica principal, complementado pelo *recall* da classe de baixo desempenho, embora esses resultados não constituam o foco da presente comparação.

3.2.4 Resultados

Os experimentos consideraram três cenários distintos — condições estacionárias, de caminhada e de deslocamentos veiculares — além de um conjunto de dados global que reúne informações de todas as áreas com localização conhecida dos painéis 5G.

Foram comparados dois modelos de naturezas distintas: o GDBT, que utiliza apenas atributos instantâneos, e o Seq2Seq, capaz de explorar sequências temporais de medições para capturar dependências mais complexas. Os resultados mostram que o uso isolado de variáveis de localização não é suficiente para garantir boa acurácia, sobretudo em contextos de mobilidade. A incorporação de atributos de mobilidade e conexão aumentou significativamente o desempenho, com o Seq2Seq superando o GDBT por considerar o histórico temporal e apresentar maior capacidade de representação. Essa superioridade é evidente nos experimentos de regressão, em que o Seq2Seq obteve erros substancialmente menores (MAE e RMSE) na maioria dos cenários, conforme apresentado na Tabela 7 e na Tabela 8.

A análise de transferibilidade reforça esse resultado ao mostrar que os atributos baseados em torre (T), por serem independentes da localização absoluta, oferecem desempenho semelhante aos de localização (L), mas com a vantagem de possibilitar a generalização entre áreas com características semelhantes, evidenciando o potencial de reutilização de modelos em ambientes similares.

Outro ponto relevante é a análise de importância dos atributos, que indicou não haver uma variável isolada ou grupo de variáveis dominante na predição do desempenho. O *throughput* em redes 5G mmWave resulta, portanto, da interação complexa entre múltiplos fatores, e não de um único preditor.

Por fim, a comparação com modelos clássicos utilizados em 3G/4G confirma a superioridade da combinação criteriosa de diferentes atributos e a expressividade das arquiteturas, em especial do Seq2Seq profundo, garantiram ganhos substanciais em relação a métodos como *Random Forest* e KNN.

3.2.5 Análise crítica

Os resultados evidenciam uma clara relação de compromisso entre capacidade preditiva e custo computacional. O GDBT, por não explorar o contexto temporal de janelas de

Tabela 7 – Resultados de regressão (MAE) para diferentes modelos e grupos de variáveis

Grupo de variáveis	KNN	RF	GDBT	Seq2Seq
L	285	300	225	208
L+M	229	256	127	74
T+M	252	173	115	52
L+M+C	223	162	109	49
T+M+C	228	163	100	57

Fonte: Adaptado de Narayanan et al. (2020).

Tabela 8 – Resultados de regressão (RMSE) para diferentes modelos e grupos de variáveis

Grupo de variáveis	KNN	RF	GDBT	Seq2Seq
L	362	378	314	273
L+M	303	330	186	144
T+M	326	253	173	109
L+M+C	311	241	166	112
T+M+C	320	241	154	119

Fonte: Adaptado de Narayanan et al. (2020).

entrada, apresentou desempenho inferior ao modelo Seq2Seq, mas destacou-se pelo tempo de treinamento bastante reduzido, limitado a poucos minutos. O Seq2Seq, por sua vez, alcançou os menores erros, MAE e RMSE, na maioria dos cenários, beneficiando-se da captura de dependências temporais mais complexas. Esse ganho, contudo, veio acompanhado de elevado custo computacional, com tempos de treinamento entre 6 e 44 horas, o que compromete sua aplicabilidade em contextos que exigem atualizações frequentes ou contam com recursos de hardware limitados. Nesse sentido, embora menos preciso, o GDBT não pode ser descartado, permanecendo como uma alternativa prática e servindo como referência comparativa frente a modelos de aprendizado profundo ao longo das análises desenvolvidas nesta tese.

Outro aspecto relevante é a questão da transferibilidade. O uso exclusivo de variáveis de localização (L) restringe a generalização dos modelos, pois as coordenadas absolutas são específicas de cada cenário. Para mitigar essa limitação, os autores sugerem complementar L com atributos de torre (T), que independem da posição absoluta do usuário e favorecem a adaptação entre áreas com características semelhantes.

Por fim, destaca-se o papel do grupo de atributos de conexão (C), que reúne o valor da última medida de *throughput*, além de métricas de QoS e eventos de *handover*. Esse conjunto mostrou impacto positivo quando combinado a L e M, reforçando a relevância do contexto de rede para melhorar a previsão. Ainda assim, como os autores não avaliaram C de forma isolada, permanece em aberto a análise mais precisa de sua contribuição individual.

3.3 Uso de aprendizado profundo na previsão de *throughput* em redes 5G

O estudo de Batool, Fouda e Fadlullah (2024) propõe modelos de previsão de *throughput* em redes 5G, utilizando o conjunto de dados público de Raca et al. (2020). Foram avaliadas três arquiteturas de aprendizado profundo — Rede Neural Artificial (RNA), LSTM e LSTM Bidirecional (BiLSTM) — em comparação ao algoritmo *Random Forest*, representante dos métodos baseados em *ensemble*. A variável-alvo foi definida a partir da combinação das taxas de *downlink* e *uplink*, por meio da média geométrica, devido à similaridade em escala e distribuição.

3.3.1 Pré-processamento dos dados

No pré-processamento, os autores aplicaram técnicas de agrupamento de dados para eliminar valores discrepantes e converter atributos categóricos em numéricos. Além disso, utilizaram a normalização pelo método *min-max scaling*, que reescala os valores para um intervalo definido, conforme a Equação 18:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (18)$$

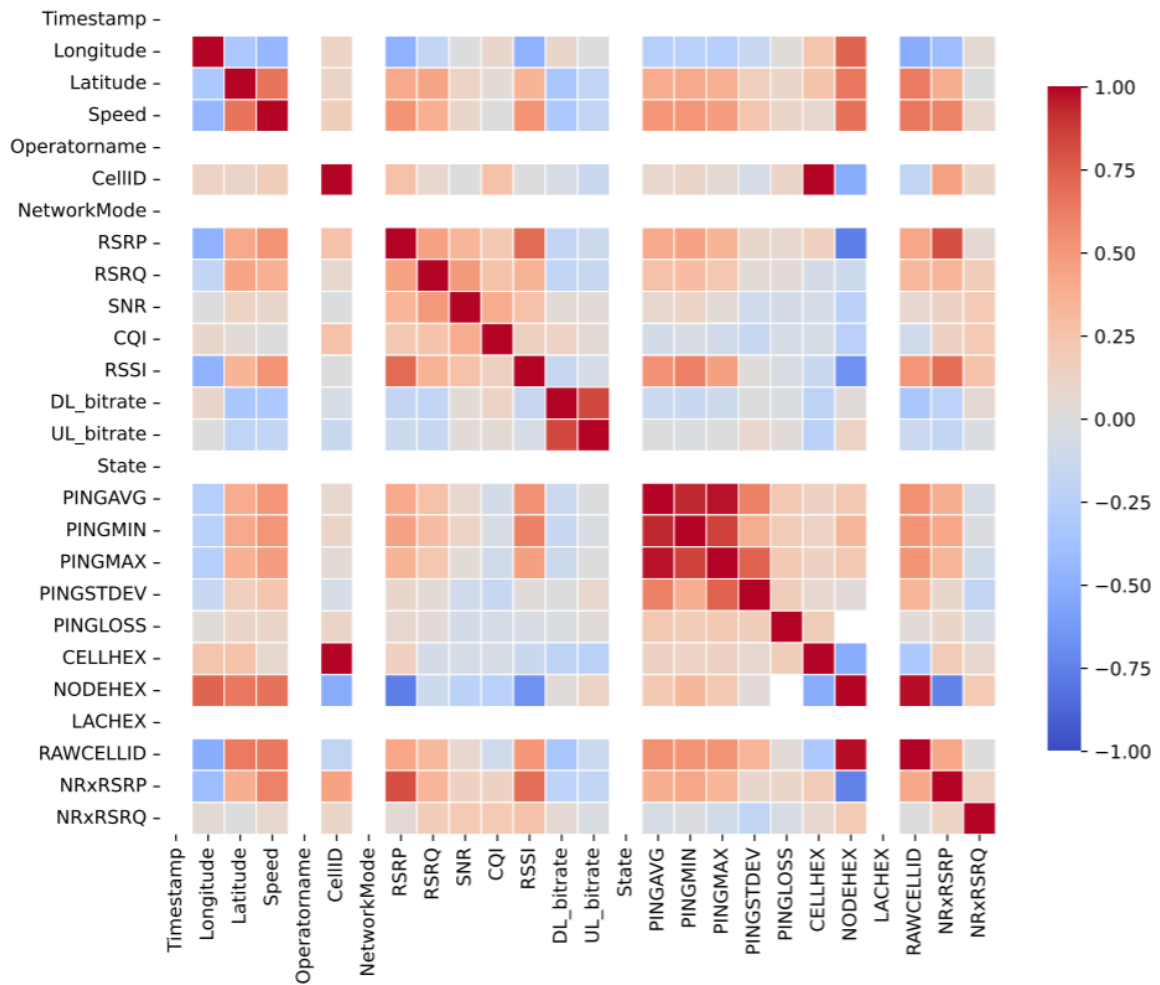
Foi também conduzida uma análise de correlação de Spearman entre as métricas de rádio e variáveis contextuais, ilustrada na Figura 14. Os resultados mostraram correlação positiva entre RSRP, RSRQ, SNR e CQI, tanto entre si quanto com o *throughput*, enquanto o RSSI apresentou associação mais discreta. Já as métricas de latência (PINGAVG, PINGMIN, PINGMAX e PINGSTDEV) exibiram correlação negativa, indicando que maiores atrasos reduzem a taxa de transmissão.

Entre as variáveis contextuais, a velocidade (*Speed*) mostrou impacto significativo tanto no *throughput* quanto nos indicadores de qualidade de sinal, refletindo a influência da mobilidade. Latitude e longitude também apresentaram forte correlação, reforçando a importância da localização geográfica. Já os identificadores de célula (CELLHEX, NODEHEX, LACHEX e RAWCELLID) foram considerados redundantes, atuando mais como marcadores de contexto do que como determinantes diretos da qualidade de canal.

3.3.2 Modelos e métricas de avaliação

Foram avaliados três modelos de aprendizado profundo (RNA, LSTM e BiLSTM) e um modelo clássico (*Random Forest*). O desempenho foi medido em termos de erro quadrático médio (MSE) e coeficiente de determinação (R^2), conforme resultados na Tabela 9.

Figura 14 – Correlação de Spearman entre características atributos da base dados celular 5G.



Fonte: Elaborado por Batool, Fouda e Fadlullah (2024).

3.3.3 Resultados

Segundo Batool, Fouda e Fadlullah (2024), o BiLSTM obteve o menor MSE, sugerindo maior precisão em relação aos demais métodos. O LSTM apresentou desempenho próximo, reforçando o potencial das arquiteturas recorrentes na captura de dependências temporais. A RNA, embora tenha registrado erro superior, alcançou R^2 semelhante ao *Random Forest*, mostrando que mesmo arquiteturas mais simples podem oferecer poder explicativo comparável a *ensemble*. O *Random Forest*, por sua vez, apresentou MSE elevado, ainda que com R^2 próximo ao da RNA, sendo considerado limitado frente à complexidade dos dados de redes 5G.

Tabela 9 – Melhores resultados obtidos com os modelos LSTM, BiLSTM, ANN e *Random Forest*

Modelo	Camadas	Épocas	Tamanho do Batch	MSE	R^2
LSTM	3	250	32	0.00012	0.31
BiLSTM	3	250	32	0.00009	0.28
RNA	3	100	32	0.0052	0.36
<i>Random Forest</i>	–	–	–	1.21	0.36

Fonte: Adaptado de Batool, Fouda e Fadlullah (2024).

3.3.4 Análise crítica

Os autores concluíram que modelos de aprendizado profundo, em especial os recorrentes, possuem maior potencial para previsão de *throughput* em cenários de alta variabilidade. Entretanto, ao analisar os resultados, observa-se que, embora o BiLSTM tenha alcançado o menor MSE, tanto a RNA quanto a *Random Forest* obtiveram os maiores valores de R^2 , sugerindo equilíbrio entre precisão e variância explicada. Nota-se ainda a ausência de comparações com modelos estatísticos ou estratégias ingênuas (*naive*), bem como a falta de exploração do impacto das covariáveis, pontos que poderiam oferecer uma avaliação mais abrangente do ganho proporcionado pelas arquiteturas propostas.

3.4 Modelos de aprendizado de máquina com pré-processamento e seleção de atributos

Sharma, Pandit e Talluri (2025) destacam que grande parte das abordagens no estado da arte tratam a previsão de *throughput* exclusivamente como um problema contínuo, formulado como regressão. Diferentemente, os autores consideram o *throughput* tanto como variável contínua quanto discreta, ressaltando que a formulação discreta possibilita antecipar valores em instantes específicos, algo inviável apenas com a abordagem contínua.

No entanto, trabalhos anteriores, como o de Narayanan et al. (2020), já haviam explorado essa perspectiva mista, tratando o *throughput* em termos contínuos e categóricos para aproximar a previsão da experiência do usuário. Ainda assim, tais abordagens se fundamentam em atributos estáticos da rede e na regressão direta da métrica alvo, sem considerar a modelagem temporal.

Os autores também propuseram um método baseado em árvores de decisão, denominado *Enhanced Sequential Decision Tree* (ESDT). Nesse modelo, busca-se reduzir iterativamente o erro residual ao ajustar previsões com base na derivada da função de perda em cada estágio, comparando tal abordagem com métodos considerados tradicionais.

3.4.1 Pré-processamento dos dados

Sharma, Pandit e Talluri (2025) destacam o tratamento de valores ausentes como etapa fundamental do pré-processamento, optando pela imputação em vez da remoção de registros incompletos, de modo a preservar observações informativas e reduzir vieses nos modelos. Os autores mostram que, após a imputação, a distribuição da medida RSRQ manteve-se consistente, evidenciada pela sobreposição entre os dados originais e imputados. Contudo, o método de imputação empregado não é explicitado, o que limita a reprodutibilidade e dificulta comparações. Além disso, a alegação de que o tema é pouco tratado na literatura revela-se imprecisa, visto que Elsherbiny et al. (2020) já aplicaram imputação por média/mediana, enquanto Batool, Fouda e Fadlullah (2024) descreveram técnicas de normalização e outros procedimentos de preparação dos dados, ainda que sem detalhar o processo de imputação.

No mesmo trabalho, os autores também relatam o uso de *one-hot encoding* para converter variáveis categóricas em numéricas, representando as categorias de mobilidade (caminhando ou dirigindo) por vetores binários. Já em relação a distribuição dos atributos, tanto Sharma, Pandit e Talluri (2025) quanto Elsherbiny et al. (2020) empregam a padronização pelo *score-z* (Equação 16), garantindo média zero e desvio-padrão unitário, enquanto Batool, Fouda e Fadlullah (2024) utilizam a normalização *min-max* (Equação 18) para reescalar os valores no intervalo $[0, 1]$. Cabe destacar que a literatura não apresenta consenso sobre qual técnica é mais adequada, sendo a escolha dependente do contexto da aplicação.

Por fim, Sharma, Pandit e Talluri (2025) realizaram uma seleção de atributos utilizando o algoritmo *Random Forest* em conjunto com uma técnica de seleção recursiva, identificando as variáveis mais relevantes para a previsão de *throughput*. Entre elas destacam-se: a Situação de Novo Rádio (*New Radio Status*)⁴, latitude, longitude, velocidade de movimento, direção da bússola e RSRP, enquanto variáveis como Identificação da Torre de Rádio e Modo de Mobilidade, derivado pelo *one-hot encoding*, apresentaram baixa importância e, segundo os autores, podem ser descartadas sem impacto significativo no modelo.

3.4.2 Modelos e métricas de avaliação

Para a avaliação dos modelos de regressão, Sharma, Pandit e Talluri (2025) utilizaram as métricas MAE, RMSE e R^2 , amplamente empregadas na análise de predição e regressão. Essas medidas foram utilizadas para comparar o desempenho dos diferentes métodos de

⁴ O atributo *New Radio Status* representa o estado da conexão entre o dispositivo e a rede 5G, podendo assumir três valores: *Conectado*, quando há uma conexão ativa; *Não restrito*, quando a conexão está estabelecida, mas a taxa de dados pode ser degradada devido à baixa qualidade do sinal; e *Nenhuma*, quando não existe conexão entre o dispositivo e a estação-base.

previsão de *throughput*, permitindo identificar aqueles que apresentaram maior acurácia e consistência.

O modelo proposto pelos autores, denominado ESDT, é descrito como um processo iterativo de construção sequencial de árvores de decisão. Nesse esquema, um modelo-base gera uma previsão inicial de *throughput*, a partir da qual é calculado o erro residual. Em seguida, uma nova árvore é treinada para modelar esse erro, reduzindo-o progressivamente. O procedimento é repetido em múltiplos estágios, de modo que cada árvore subsequente corrige as limitações da anterior, até a obtenção de uma previsão com erro acumulado mínimo. Tal abordagem configura um *ensemble* sequencial de árvores, conceitualmente semelhante ao GBDT, embora apresentada sob a denominação de ESDT.

A Tabela 10 apresenta os hiperparâmetros adotados para os modelos de comparação (SVM, DT, RF e KNN) e para o próprio ESDT. Observa-se, contudo, que para este último os autores reportam apenas a taxa de aprendizado ($\alpha = 0,02$), sem especificar parâmetros estruturais típicos de árvores, como profundidade máxima, número de nós ou critérios de divisão. Essa ausência de informações compromete a reprodutibilidade do modelo e dificulta comparações consistentes com outros trabalhos da literatura.

Tabela 10 – Hiperparâmetros definidos para os modelos comparados pelo modelo ESDT.

Modelo	Hiperparâmetros
SVM	Parâmetro de penalidade = 20
DT	Profundidade máxima da árvore = 55
RF	Profundidade máxima da árvore = 60 Número máximo de árvores = 55
KNN	Número de vizinhos mais próximos = 7
ESDT	Taxa de aprendizado (α) = 0,02

Fonte: Adaptado de Sharma, Pandit e Talluri (2025).

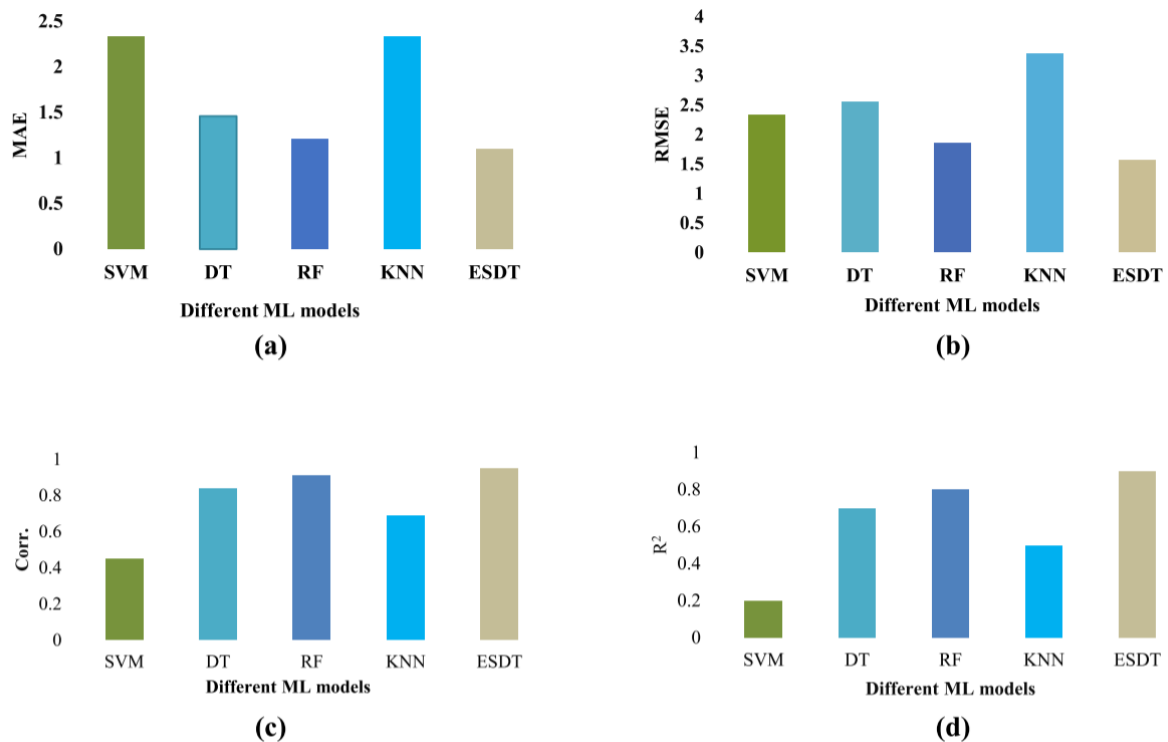
3.4.3 Resultados

Sharma, Pandit e Talluri (2025) avaliaram o desempenho dos modelos SVM, DT, RF, KNN e ESDT em dois conjuntos de dados distintos: no primeiro, os registros com valores ausentes foram removidos, enquanto no segundo foi aplicado um pré-processamento mais abrangente, incluindo imputação de valores, codificação de variáveis categóricas por meio de *one-hot encoding*, seleção híbrida de atributos e ajuste de hiperparâmetros.

Os resultados consolidados do segundo conjunto de dados estão representados na Figura 15, que reúne as métricas MAE, RMSE, Correlação e R^2 . O ESDT apresentou desempenho superior aos demais modelos, registrando menores valores de MAE e RMSE, além de R^2 mais elevado, o que indica maior proximidade entre os valores previstos e observados. Em contraste, modelos como SVM e KNN exibiram erros mais altos e menor qualidade de ajuste, enquanto DT e RF obtiveram desempenho intermediário.

Conforme evidenciado pelos autores, a aplicação do pré-processamento no segundo conjunto de dados resultou em ganhos significativos para todos os modelos, com destaque para o ESDT, que obteve os melhores resultados em relação às abordagens tradicionais. O estudo reforça que a preparação adequada dos dados, aliada à abordagem sequencial de correção de resíduos, foi determinante para a melhoria do desempenho observado.

Figura 15 – Resultados de regressão com diferentes modelos de ML em redes móveis: (a) MAE, (b) RMSE, (c) Correlação e (d) R^2 .



Fonte: Elaborado por Sharma, Pandit e Talluri (2025).

3.4.4 Análise crítica

Como Narayanan et al. (2020), o estudo de Sharma, Pandit e Talluri (2025) concentra-se principalmente no problema de classificação, tratando a previsão de *throughput* apenas como regressão linear, sem explorar adequadamente as características temporais.

No que se refere ao pré-processamento, a imputação é descrita de forma vaga, sem indicar se foi utilizada média, mediana ou outro método, o que compromete a reprodutibilidade. De forma semelhante, no modelo ESDT apenas a taxa de aprendizado ($\alpha = 0,02$) é reportada, sem detalhamento de parâmetros estruturais como profundidade ou critérios de divisão, ainda que a proposta apresente forte semelhança conceitual com o GBDT. Além disso, embora os autores afirmem aplicar normalização e seleção de atributos, a descrição não especifica claramente os procedimentos adotados. Por fim, os resultados

são apresentados exclusivamente em gráficos (MAE, RMSE, R^2 , Correlação), sem valores tabulares, o que dificulta comparações quantitativas diretas com outros trabalhos.

3.5 Previsão de *throughput* em comunicações 5G NR V2X com aprendizado profundo híbrido

O estudo de Yeaser e Hassan (2025) aborda a quinta geração de comunicações veiculares (5G V2X), destacando a previsão de *throughput* como elemento estratégico para otimizar a alocação de recursos e o ajuste da potência de transmissão, de modo a mitigar perdas de percurso, interferências e problemas de *handover*.

Para alcançar esse objetivo, Yeaser e Hassan (2025) propuseram um modelo híbrido de aprendizado profundo, treinado com dados reais de 5G, capaz de estimar o *throughput* tanto no *uplink* quanto no *downlink*. A arquitetura apresentada pelos autores combina LSTM e GRU, explorando as vantagens de cada uma dessas redes neurais recorrentes. Ressalta-se que os autores utilizaram em seus experimentos o conjunto de dados de Raca et al. (2020).

3.5.1 Pré-processamento dos dados

Yeaser e Hassan (2025) relatam ter utilizado como variáveis de entrada métricas de qualidade de canal, como SNR, RSRP e RSRQ, além da velocidade da UE. Na base de dados empregada, informam que foi realizado o escalonamento das variáveis e a remoção das linhas com valores ausentes, a fim de evitar vieses no treinamento. Contudo, não especificam quais técnicas foram adotadas nesses procedimentos.

3.5.2 Modelos e métricas de avaliação

Yeaser e Hassan (2025) destacam que diferentes modelos de aprendizado profundo apresentam eficiências distintas conforme as condições e o conjunto de dados, de modo que não existe um método universalmente superior. Nesse contexto, justificam o uso de uma arquitetura híbrida que combina LSTM e GRU, explorando as vantagens complementares de cada abordagem para aumentar a eficácia da previsão de *throughput*.

A estrutura do modelo híbrido é apresentada na Figura 16, composta por uma camada densa inicial com 80 neurônios, seguida por uma camada LSTM com 40 neurônios e uma camada GRU com 20 neurônios, finalizando com camadas densas adicionais para a saída. Segundo os autores, a LSTM foi empregada para mitigar o problema da dissipação do gradiente e capturar dependências temporais, enquanto a GRU foi incorporada pela sua maior eficiência em cenários com menor volume de dados.

Figura 16 – Estrutura do modelo híbrido LSTM–GRU proposto.



Fonte: Elaborado por Yeaser e Hassan (2025).

Tabela 11 – Hiperparâmetros definidos para o modelo híbrido LSTM–GRU.

Modelo	Hiperparâmetros
Híbrido (LSTM–GRU)	Camadas: Densa (80 unidades) LSTM (40 unidades) GRU (20 unidades) Densa (20 unidades) Épocas = 150 Taxa de aprendizado = 0,001 Função de ativação = ReLU Otimizador = Adam

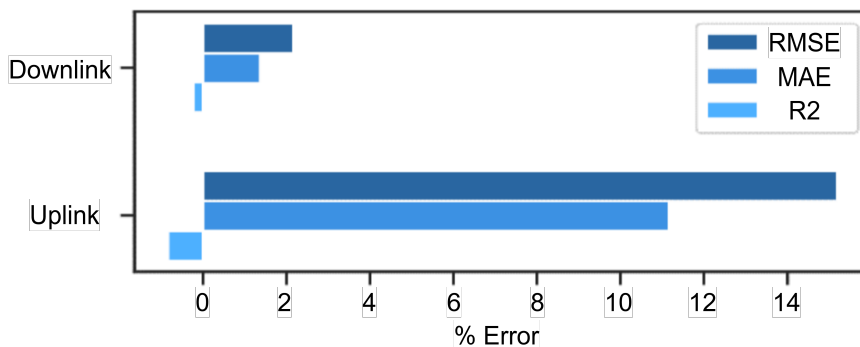
Fonte: Adaptado de Yeaser e Hassan (2025).

Os principais hiperparâmetros adotados encontram-se resumidos na Tabela 11. Para avaliação do desempenho, foram empregadas métricas de regressão como MAE e RMSE, complementadas pela análise da tendência do *throughput* em diferentes cenários de SNR e velocidade veicular.

3.5.3 Resultados

Os resultados de Yeaser e Hassan (2025) mostram que o modelo híbrido capturou de forma eficaz a influência dos parâmetros de canal e da mobilidade na previsão do *throughput*. No *downlink*, observou-se aumento da taxa com o SNR e queda com a elevação da velocidade do veículo. Por fim, os autores destacam que o modelo apresentou estabilidade nas previsões em ambos os sentidos de transmissão, alcançando o objetivo central do estudo.

Figura 17 – Análise de erro do modelo híbrido na previsão de *throughput*.



Fonte: Adaptado de Yeaser e Hassan (2025).

A Figura 17 mostra que o modelo apresentou bom desempenho na previsão da taxa de dados em *downlink*, enquanto no *uplink* os resultados foram inferiores. Os autores atribuem essa diferença ao fato de que a UE envia dados à rede apenas quando necessário, e não continuamente, o que dificulta a previsão precisa. Ressalta-se, contudo, que os resultados são apresentados apenas em forma gráfica e em porcentagem, o que dificulta a identificação dos valores reais e compromete a reprodutibilidade da análise.

3.5.4 Análise crítica

O estudo de Yeaser e Hassan (2025) demonstra que a arquitetura, que combina LSTM e GRU, foi capaz de capturar a influência de parâmetros de canal, como SNR, RSRP e RSRQ, em conjunto com a velocidade da UE em um veículo, alcançando estabilidade na previsão tanto em *uplink* quanto em *downlink*.

Entretanto, algumas limitações metodológicas devem ser destacadas. Os resultados são apresentados apenas em figuras, expressos em porcentagem, sem valores tabulares, o que compromete tanto a reprodutibilidade quanto a comparação quantitativa com outros trabalhos. Soma-se a isso a falta de detalhamento sobre o pré-processamento aplicado, em especial no que se refere às estratégias de normalização e imputação.

Apesar dessas restrições, o trabalho representa um ponto de comparação relevante para esta tese, pois utiliza a mesma base de dados de Raca et al. (2020) e foca no grupo

de variáveis de conexão (C), definido em Narayanan et al. (2020). Esse grupo reúne histórico de *throughput* e métricas da camada física, como intensidade de sinal LTE e 5G, tipo de rádio e eventos de *handover*, que não haviam sido analisados isoladamente em estudos anteriores, conferindo ao trabalho caráter complementar e valor como referência comparativa.

3.6 Síntese da revisão da literatura

A Tabela 12 apresenta uma comparação entre os trabalhos mais relevantes da literatura, detalhados nas seções anteriores. A tabela destaca modelos utilizados, técnicas de pré-processamento e resultados alcançados.

Tabela 12 – Síntese dos trabalhos relacionados sobre previsão de *throughput* em redes móveis.

Autores/Ano	Modelos Com-parados	Pré-processamento	Resultados
Elsherbiny et al. (2020)	RF, SVR, KNN, ARIMA, LSTM	Imputação (média/mediana); padronização (<i>z</i> -score); remoção de <i>outliers</i> por <i>z</i> -score; agregação em janelas	RF com melhor desempenho ($R^2 \approx 0,78$; RMSE ≈ 8.200); ARIMA supera LSTM em série temporal
Batool, Fouda e Fadlullah (2024)	LSTM, BiLSTM, ANN, RF	Normalização min-max; codificação categórica; sem detalhar imputação	BiLSTM com menor MSE (9×10^{-5}); RNA e RF com $R^2 \approx 0,36$
Narayanan et al. (2020)	GBDT, Seq2Seq (LSTM)	Organização em grupos L, M, T, C; sem detalhar imputação/normatização	Seq2Seq reduz MAE/RMSE vs. GBDT quando usa L+M(+C/T); localização isolada é insuficiente
Sharma, Pandit e Talluri (2025)	SVM, DT, RF, KNN, ESDT	Imputação (não especificada); <i>one-hot</i> ; padronização (<i>z</i> -score); seleção híbrida de atributos (RF + seleção recursiva)	ESDT supera SVM/DT/RF/KNN (melhor MAE/RMSE e R^2) em figuras; sem valores tabulares reportados
Yeaser e Hasan (2025)	LSTM+GRU (híbrido)	Escalonamento; remoção de linhas com valores ausentes (sem detalhar técnica)	Estável em <i>downlink</i> ; <i>uplink</i> com maior erro; resultados apresentados em figuras percentuais

Fonte: Autoria própria.

3.7 Síntese crítica

A análise comparativa apresentada na Tabela 12 evidencia que diferentes abordagens têm contribuído para o avanço da previsão de *throughput* em redes móveis, mas cada uma apresenta limitações que afetam sua generalização.

O trabalho de Elsherbiny et al. (2020) demonstrou que modelos clássicos, como *Random Forest* e ARIMA, podem superar arquiteturas mais complexas em cenários com dados limitados, mas careceu de uma linha de base com modelos ingênuos e explorou pouco o uso de covariáveis. Já Batool, Fouda e Fadlullah (2024) reforçaram o potencial de arquiteturas recorrentes (LSTM e BiLSTM) em cenários de alta variabilidade, embora seus resultados revelem inconsistências entre MSE e R^2 e a ausência de comparações com modelos estatísticos ou tradicionais.

O *framework* Lumos5G, proposto por Narayanan et al. (2020), ofereceu contribuição importantes ao organizar atributos em grupos semânticos (L, M, T, C), demonstrando que a previsão em 5G mmWave depende da interação entre múltiplos fatores de mobilidade, torre e conexão, sendo insuficiente basear-se apenas na localização. Entretanto, o estudo não detalhou procedimentos de pré-processamento nem explorou o grupo de conexão (C) de forma isolada, deixando em aberto sua real contribuição.

Em seguida, Sharma, Pandit e Talluri (2025) propuseram o ESDT, uma variação conceitualmente próxima ao GBDT, mas sem detalhar parâmetros estruturais relevantes, o que compromete a reprodutibilidade. Seus resultados mostraram superioridade frente a modelos tradicionais (SVM, DT, RF, KNN), mas foram apresentados apenas em figuras, dificultando comparações quantitativas.

Por fim, Yeaser e Hassan (2025) exploraram a previsão em cenários 5G V2X com um modelo híbrido LSTM–GRU, alcançando estabilidade na previsão de *throughput* em ambos os sentidos de transmissão e destacando a relevância das variáveis de conexão (C). No entanto, os resultados foram reportados apenas em gráficos percentuais, sem valores tabulares, e o pré-processamento foi descrito de forma superficial, sem indicar técnicas específicas de normalização ou imputação.

De forma geral, os trabalhos mais recentes mostram avanços no emprego de arquiteturas híbridas e no aproveitamento de atributos contextuais, mas ainda possuem limitações importantes quanto à transparência metodológica, ao custo computacional e à reprodutibilidade dos resultados.

A revisão da literatura evidencia: (i) ausência sistemática de linhas de base (*baselines*) ingênuas e estatísticas; (ii) pouca distinção entre modelos locais e globais; (iii) comparação limitada entre horizontes de previsão; (iv) uso heterogêneo — e raramente isolado — das covariáveis L/M/T/C; e (v) protocolos de validação predominantemente aleatórios, com risco de vazamento temporal. Nesta dissertação, essas limitações são abordadas sistematicamente por meio de um protocolo reprodutível de validação temporal; comparação entre modelos locais e globais; avaliação com e sem covariáveis (L/M/T/C); e inclusão

explícita de linhas de base (*baselines*) ingênuas e estatísticas, com relato completo dos hiperparâmetros e do conjunto de métricas a fim de promover a plena reprodutibilidade dos experimentos realizados.

Capítulo 4

Metodologia

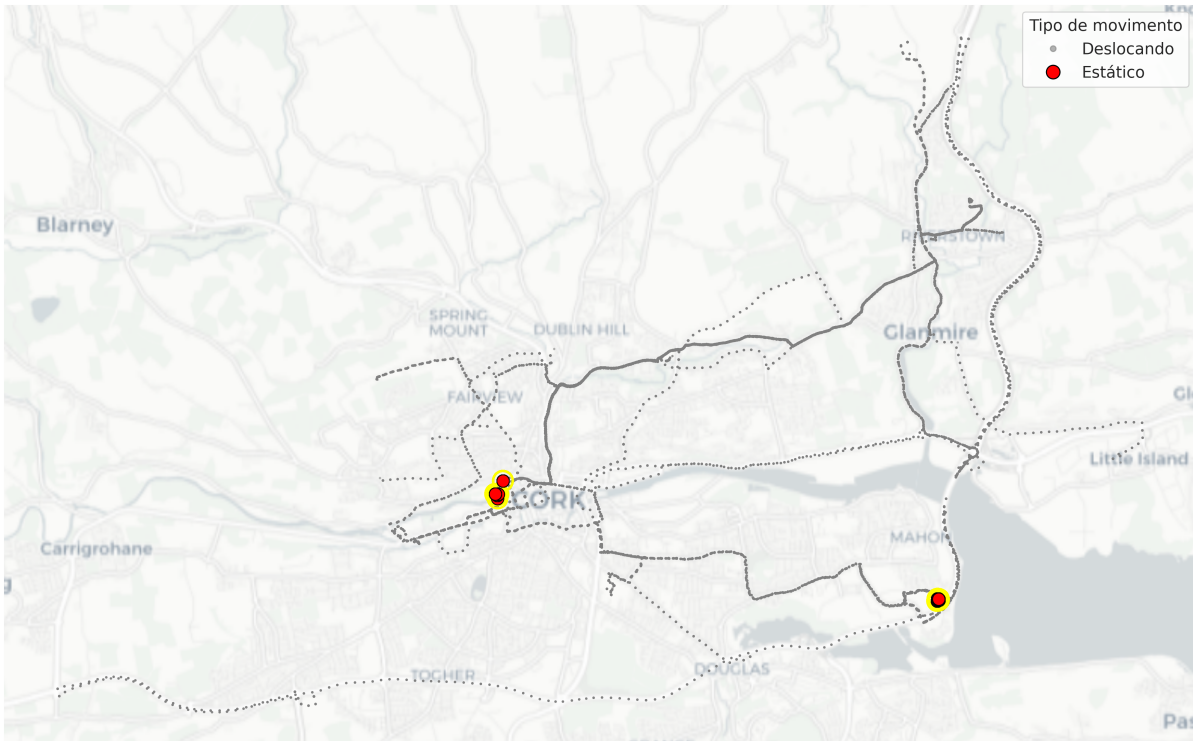
Esta seção detalha a metodologia para a previsão do *throughput* utilizando métricas de qualidade do canal e a velocidade do UE como covariáveis. As seções seguintes abordam o conjunto de dados utilizado, as etapas de pré-processamento aplicadas aos dados, as métricas de desempenho empregadas para avaliar os modelos e o protocolo que detalha os procedimentos de treinamento e teste.

4.1 Explorando o conjunto de dados

Todos os experimentos foram realizados com o conjunto de dados criado e disponibilizado por Raca et al. (2020), o qual reúne 83 trajetórias de medição, totalizando mais de 3.000 minutos de observações coletadas com resolução temporal de um segundo. As coletas contemplam diferentes cenários de mobilidade e aplicação, abrangendo tanto condições estáticas quanto em movimento, por meio do uso de uma UE em condução veicular. Além disso, incluem distintos perfis de aplicação, como transmissão de vídeo (*streaming*) e *download* de arquivos. A Figura 18 apresenta a localização geográfica de todas as medições registradas no conjunto de dados, distinguindo os pontos referentes a cenários estáticos daqueles obtidos durante deslocamento.

O trajeto da UE, ilustrado na Figura 18, contempla percursos em áreas urbanas e suburbanas da Irlanda, abrangendo tanto medições em ambiente interno estático quanto deslocamentos realizados por condução veicular (RACA et al., 2020). Essa distribuição espacial está naturalmente associada às coordenadas de latitude e longitude, uma vez que a cobertura do sinal varia com a localização geográfica (NARAYANAN et al., 2020; ELSHERBINY et al., 2020). Ainda assim, a análise conduzida não incorpora essa dependência espacial, pois o foco recai sobre o desenvolvimento de modelos capazes de ge-

Figura 18 – Localização geográfica de todas as medições do conjunto de dados de Raca et al. (2020), com distinção entre pontos coletados em cenários estáticos e em deslocamento.



Fonte: Autoria própria.

neralizar a previsão de métricas de rede com base exclusivamente nas condições temporais do sinal. Tal escolha metodológica está alinhada a discussões recentes que recomendam mitigar vieses geográficos em modelos de previsão (BATOOL; FOUDA; FADLULLAH, 2024; SHARMA; PANDIT; TALLURI, 2025), ampliando a aplicabilidade dos resultados em distintos cenários e contextos de mobilidade.

Raca et al. (2020) destacam que, embora o conjunto de dados tenha sido coletado em uma rede comercial já em funcionamento, a cobertura do sinal não se estendia a todo o trajeto percorrido. Essa limitação resultou em transições recorrentes para tecnologias legadas de redes móveis, notadamente 4G, nos cenários de mobilidade, em razão da restrita disponibilidade do 5G. A Tabela 13 apresenta as estatísticas descritivas do *throughput* em diferentes condições de mobilidade e atividade do usuário, evidenciando contrastes significativos entre os cenários analisados. Em ambiente estático, o desempenho médio em tarefas de *download* foi de 71,14 Mbps, com picos de até 333,07 Mbps, enquanto na transmissão de vídeo a média reduziu para 2,46 Mbps, ainda assim alcançando valores superiores a 190 Mbps. Em contrapartida, sob mobilidade, a média do *throughput* em *download* caiu para 29,67 Mbps, embora com picos de 532,90 Mbps, e para apenas 0,75 Mbps em transmissão de vídeo, refletindo forte instabilidade na vazão de dados. Essa

diferença entre *download* e *streaming* pode estar associada ao próprio perfil de consumo: enquanto o *download* de arquivos demanda tráfego contínuo e de alta taxa, o *streaming* depende de mecanismos de *buffer*, que reduzem a média observada ao longo do tempo, mas ainda permitem picos ocasionais. Tais desempenhos evidenciam o impacto conjunto das condições de mobilidade e da natureza da aplicação sobre a qualidade da conexão, reforçando a heterogeneidade característica de redes móveis em processo de transição tecnológica.

Tabela 13 – Estatísticas descritivas do *throughput* (em Mbps) por mobilidade e atividade do usuário. Nota: D.P. = desvio-padrão; N = número de amostras.

Mobilidade	Download de Arquivo				Transmissão de Vídeo			
	Média ± D.P.	Min – Máx	N	Média ± D.P.	Min – Máx	N		
Condução	29,67	58,78	0,00–532,90	24.446	0,75	3,08	0,00–111,62	72.336
Estático	71,14	77,85	0,00–333,07	14.490	2,46	6,63	0,00–194,26	69.471

Fonte: Autoria própria.

A análise das matrizes de correlação de Pearson (PEARSON, 1895; RODGERS; NICEWANDER, 1988) apresentadas nas Figuras 19, 20 e 21 evidencia relações relevantes entre as métricas de qualidade de canal e o *throughput*. No conjunto global (Figura 19), observa-se forte associação entre RSRP e RSSI (Corr $\approx 0,72$), além de correlações moderadas entre RSRQ, SNR e CQI, confirmando a interdependência estrutural dessas métricas na caracterização do enlace de rádio. Por outro lado, o *throughput* apresenta apenas correlações fracas, com destaque para a relação negativa com o RSSI em condição estática (Figura 20), onde se verificam valores de Corr $\approx -0,53$. Já no cenário de mobilidade (Figura 21), as correlações entre o *throughput* e as métricas de canal tornam-se ainda menos expressivas, refletindo maior instabilidade e variabilidade do enlace em deslocamento. Esses resultados reforçam que, embora métricas como RSRP, RSRQ, SNR, CQI e RSSI sejam fundamentais para caracterizar as condições de rádio, sua capacidade de explicar diretamente a variação do *throughput* é limitada. É importante destacar que a correlação de Pearson captura apenas relações lineares entre as variáveis, de modo que possíveis dependências não lineares relevantes na dinâmica do canal podem não ser identificadas por essa análise.

Todas as 83 séries temporais foram inicialmente avaliadas quanto à presença de tendência e sazonalidade. Entretanto, devido à elevada granularidade das medições — realizadas em intervalos de apenas um segundo — a maioria não apresentou padrões de longo prazo ou ciclos sazonais facilmente identificáveis. Para examinar a estacionariedade do *throughput*, aplicou-se diretamente o teste de Dickey-Fuller Aumentado (ADF) (DICKEY; FULLER, 1979). Os resultados, apresentados na Figura 22, revelam que aproximadamente 95% das séries podem ser consideradas estacionárias, restando apenas uma pequena fração que requer diferenciação mínima para atingir esse estado. Esse resultado é

Figura 19 – Correlação de Pearson calculada para todo o conjunto de dados (condições estáticas e em condução), abrangendo métricas de canal.



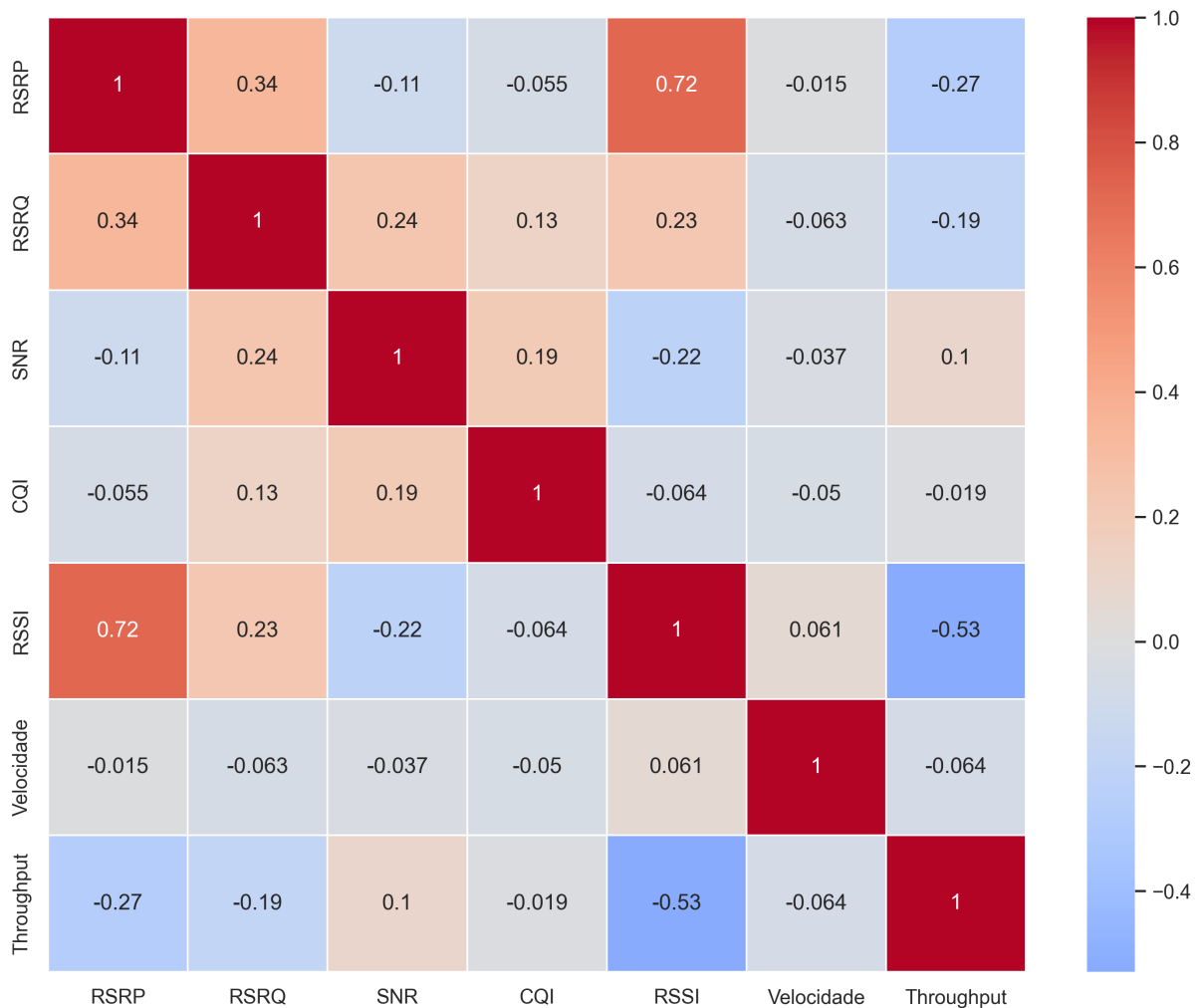
Fonte: Autoria própria.

particularmente relevante, pois indica que a maior parte das séries já se encontra em condições adequadas para aplicação de modelos estatísticos e de aprendizado de máquina sem a necessidade de transformações adicionais (HYNDMAN; ATHANASOPOULOS, 2021). Ademais, a predominância de estacionariedade sugere que as variações de curto prazo refletem de forma mais pronunciada as flutuações do ambiente de rede, ao passo que tendências estruturais de longo prazo exercem influência menos significativa nesse conjunto de dados.

4.2 Pré-processamento

A Figura 23 apresenta a proporção de valores ausentes em cada métrica do conjunto de dados analisado. Observa-se que o RSSI concentra a maior parcela de falhas de regis-

Figura 20 – Matriz de correlação de Pearson calculada apenas para as medições em condição estática.

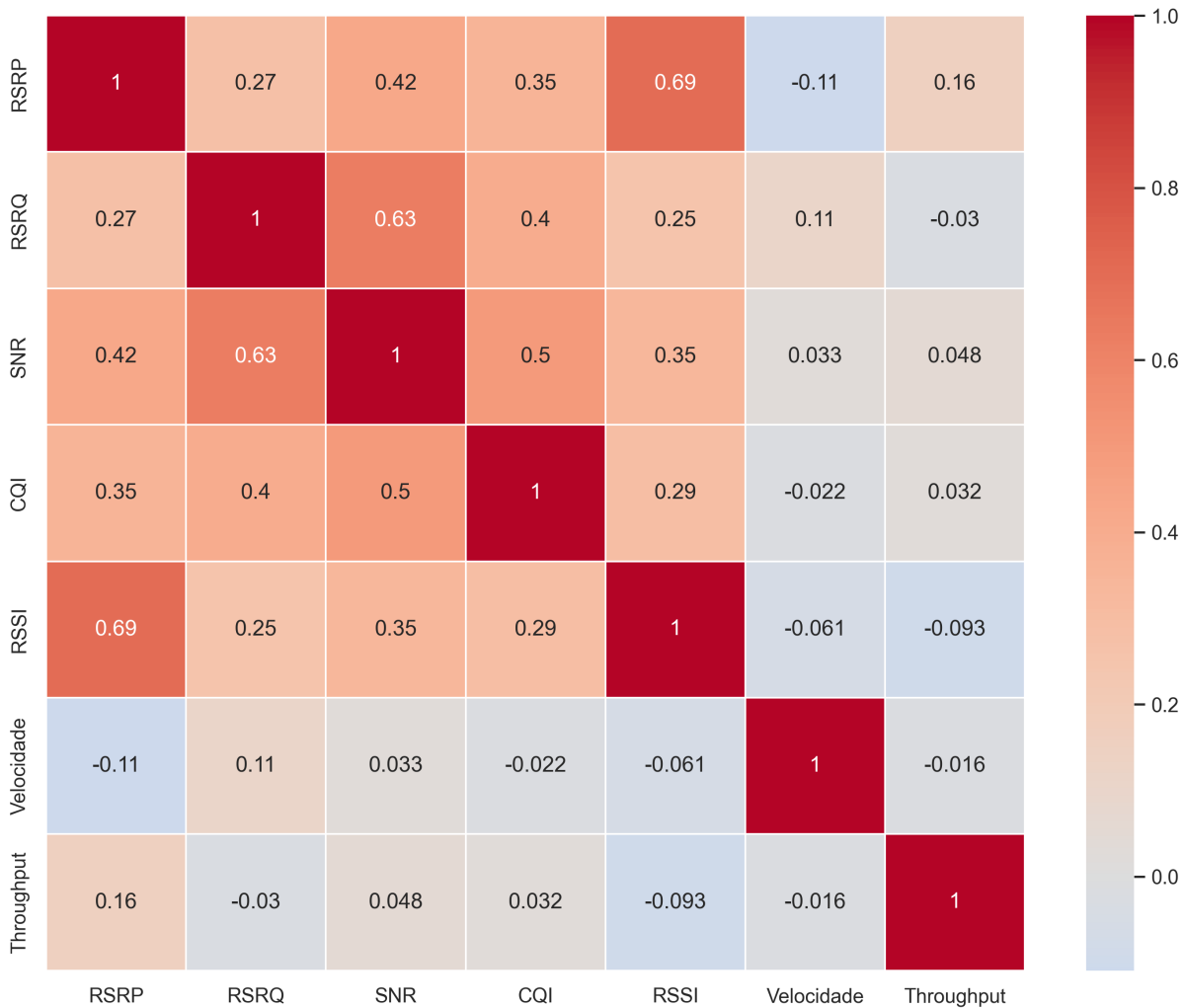


Fonte: Autoria própria.

tro, correspondendo a aproximadamente 28,8% do total de observações, seguido por CQI (8,9%) e SNR (8,8%). Já o RSRQ apresenta apenas 1,8% de valores ausentes, enquanto o RSRP, o *throughput* e a velocidade da UE não registraram ausência de dados. Esses resultados evidenciam que as lacunas estão concentradas em métricas relacionadas à intensidade e qualidade do sinal, reforçando a necessidade de técnicas de imputação para assegurar consistência nas análises subsequentes.

O conjunto de dados foi submetido a um fluxo padronizado de pré-processamento, visando sua adequação à modelagem de séries temporais e à aplicação de modelos estatísticos e de aprendizado de máquina. Inicialmente, aplicou-se a re-amostragem em intervalos de um segundo para garantir uniformidade na granularidade temporal, seguida da remoção de registros duplicados. Em seguida, os valores ausentes foram tratados exclusivamente por interpolação linear, assegurando a continuidade das séries temporais.

Figura 21 – Matriz de correlação de Pearson calculada apenas para as medições em condição de deslocamento.



Fonte: Autoria própria.

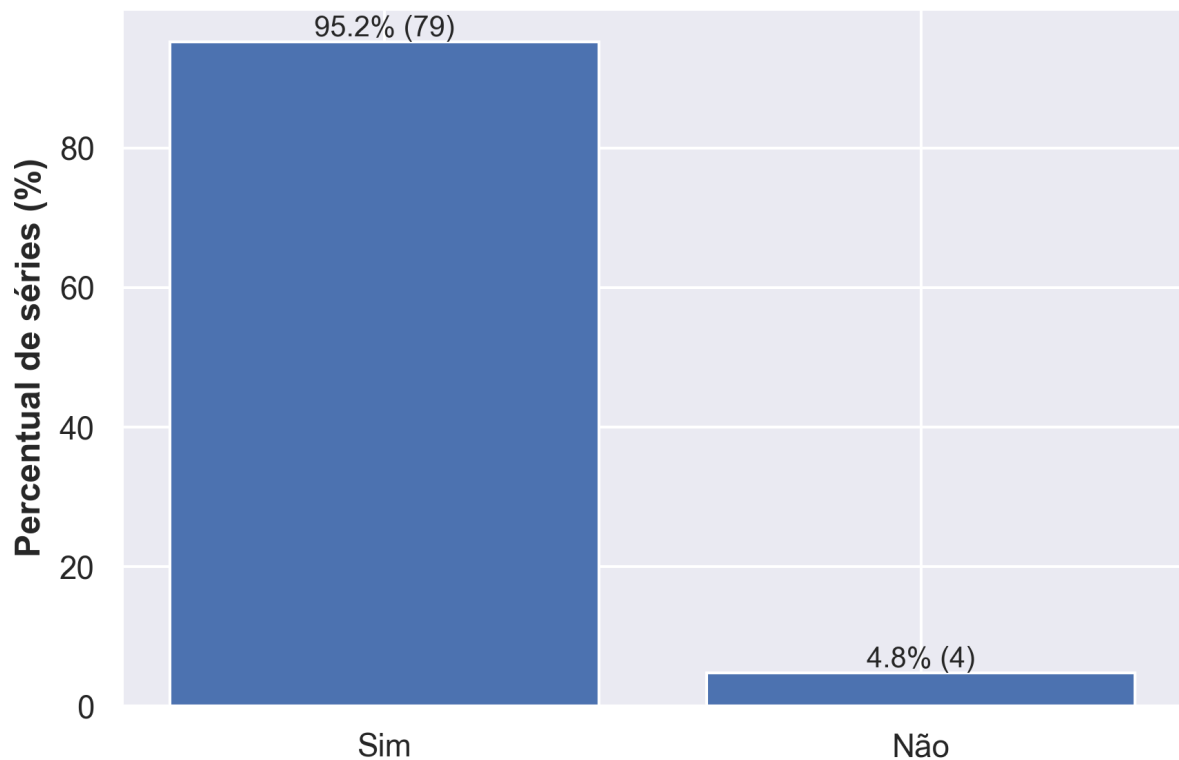
Por fim, todas as métricas foram escalonadas pela normalização *min-max* (Equação 18), em conformidade com práticas consolidadas na literatura (BOX et al., 2015; HYNDMAN; ATHANASOPOULOS, 2021; BATOOL; FOUDA; FADLULLAH, 2024).

A interpolação aplicada adota o método linear, no qual os valores faltantes são estimados a partir da relação proporcional entre pontos adjacentes em uma série ordenada. Seja (τ_0, v_0) e (τ_1, v_1) dois pontos consecutivos observados e τ um instante entre τ_0 e τ_1 em que v é desconhecido. O valor interpolado é obtido por:

$$v = v_0 + \frac{(\tau - \tau_0)}{(\tau_1 - \tau_0)} \cdot (v_1 - v_0) \quad (19)$$

Esse procedimento preserva a coerência local e evita descontinuidades abruptas, sendo mais consistente com a dinâmica temporal imediata quando comparado a imputações

Figura 22 – Proporção de séries temporais estacionárias e não estacionárias, considerando os resultados do teste de ADF aplicado ao *throughput*.



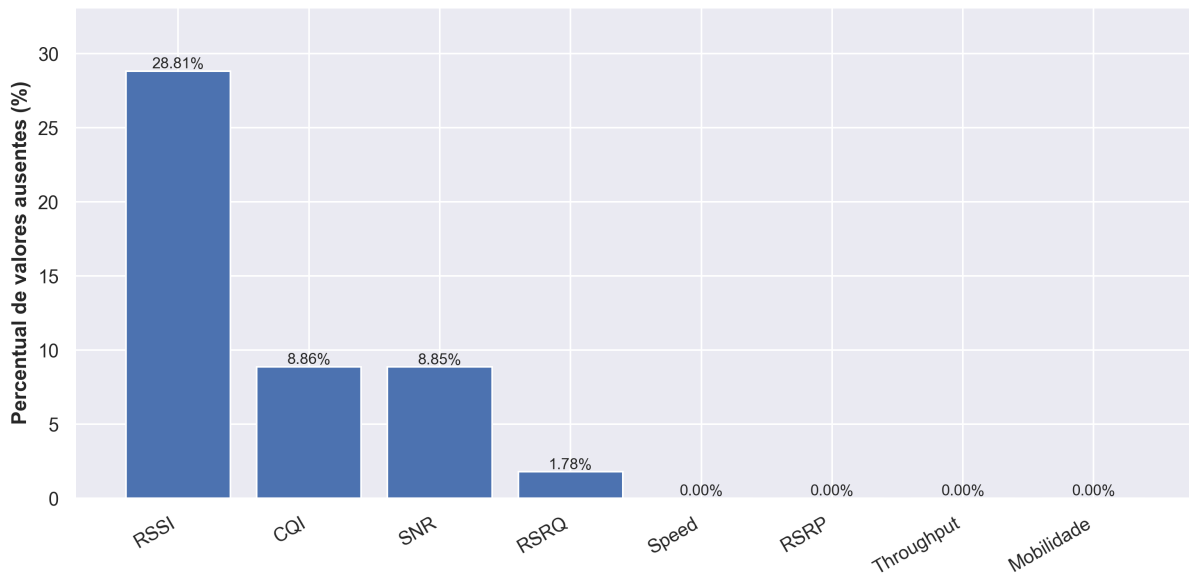
Fonte: Autoria própria.

por medidas globais de tendência central (HYNDMAN; ATHANASOPOULOS, 2021). É importante destacar que apenas as covariadas apresentaram valores ausentes. Dessa forma, a interpolação não impacta diretamente a variável alvo (*throughput*), mas pode beneficiar as covariadas, constituindo uma limitação metodológica a ser considerada na interpretação dos resultados (SHARMA; PANDIT; TALLURI, 2025).

Além disso, em lugar do uso de uma variável binária para distinguir cenários estáticos e de mobilidade, este estudo utilizou a velocidade da UE como covariada contínua. Essa decisão permite capturar diferentes níveis de deslocamento de forma mais realista, evitando reduções dicotômicas e oferecendo maior riqueza descritiva sobre o impacto da mobilidade no *throughput*.

Por fim, diferentemente de trabalhos que incorporam variáveis contextuais, como coordenadas geográficas ou identificadores de operadora (NARAYANAN et al., 2020; BATOOL; FOUUDA; FADLULLAH, 2024), este estudo optou por não empregá-las. Essa decisão visa reduzir vieses geográficos e ampliar a capacidade de generalização dos modelos, concentrando a previsão exclusivamente nas métricas de qualidade do sinal, o que resulta em entradas mais limpas, temporalmente consistentes e adequadas para modelos de previsão tanto locais quanto globais.

Figura 23 – Proporção de valores ausentes por métrica do conjunto de dados.



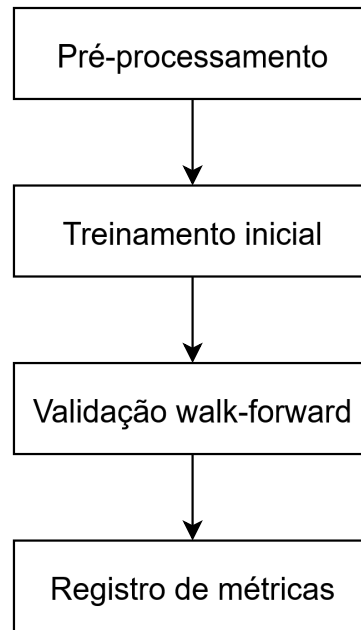
Fonte: Autoria própria.

4.3 Protocolo experimental

O protocolo experimental adotado nesta dissertação está representado de forma geral na Figura 24. Para avaliar a capacidade de generalização dos modelos em condições realistas, foi utilizada a estratégia de validação de avanço em janelas (*walk-forward*), ou seja, origem rolante, que preserva a dependência temporal inerente às séries. Cada série temporal foi considerada de forma independente, sem concatenação entre diferentes sequências, sendo dividida em 80% para treinamento e 20% para validação. Nesta fase, as previsões foram geradas iterativamente em um horizonte de 10 passos à frente, com a janela de validação avançando sequencialmente para avaliar o desempenho em todos os pontos possíveis do conjunto reservado.

Há uma distinção fundamental entre modelos locais e globais, conforme ilustrado na Figura 25 e sistematizado na Tabela 14. Os modelos locais, em sua maioria de natureza estatística, apresentam a limitação de exigir o treinamento de um modelo independente para cada série temporal, sem qualquer compartilhamento de informação entre elas. No contexto da validação *walk-forward*, esses modelos precisam ser reestimados a cada passo, o que resulta na geração de múltiplos modelos por série. Embora tal estratégia possibilite maior adaptação às particularidades de cada sequência, ela eleva a complexidade computacional e amplia a suscetibilidade a variações abruptas nos dados (*concept drift*). Em contraste, os modelos globais são treinados uma única vez e aplicados de forma consistente a todas as janelas de validação, partindo de um modelo capaz de capturar padrões latentes comuns a diferentes séries e, em muitos casos, de incorporar covariadas passadas,

Figura 24 – Protocolo experimental.



Fonte: Autoria própria.

o que tende a conferir maior robustez frente a mudanças localizadas.

Tabela 14 – Modelos de previsão indicando suporte a covariadas passadas e escopo (local ou global).

Categoria	Modelo	Covariadas / Escopo
<i>Naive</i> (Locais)	<i>Naive Mean</i>	Não / Local
	<i>Drift</i>	Não / Local
	<i>Seasonal</i>	Não / Local
<i>Naive</i> (Global)	<i>Aggregate</i>	Não / Global
Simples	<i>Exponential Smoothing</i>	Não / Local
Avançados	AutoARIMA	Não / Local
<i>Baselines/Ensemble</i>	<i>Linear Regression</i>	Sim / Global
	<i>Random Forest</i>	Sim / Global
	<i>LightGBM</i>	Sim / Global
<i>Deep Learning</i>	<i>BlockRNN</i>	Sim / Global
	NBEATS	Sim / Global
	<i>Transformer</i>	Sim / Global

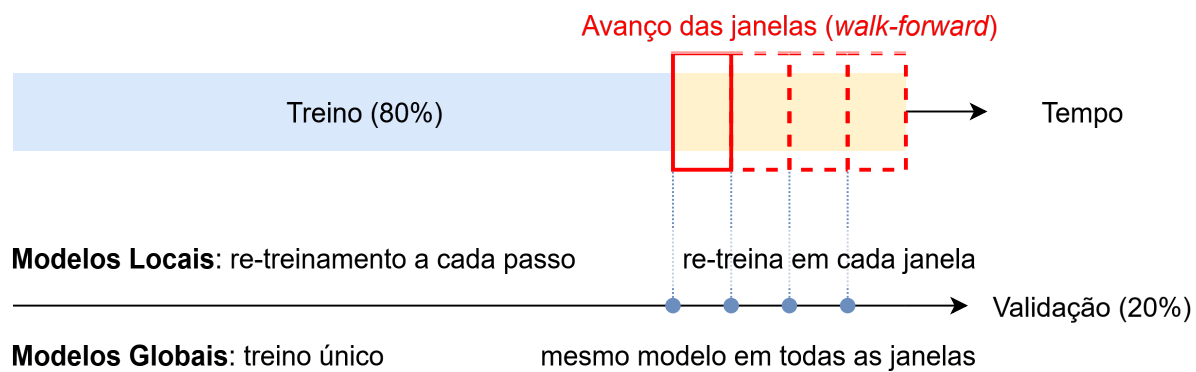
Fonte: Autoria própria.

Além disso, para os modelos globais que oferecem suporte ao uso de covariadas passadas¹, foram conduzidos experimentos em dois cenários distintos: (i) utilizando apenas

¹ No contexto de previsão de séries temporais, covariadas passadas (ou observadas) correspondem a variáveis exógenas disponíveis até o instante presente, registradas no mesmo período da série alvo ou em períodos anteriores. Já as covariadas futuras (ou conhecidas) referem-se a variáveis cujo valor é previamente determinado e acessível no horizonte de previsão, como datas de calendário, feriados ou eventos programados (HYNDMAN; ATHANASOPOULOS, 2021).

a série alvo e (ii) incorporando as covariadas disponíveis. Esse procedimento tem como objetivo avaliar de forma sistemática o impacto das covariadas no desempenho preditivo, permitindo identificar em que medida tais informações adicionais contribuem para a melhoria das previsões.

Figura 25 – Validação *walk-forward*: divisão treino/validação, avanço das janelas e diferença de fluxo entre modelos locais (re-treinamento) e globais (treino único).



Fonte: Autoria própria.

A adoção da validação *walk-forward* foi inspirada no trabalho de Hewamalage, Ackermann e Bergmeir (2023), uma vez que, na literatura, muitos estudos recorrem a práticas simplificadas de avaliação em séries temporais, como divisões fixas de treino-teste, que tendem a produzir resultados excessivamente otimistas em cenários sujeitos a dependências, tendências ou sazonalidades variáveis. De acordo com Hewamalage, Ackermann e Bergmeir (2023), métodos baseados em múltiplas janelas, como o *walk-forward*, permitem aferir o desempenho dos modelos em diferentes pontos do tempo, revelando variações no erro e instabilidades que não seriam observadas em avaliações pontuais, ainda que impliquem maior custo computacional e exijam reestimativas frequentes.

Os experimentos foram conduzidos em uma estação de trabalho dedicada, equipada com processador AMD Ryzen 7 5800X, 62 GB de memória RAM e GPU NVIDIA GeForce RTX 3070 (8 GB), operando em ambiente Linux (kernel versão 6.12.10). A GPU foi utilizada para acelerar o treinamento dos modelos baseados em aprendizado profundo, enquanto os métodos estatísticos e de *ensemble* foram processados na CPU. Para assegurar consistência e reprodutibilidade, foram mantidos os hiperparâmetros padrão disponibilizados pela biblioteca Darts², com exceções documentadas na Tabela 15.

² Biblioteca Darts, disponível em <<https://unit8co.github.io/darts/>>.

Tabela 15 – Parâmetros configurados para os modelos de previsão.

Categoria	Modelo	Hiperparâmetros
Comuns a todos	-	<code>predict_horizon=10,</code> <code>input_chunk_length=10,</code> <code>output_chunk_length=10</code>
<i>Naive</i> (Locais)	<i>Naive Mean</i>	Sem parâmetros adicionais
<i>Naive</i> (Globais)	<i>Aggregate</i>	<code>agg_fn="mean"</code> Demais parâmetros iguais aos comuns.
	<i>Drift</i>	Apenas parâmetros comuns.
	<i>Seasonal</i>	Apenas parâmetros comuns.
Simple	<i>Exponential Smoothing</i>	<code>trend=None</code> <code>damped=False</code> <code>seasonal=None</code> <code>seasonal_periods=None</code>
Avançados	AutoARIMA	<code>start_p=0, start_q=0</code> <code>max_p=5, max_q=5</code> <code>d=None</code> <code>seasonal=False, stepwise=True</code>
<i>Baselines/Ensemble</i>	<i>Linear Regression</i>	<code>lags=10, lags_past_covariates=10</code>
	<i>Random Forest</i>	Idem ao <i>Linear Regression</i> .
	<i>LightGBM</i>	Idem ao <i>Linear Regression</i> .
<i>Deep Learning</i>	<i>BlockRNN</i>	<code>batch_size=32, n_epochs=100</code>
	NBEATS	Idem ao <i>BlockRNN</i> .
	<i>Transformer</i>	Idem ao <i>BlockRNN</i> .

Fonte: Autoria própria.

4.4 Avaliação e registro de resultados

A etapa final do protocolo metodológico consistiu na avaliação do desempenho preditivo dos modelos. Para isso, foram empregadas duas métricas amplamente utilizadas em tarefas de previsão de séries temporais: o MAE (Equação 6) e o RMSE (Equação 5).

Ao todo, foram conduzidos aproximadamente 1.400 experimentos, resultantes da combinação entre o conjunto de séries temporais avaliadas, os diferentes métodos de previsão e o protocolo de validação adotado. Entre os modelos considerados, seis apresentaram suporte ao uso de covariadas, o que possibilitou a realização de cenários adicionais de teste e ampliou a abrangência da análise. Todos os resultados — incluindo previsões geradas, valores observados, erros correspondentes e tempos de execução — foram registrados de maneira sistemática. Esses registros constituem a base para a análise comparativa apresentada no próximo capítulo, permitindo examinar em detalhe o impacto de diferentes modelos e do uso de covariadas na previsão do *throughput* em redes móveis.

Todo o código desenvolvido encontra-se disponível em repositório público³, garantindo

³ Repositório da dissertação disponível em: <<https://github.com/ejs94/5g-forecasting>>

transparência, reprodutibilidade e reutilização dos resultados.

Capítulo 5

Resultados

5.1 Análise global

Esta seção apresenta os resultados consolidados da avaliação dos modelos de previsão em múltiplos cenários de mobilidade e atividade do usuário. O objetivo é fornecer uma visão abrangente do desempenho preditivo, destacando os padrões gerais de erro e a relação entre acurácia e custo computacional. Inicialmente, é avaliado o equilíbrio entre desempenho e tempo de treinamento dos modelos, complementado pela análise da distribuição global dos erros e pela comparação relativa ao modelo de referência *Naive Mean*. Essa análise global estabelece a base para as seções subsequentes, que aprofundam a investigação em cenários específicos.

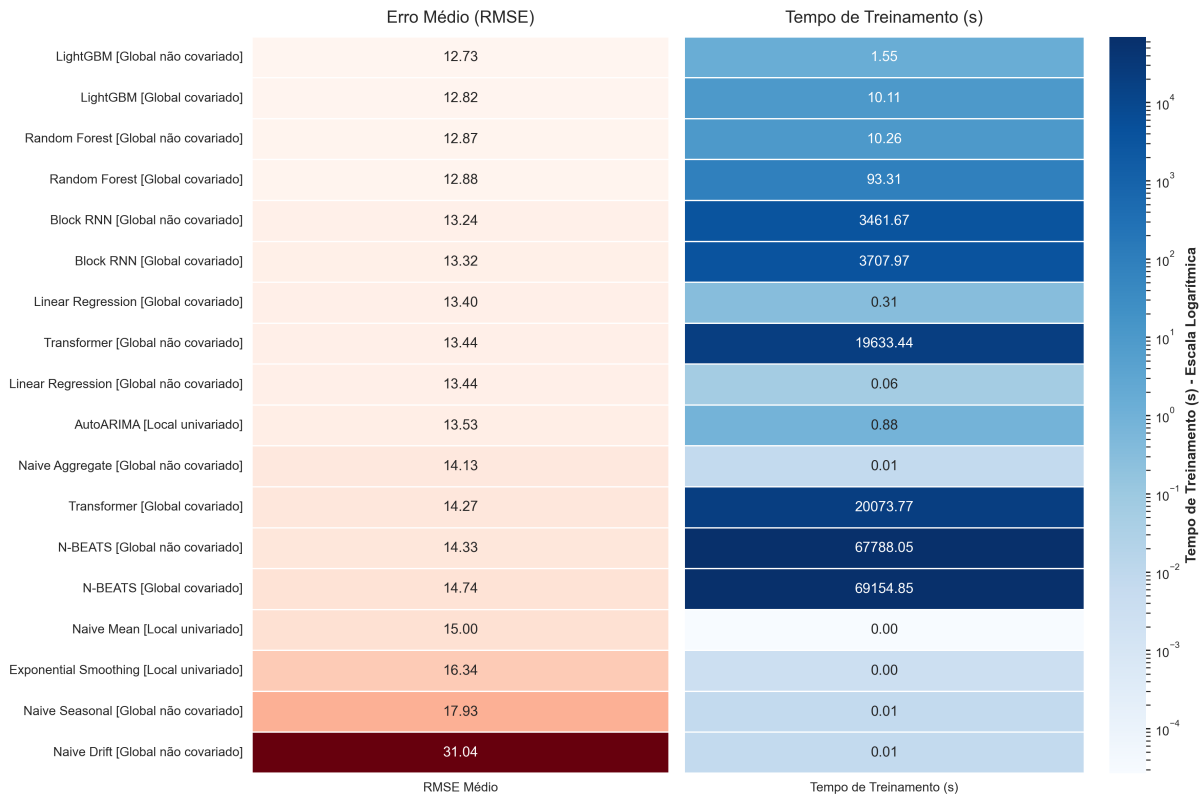
5.1.1 Equilíbrio entre acurácia e custo computacional

A relação entre erro médio (RMSE) e custo computacional de treinamento, apresentada na Figura 26, evidencia diferenças marcantes entre as abordagens. Modelos baseados em *ensemble* (LightGBM e *Random Forest*) e regressão linear destacam-se pelo equilíbrio, combinando baixos valores de erro com tempos de treinamento reduzidos. Essa característica reforça sua adequação a sistemas que demandam reestimativas frequentes dos parâmetros de mobilidade.

Já os modelos de aprendizado profundo (*Block RNN*, *Transformer* e especialmente o *N-BEATS*) alcançam desempenho médio semelhante ao dos modelos clássicos, mas a um custo de duas a quatro ordens de magnitude superior em tempo de processamento. Embora capturem dependências temporais mais complexas, esse desbalanceamento entre acurácia e custo limita sua aplicabilidade em cenários que exigem adaptações rápidas ou

em tempo quase real.

Figura 26 – Análise de desempenho dos modelos de previsão, comparando o RMSE médio com o tempo de treinamento.



Fonte: Autoria própria.

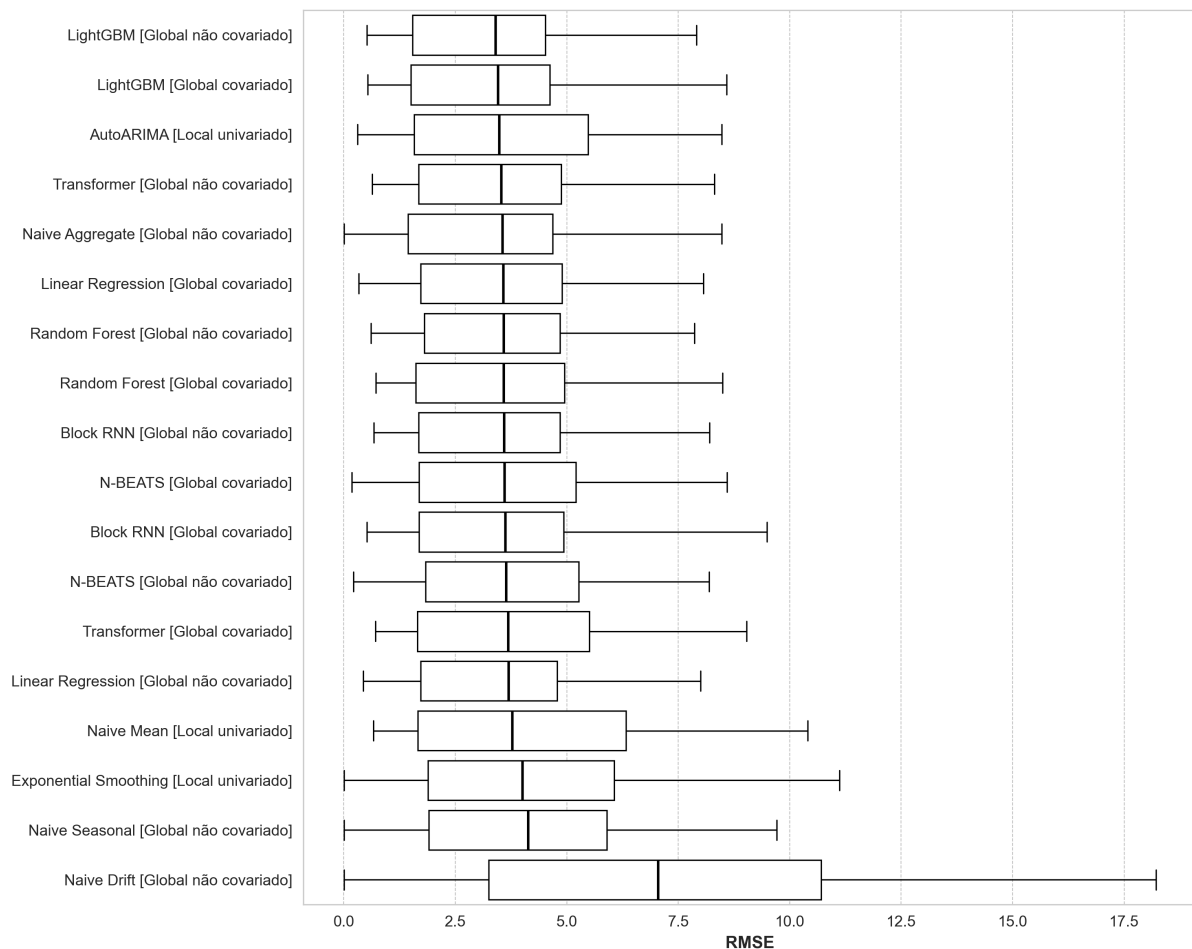
5.1.2 Distribuição global dos erros

Os modelos ingênuos e os métodos estatísticos simples foram utilizados como linha de base. Apesar do custo computacional praticamente nulo, especialmente no caso do *Naive Drift*, os erros médios obtidos foram substancialmente mais elevados, com RMSE médio superior a 15 Mbps e distribuição caracterizada por caudas longas, conforme ilustrado na Figura 27. Além disso, esses modelos apresentam limitações estruturais relevantes, pois exigem ajuste individualizado para cada série e não permitem o compartilhamento de conhecimento entre sequências, o que restringe sua escalabilidade em cenários com grandes volumes de dados.

A análise dos resultados nas Figura 26 e Figura 27 confirma a existência de um *trade-off* entre acurácia e custo computacional. Enquanto os modelos de aprendizado profundo oferecem maior capacidade de representação temporal, essa vantagem é contrabalançada pelo custo elevado de treinamento. Em contraste, os modelos de *ensemble* e de regressão

surtem como alternativas mais equilibradas, conciliando desempenho competitivo com viabilidade prática em ambientes de mobilidade.

Figura 27 – Distribuição do erro de previsão dos modelos, avaliada pela métrica RMSE. O boxplot apresenta os resultados consolidados para todos os cenários analisados, englobando séries estáticas e em mobilidade, permitindo a comparação global do desempenho entre diferentes tipos de modelos.



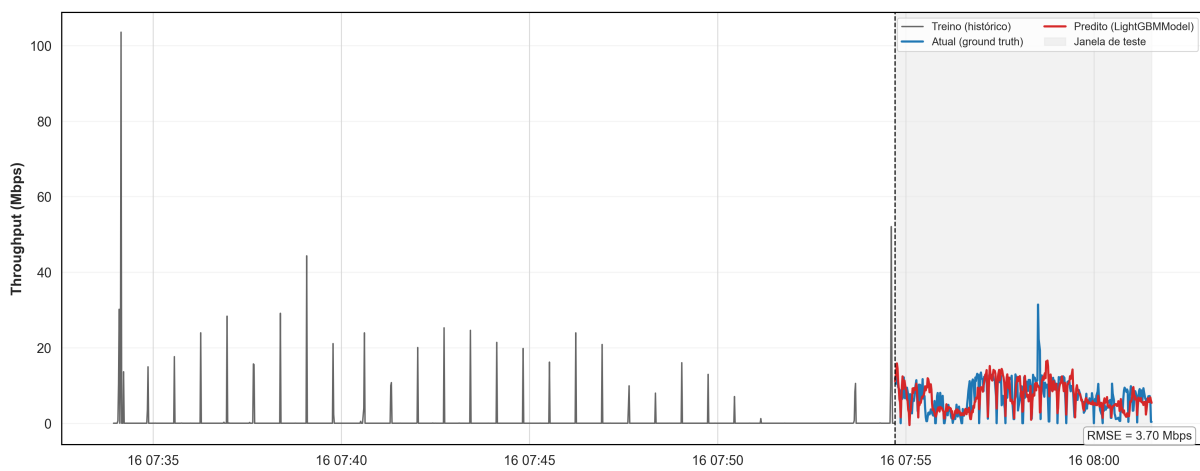
Fonte: Autoria própria.

5.1.3 Comparação entre o *LightGBM* e o *Naive Mean*

A análise individual de séries específicas evidencia tanto os benefícios quanto as limitações dos modelos em cenários contrastantes. Para isso, foram selecionadas duas séries representativas: a série 75, na qual o *LightGBM* sem covariáveis apresentou o melhor resultado entre todos os modelos, e a série 50, em que o modelo *Naive Mean* superou o desempenho do *LightGBM*. Essas duas séries permitem contrastar situações em que modelos sofisticados proporcionam ganhos substanciais em acurácia com outras em que abordagens simples se mostram mais eficazes.

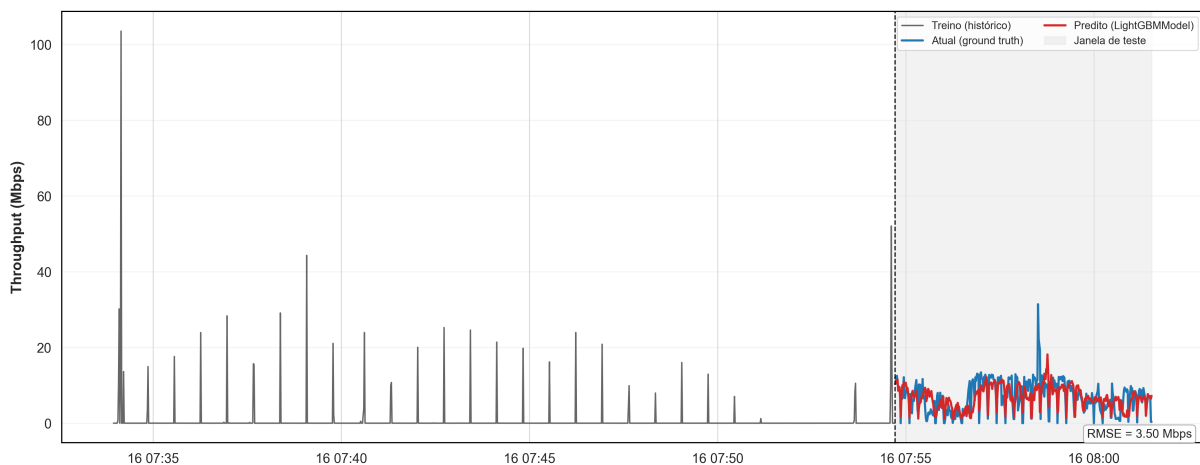
A Figura 28 apresenta a previsão da série 75 no cenário estático com transmissão de vídeo, utilizando o modelo *LightGBM* com covariáveis. Nota-se boa aderência entre valores previstos e observados, refletida em RMSE melhor ao do modelo *baseline*. Já na Figura 29, observa-se que a remoção das covariáveis, em consonância com o resultado global mostrado na Figura 27, proporciona maior precisão, com RMSE de 3,50 contra 3,70 do modelo com covariáveis. Por fim, a Figura 30 mostra a mesma série prevista pelo modelo *Naive Mean*, cujo desempenho é claramente inferior ao do *LightGBM*, reforçando a relevância de modelos mais sofisticados nesse cenário.

Figura 28 – Previsão da série 75 no cenário estático com transmissão de vídeo, utilizando o modelo *LightGBM* com covariáveis.



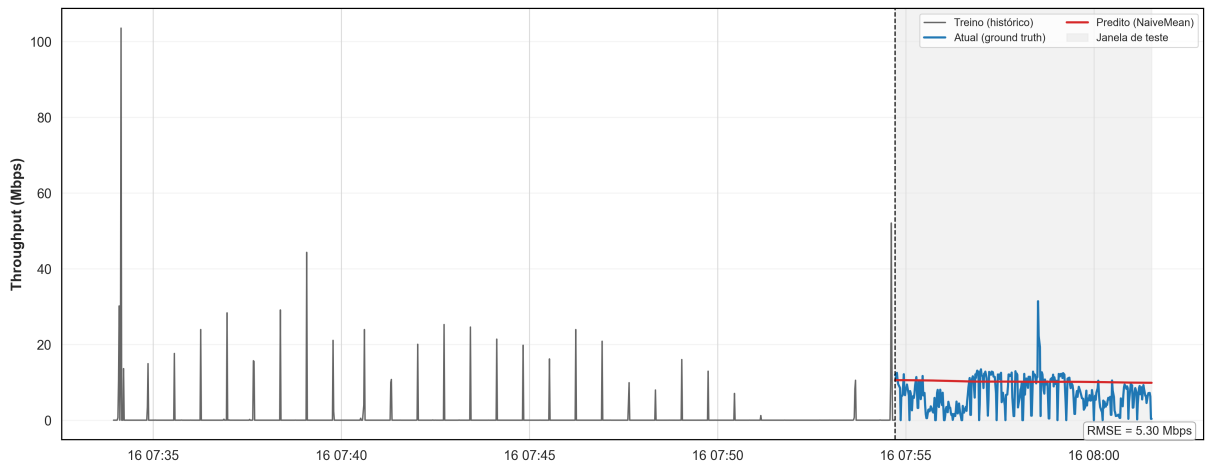
Fonte: Autoria própria.

Figura 29 – Previsão da série 75 no cenário estático com transmissão de vídeo, utilizando o modelo *LightGBM* sem covariáveis.



Fonte: Autoria própria.

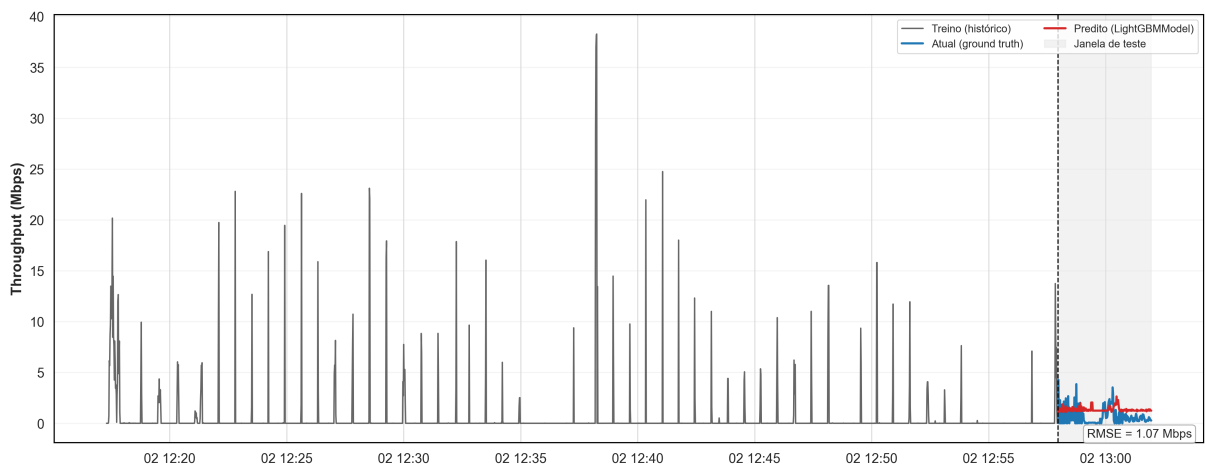
Figura 30 – Previsão da série 75 no cenário estático com transmissão de vídeo, utilizando o modelo *Naive Mean* local.



Fonte: Autoria própria.

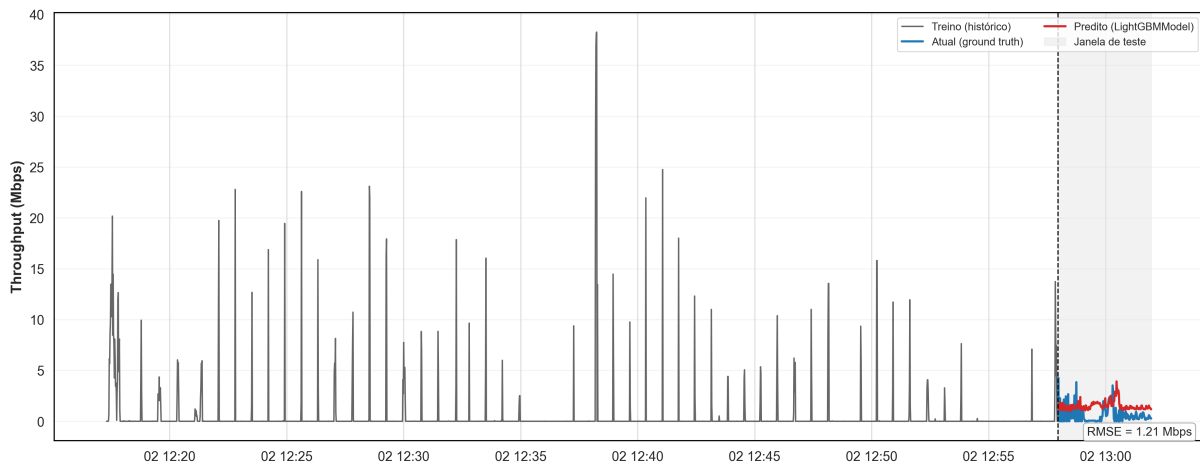
A série 50, correspondente ao cenário de condução com transmissão de vídeo, ilustra um comportamento oposto. Na Figura 31, observa-se que, embora o *LightGBM* com covariáveis capture parte das variações, o desempenho é prejudicado pela elevada irregularidade do tráfego. A Figura 32 mostra que a exclusão das covariáveis não resulta em melhora. Por outro lado, a Figura 33 revela que o *Naive Mean*, de forma contraintuitiva, alcançou menor RMSE que o *LightGBM* (com e sem covariáveis), evidenciando que, em cenários de curta predição e tráfego relativamente estável, abordagens simples ainda podem superar modelos mais sofisticados.

Figura 31 – Previsão da série 50 no cenário de condução com transmissão de vídeo, utilizando o modelo *LightGBM* com covariáveis.



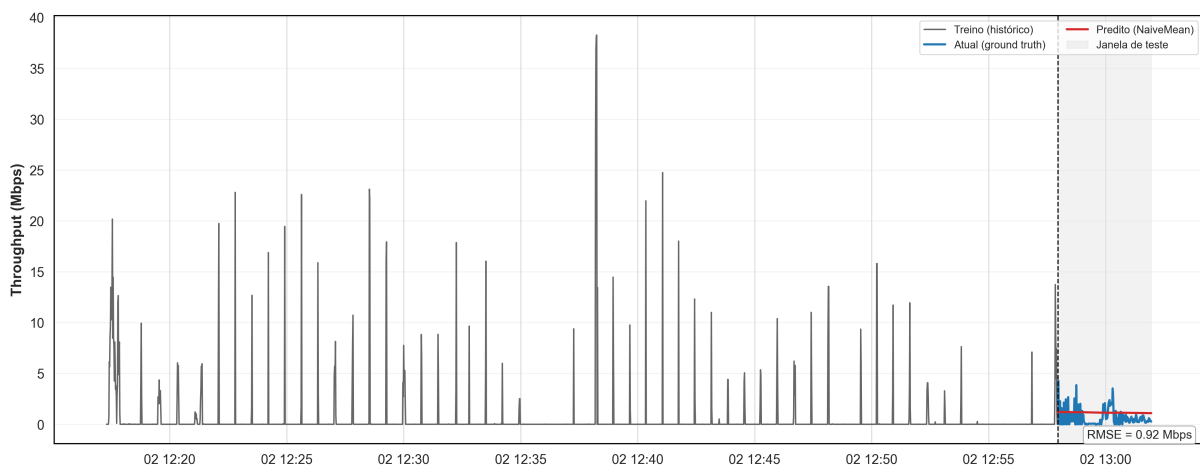
Fonte: Autoria própria.

Figura 32 – Previsão da série 50 no cenário de condução com transmissão de vídeo, utilizando o modelo *LightGBM* sem covariáveis.



Fonte: Autoria própria.

Figura 33 – Previsão da série 50 no cenário de condução com transmissão de vídeo, utilizando o modelo *Naive Mean* local.



Fonte: Autoria própria.

Em síntese, a comparação entre as séries 75 e 50 evidencia a influência do contexto de mobilidade no desempenho dos modelos. Em cenários estáticos e mais regulares, como na série 75, a remoção das covariáveis resulta em melhor desempenho, sugerindo que sua utilização pode introduzir ruídos na captura de padrões estáveis. Já em cenários de condução, as covariáveis aparentam ter maior relevância para reduzir erros de previsão; contudo, em condições de baixa taxa de *throughput* e poucos picos de variação, como na série 50, métodos simples como o *Naive Mean* podem superar modelos sofisticados. Esse contraste reforça que a eficácia dos modelos não depende apenas de sua complexidade,

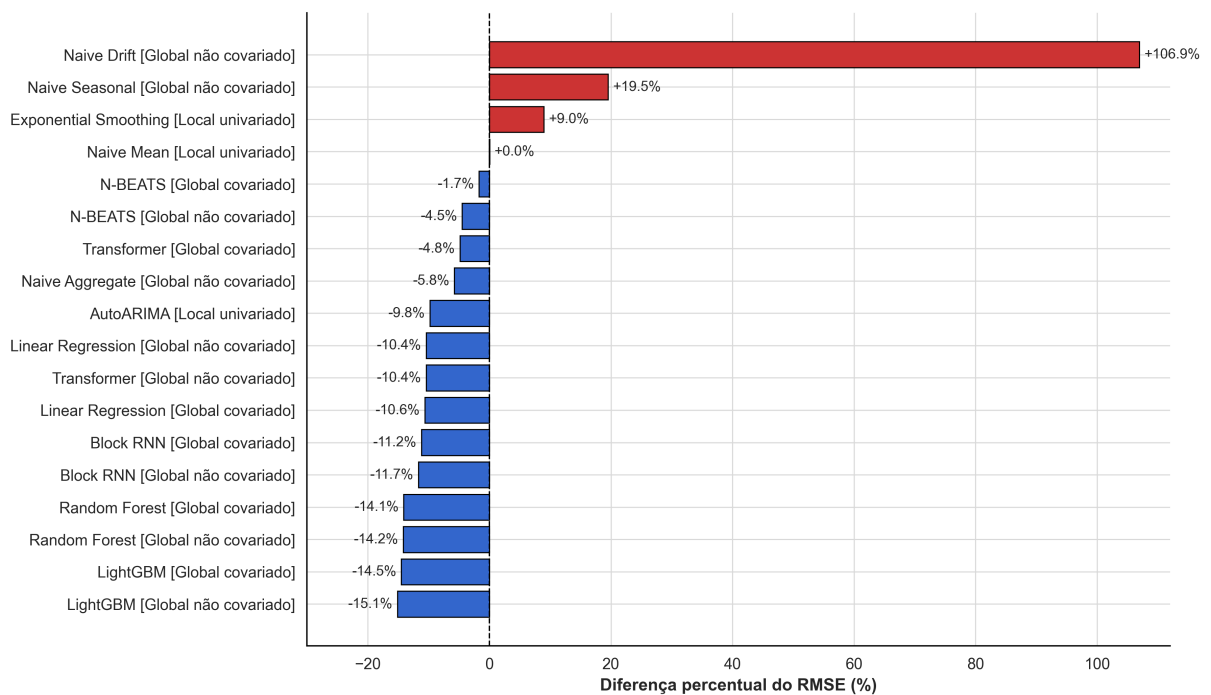
mas também da natureza do tráfego e das condições de mobilidade subjacentes, como discutido na literatura.

5.1.4 Comparando todos os modelos ao *baseline*

Por fim, a comparação global em relação ao *Naive Mean*, apresentada na Figura 34, reforça a vantagem dos modelos de aprendizado de máquina em relação às abordagens ingênuas e estatísticas simples. Modelos baseados em *ensemble*, como *LightGBM* e *Random Forest*, obtiveram reduções de até 15% no RMSE, enquanto arquiteturas de aprendizado profundo, como *Block RNN* e *Transformer*, alcançaram ganhos entre 10% e 12%. Já os métodos estatísticos (*AutoARIMA* e *Exponential Smoothing*) apresentaram desempenho semelhante, com variações inferiores a 10%. Em contraste, os modelos *Naive Drift* e *Naive Seasonal* registraram degradações expressivas, com erros até 107% superiores.

Essa abordagem, ao utilizar um método ingênuo de médias como referência, possibilita avaliar a eficácia relativa de cada modelo, e também permite comparar o desempenho entre diferentes séries de forma padronizada. Assim, consolida-se uma base comum para analisar em que medida modelos mais sofisticados superam soluções triviais, destacando em quais contextos os ganhos de acurácia justificam a complexidade adicional.

Figura 34 – Comparação percentual do desempenho de todos os modelos em relação ao *Naive Mean*, utilizando como referência a métrica RMSE. Valores negativos indicam que o modelo obteve erro médio inferior (melhor desempenho) em comparação ao *baseline*, enquanto valores positivos indicam desempenho pior.



Fonte: Autoria própria.

5.2 Análises por cenários

Após a apresentação dos resultados globais, esta seção aprofunda a análise considerando separadamente os cenários de mobilidade e de atividade do usuário. O objetivo é analisar de que forma as condições de condução, de estado estático, de transmissão de vídeo e de *download* de arquivos impactam o desempenho dos modelos de previsão. Para cada cenário, os resultados são comparados em relação ao *Naive Mean*, utilizando como referência a métrica RMSE. Dessa forma, é possível identificar em quais contextos os modelos sofisticados apresentam ganhos consistentes e em quais situações abordagens simples permanecem competitivas.

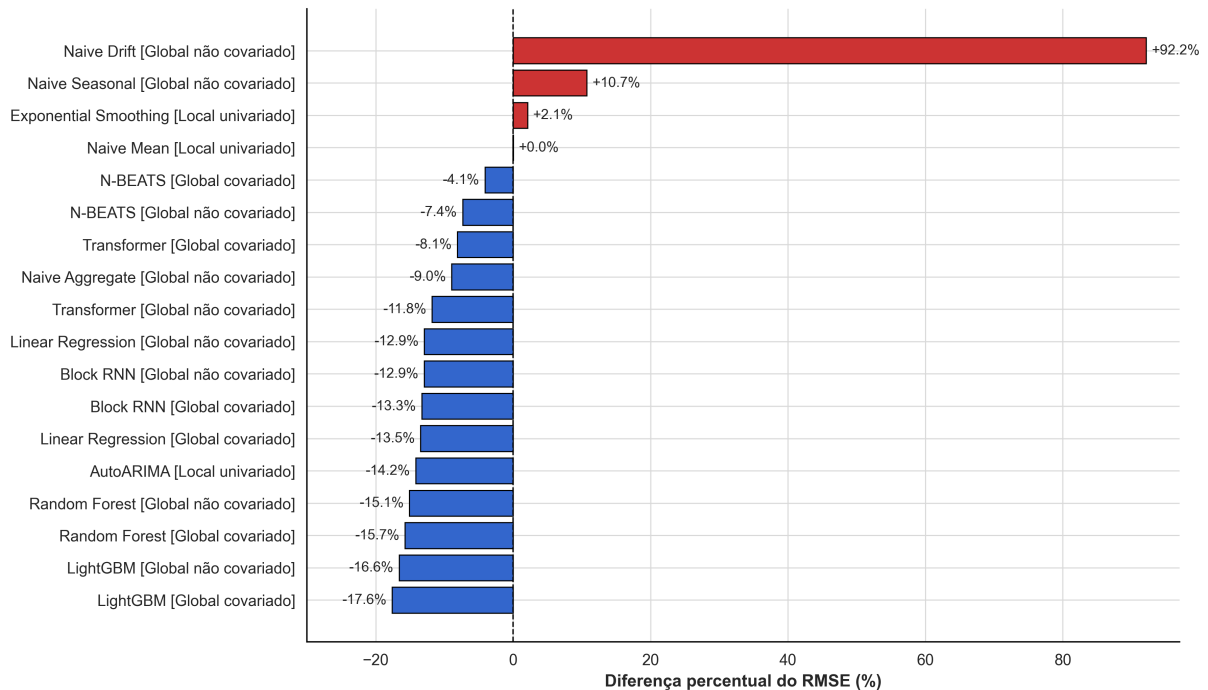
5.2.1 Cenário de condução

No cenário de mobilidade em condução, a maioria dos modelos de aprendizado de máquina e redes neurais supera o *Naive Mean*, com reduções expressivas no RMSE. Destacam-se os modelos baseados em *ensemble* (LightGBM e *Random Forest*) e a regressão linear, que aliam acurácia e baixo custo computacional. Já arquiteturas profundas, como *Block RNN* e *Transformer*, também apresentaram ganhos relevantes, embora demandem maior tempo de treinamento. Em contrapartida, métodos estatísticos simples (*Exponential Smoothing* e *Naive Seasonal*) e, sobretudo, o *Naive Drift* resultaram em piora significativa em relação ao *baseline*, com erros chegando a duplicar. Esses resultados, sintetizados na Figura 35, confirmam o padrão observado na análise global: em cenários dinâmicos, modelos mais sofisticados são capazes de capturar de maneira robusta as variações do tráfego.

5.2.2 Cenário estático

No cenário de mobilidade estática, os resultados na Figura 36 revelam maior heterogeneidade no desempenho dos modelos em comparação ao *Naive Mean*. Enquanto abordagens avançadas, como LightGBM, *Random Forest* e *Block RNN*, alcançaram reduções de até 19% no RMSE, métodos estatísticos simples (*Exponential Smoothing* e *Naive Seasonal*) apresentaram degradações expressivas, com erros de 14% a 18% superiores ao *baseline*. O *Naive Drift*, em especial, registrou erros mais que duplicados em relação à referência. Diferentemente do cenário de condução, aqui a maior variabilidade do tráfego em estado estacionário torna o contraste entre modelos sofisticados e ingênuos ainda mais evidente. Esses achados reforçam, em linha com a análise global, que modelos de aprendizado de máquina e redes neurais oferecem vantagens consistentes, enquanto abordagens ingênuas se mostram inferiores para capturar a complexidade do tráfego.

Figura 35 – Comparação percentual do desempenho dos modelos em relação ao *Naive Mean*, considerando exclusivamente o cenário de mobilidade em condução. A métrica de referência é o RMSE, com valores negativos representando melhora em relação ao *baseline* e positivos indicando pior desempenho.



Fonte: Autoria própria.

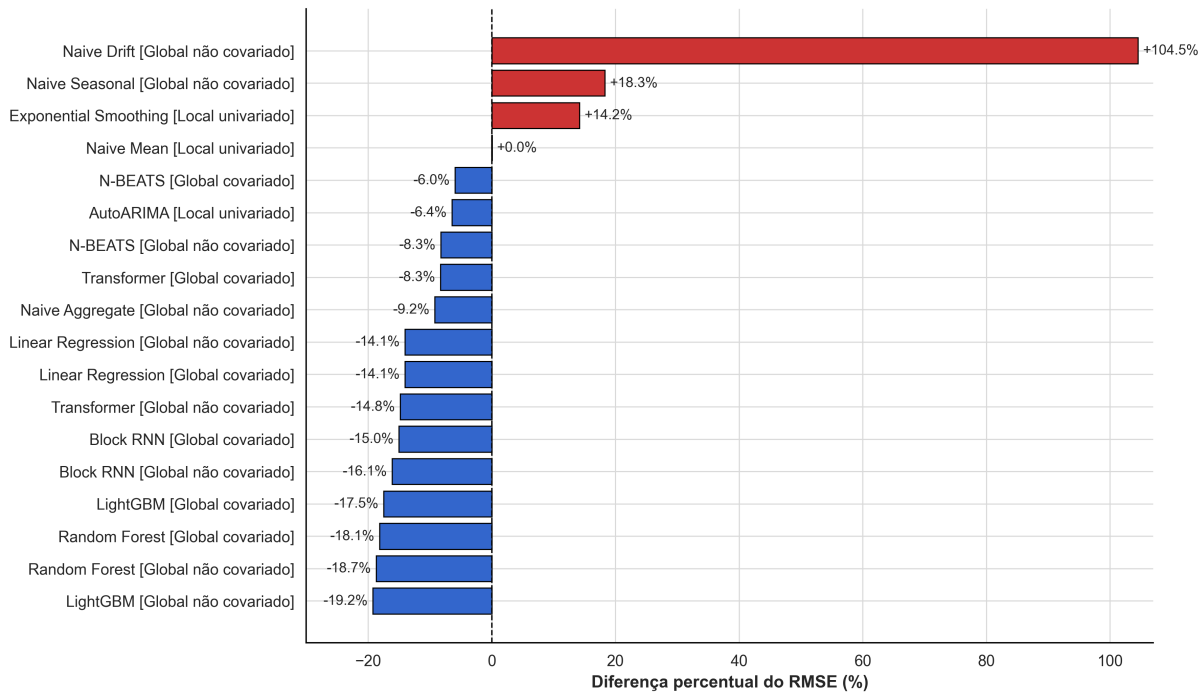
5.2.3 Cenário transmissão de vídeo

No caso da transmissão de vídeo, ilustrado na Figura 37, verifica-se um comportamento mais próximo ao da condução do que ao do *download*. Modelos de aprendizado de máquina e redes neurais, como LightGBM, *Random Forest* e *Block RNN*, alcançaram reduções superiores a 10% no RMSE em relação ao *baseline*. Esse ganho decorre da capacidade desses modelos em lidar com o caráter dinâmico e irregular do tráfego de vídeo, marcado por flutuações rápidas e padrões de difícil generalização. Em contraste, abordagens estatísticas simples, como *Naive Drift* e *Naive Seasonal*, apresentaram degradações significativas. Diferente do cenário estático, aqui os resultados estão bastante alinhados com os padrões da análise global, reforçando que cenários de alta variabilidade favorecem modelos mais sofisticados.

5.2.4 Cenário *download* de arquivo

Por fim, no cenário de *download* de arquivos, os resultados da Figura 38 indicam que os modelos mais sofisticados não conseguiram superar de forma consistente o *Naive Mean*. Com exceção do *AutoARIMA*, que apresentou leve ganho, a maioria dos métodos

Figura 36 – Comparação percentual do desempenho dos modelos em relação ao *Naive Mean*, no cenário de mobilidade estática. A referência é a métrica RMSE; valores negativos indicam modelos mais precisos que o *baseline*, enquanto valores positivos indicam maior erro em relação à referência.



Fonte: Autoria própria.

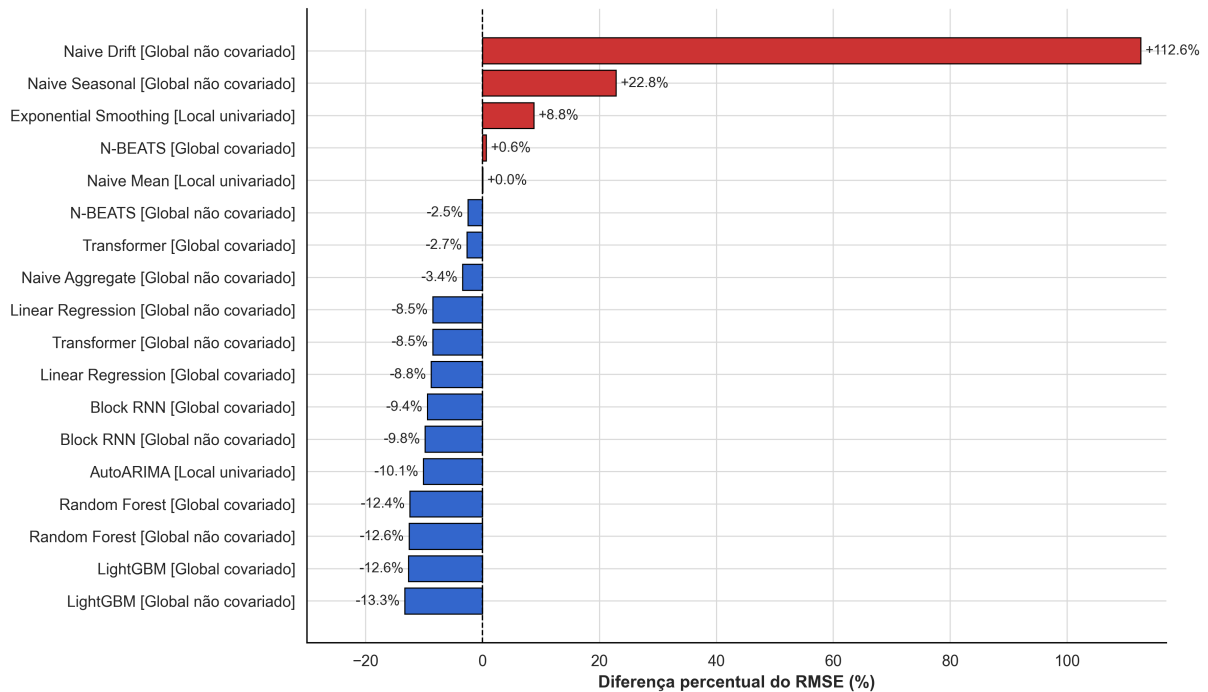
— estatísticos, de aprendizado de máquina e de redes neurais — obteve valores de RMSE superiores ao *baseline*. Esse comportamento contrasta com os demais cenários e com a análise global, pois aqui o perfil mais regular e previsível do tráfego reduz substancialmente a vantagem de modelos complexos. Dessa forma, no caso do *download* de arquivos, a simplicidade de abordagens como o *Naive Mean* se mostra suficiente para capturar as principais variações da série, reforçando que a sofisticação dos modelos só se justifica em contextos mais dinâmicos e irregulares.

5.3 Síntese dos resultados

A análise global e as comparações por cenário permitem sintetizar um conjunto de padrões consistentes que orientam a avaliação do desempenho dos modelos de previsão em redes móveis:

- ❑ **Trade-off entre acurácia e custo computacional:** modelos de aprendizado profundo, como *Block RNN*, *Transformer* e *N-BEATS*, apresentam ganhos moderados em acurácia, mas com custo de treinamento até quatro ordens de magnitude

Figura 37 – Comparação percentual do desempenho dos modelos em relação ao *Naive Mean*, considerando exclusivamente o cenário de transmissão de vídeo. A métrica utilizada é o RMSE, onde valores negativos indicam melhora em relação ao *baseline* e positivos representam pior desempenho.

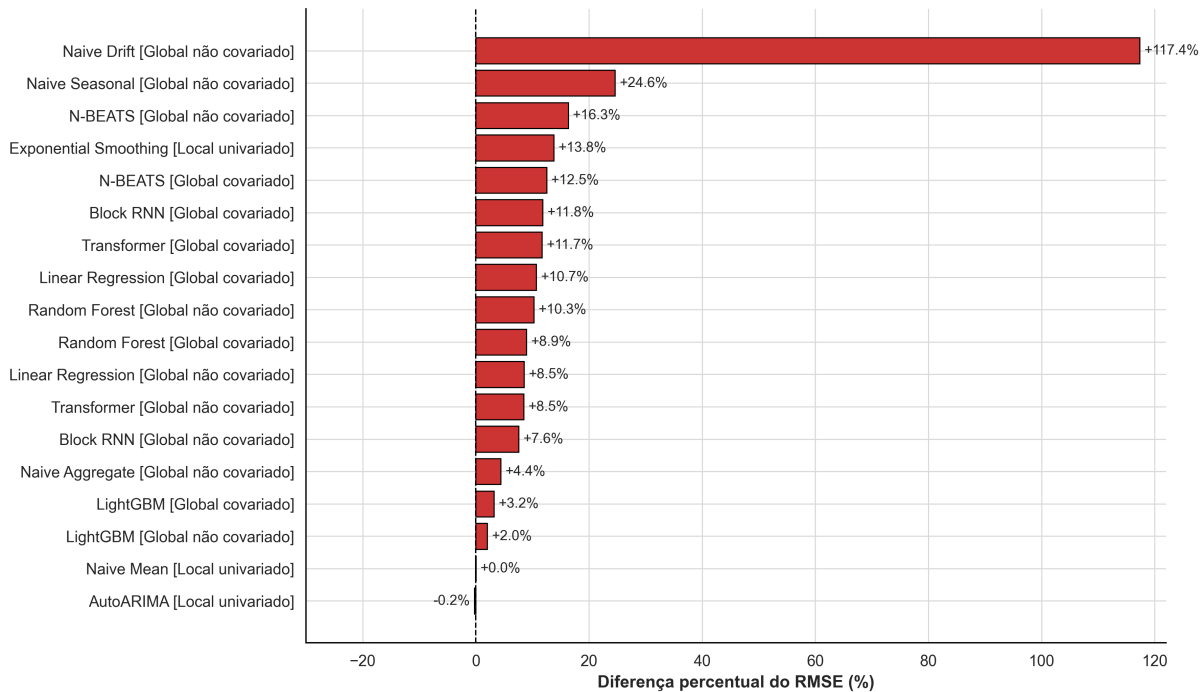


Fonte: Autoria própria.

superior, o que limita sua aplicabilidade em contextos que exigem reconfiguração frequente ou predição em tempo quase real.

- ❑ **Modelos de *ensemble* e regressão linear como equilíbrio:** LightGBM, *Random Forest* e regressão linear se destacam pela combinação entre baixo erro médio e custo computacional reduzido, mostrando-se especialmente adequados para sistemas adaptativos de mobilidade que demandam atualizações contínuas.
- ❑ **Limitações das abordagens ingênuas e estatísticas simples:** métodos como *Naive Drift*, *Naive Seasonal* e *Exponential Smoothing* mostraram-se inadequados, apresentando degradações significativas em praticamente todos os cenários. Apesar da baixa complexidade, sua incapacidade de capturar variações temporais mais complexas e de compartilhar conhecimento entre séries restringe fortemente sua utilidade.
- ❑ **Influência do contexto de mobilidade:** em cenários de condução e transmissão de vídeo, marcados por maior variabilidade, os modelos sofisticados demonstram vantagens claras sobre o *baseline*. Já em cenários estáticos, a remoção das covariáveis pode melhorar os resultados, sugerindo que essas informações introduzem ruídos em

Figura 38 – Comparação percentual do desempenho dos modelos em relação ao *Naive Mean*, considerando exclusivamente o cenário de download de arquivos. A métrica utilizada é o RMSE, com valores negativos representando ganho em relação ao *baseline* e positivos indicando perda de desempenho.



Fonte: Autoria própria.

condições mais regulares. No caso do *download* de arquivos, o perfil previsível do tráfego favorece abordagens simples, tornando o *Naive Mean* competitivo frente a modelos mais complexos.

Em síntese, os resultados reforçam que a eficácia de um modelo não depende exclusivamente de sua complexidade, mas está fortemente condicionada ao contexto de mobilidade e ao perfil do tráfego. A escolha do método deve, portanto, equilibrar acurácia, custo computacional e adequação ao cenário de aplicação.

Capítulo 6

Conclusão

Este estudo comparou múltiplas abordagens de previsão de *throughput* em redes móveis 5G, avaliando o compromisso entre acurácia, robustez e custo computacional em diferentes cenários de mobilidade (condução/estático) e atividade (transmissão de vídeo/*download*). Os resultados mostraram que a eficácia dos modelos depende tanto de sua complexidade quanto do contexto de tráfego: em condições dinâmicas e irregulares, métodos sofisticados tendem a superar o *baseline*; já em cenários mais regulares, abordagens simples podem permanecer competitivas. Além disso, o uso do *Naive Mean* como referência forneceu uma base padronizada para comparar modelos e séries, evidenciando com clareza quando os ganhos de acurácia justificam a complexidade adicional.

Os resultados apresentados e a discussão realizada oferecem subsídios para retomar as questões de pesquisa propostas e avaliar em que medida podem ser respondidas de forma fundamentada.

QP1 Quais métodos de previsão são mais eficazes sob diferentes condições de rede e mobilidade?

QP2 Quais são as vantagens e as limitações comparativas entre as abordagens de modelagem local e global?

Resposta à QP1:

O *LightGBM* apresentou o melhor desempenho geral, sobretudo em configuração global sem covariáveis, graças ao equilíbrio entre: (i) baixa média de erro, (ii) robustez

a assimetrias e *outliers*, e (iii) eficiência computacional (treinando até duas ordens de grandeza mais rápido que redes profundas). Modelos de *ensemble* (LightGBM, *Random Forest*) e regressão linear mantiveram, de forma consistente, um bom compromisso acurácia–custo. Em contraste, modelos de aprendizado profundo (*Block RNN*, *Transformer*, *N-BEATS*) obtiveram erros comparáveis aos clássicos, porém com *overhead* de treinamento que chega a três–quatro ordens de magnitude, o que inviabiliza reconfigurações frequentes ou operação quase em tempo real. Os cenários confirmaram a dependência do contexto: na série 75 (estático, vídeo), o LightGBM sem covariáveis superou as demais variantes (RMSE \approx 3,50 vs. 3,70 com covariáveis), sugerindo que covariáveis de qualidade de canal podem introduzir ruído em padrões estáveis. Já na série 50 (condução, vídeo), caracterizada por baixo *throughput* médio, poucos picos e dados mais espaçados pelo efeito de *buffer*, o *Naive Mean* superou o LightGBM (com e sem covariáveis), mostrando que, para horizontes curtos e tráfego com baixa variabilidade, métodos simples ainda podem ser mais adequados.

Resposta à QP2:

A modelagem global mostrou-se mais vantajosa que a local. Modelos globais exploram padrões compartilhados entre séries, generalizando melhor e reduzindo o custo de manutenção, enquanto abordagens locais (p. ex., *AutoARIMA*, *Exponential Smoothing*) exigem ajuste fino por série e não compartilham conhecimento, limitando a escalabilidade. Adicionalmente, o uso exclusivo de métricas de qualidade de canal (RSRP, RSRQ, SNR, CQI, RSSI) como covariáveis não trouxe ganhos consistentes — em cenário estático chegou a degradar o resultado — corroborando a literatura (NARAYANAN et al., 2020) quanto à necessidade de incluir covariáveis contextuais (mobilidade, localização, condições de tráfego e rede) para ganhos estáveis.

Em síntese, modelos baseados em árvores (especialmente LightGBM) constituem a solução mais equilibrada para previsão de *throughput* em 5G, conciliando acurácia, robustez e viabilidade computacional. A escolha do método deve considerar o contexto de mobilidade e o perfil do tráfego: covariáveis tendem a ser mais úteis em condução; em cenários estáticos, podem ser dispensadas; e, em situações de *throughput* baixo e pouca variabilidade, métodos simples podem ser suficientes.

Limitações deste trabalho:

As principais limitações identificadas neste estudo são:

- **Escopo do conjunto de dados:** utilização de um único *dataset* (Irlanda, rede comercial em transição 4G/5G), o que pode restringir a generalização dos resultados para outros contextos de rede, regiões e cenários de uso.

- ❑ **Covariáveis e contexto:** foram consideradas apenas métricas de canal e a velocidade da UE, sem a exploração de diferentes combinações ou configurações dessas variáveis. Uma análise comparativa entre conjuntos distintos de métricas poderia verificar se os resultados permaneceriam consistentes, mesmo sem incluir informações de localização ou outros dados contextuais adicionais.
- ❑ **Pré-processamento:** adoção de uma estratégia única de imputação e normalização para garantir consistência entre os experimentos, o que limitou a comparação com alternativas de pré-processamento.
- ❑ **Horizonte e tarefa:** avaliação restrita a horizonte curto ($H=10$) e previsão pontual, não contemplando cenários multi-horizonte ou probabilísticos, que poderiam ampliar a aplicabilidade prática.
- ❑ **Hiperparâmetros:** ausência de ajuste sistemático dos modelos, especialmente nas arquiteturas de aprendizado profundo. Embora o uso dos parâmetros padrão favoreça comparabilidade e reprodutibilidade, a otimização individualizada poderia potencialmente melhorar o desempenho, ainda que traga desafios adicionais, sobretudo para modelos locais com características específicas de cada série.

Trabalhos futuros:

Com base nas limitações identificadas, são destacadas algumas direções promissoras para aprofundar esta linha de pesquisa e ampliar a aplicabilidade dos modelos em cenários reais de redes móveis:

- ❑ Investigar covariáveis contextuais (mobilidade fina, posicionamento geográfico, carga da célula, métricas de congestionamento e *handover*).
- ❑ Estudar horizontes de previsão e janelas de amostragem que mitiguem efeitos de *buffering* em vídeo.
- ❑ Incluir etapa sistemática de otimização de hiperparâmetros para os modelos de aprendizado de máquina, avaliando o impacto no equilíbrio entre acurácia e custo computacional.
- ❑ Explorar modelos híbridos (modelos simples + complexos) e adaptação *on-line* orientada a custo.
- ❑ Avaliar métricas de utilidade fim-a-fim (QoE) além de RMSE/MAE, aproximando a modelagem das necessidades operacionais.

Contribuições científicas:

Parte dos resultados preliminares desta dissertação culminou na publicação do artigo “From Statistics to Deep Learning: Forecasting Mobile Throughput“, aceito para apresentação no *13th Symposium on Knowledge Discovery, Mining and Learning (KDMiLe’25)*. Adicionalmente, todo o código-fonte desenvolvido ao longo do trabalho encontra-se disponível em repositório público, de modo a assegurar transparência, reprodutibilidade e potencial de reutilização dos resultados¹.

¹ Repositório da dissertação: <<https://github.com/ejs94/5g-forecasting>>

Referências

- AFROZ, F. et al. SINR, RSRP, RSSI and RSRQ measurements in long term evolution networks. **International Journal of Wireless & Mobile Networks**, Academy and Industry Research Collaboration Center (AIRCC), v. 7, p. 113–123, 08 2015.
- AGIWAL, M.; ROY, A.; SAXENA, N. Next generation 5G wireless networks: A comprehensive survey. **IEEE Communications Surveys and Tutorials**, v. 18, p. 1617–1655, 2016.
- ALRAIH, S. et al. A survey on handover optimization in beyond 5g mobile networks: Challenges and solutions. **IEEE Access**, Institute of Electrical and Electronics Engineers Inc., v. 11, p. 59317–59345, 2023. ISSN 21693536.
- ANDREWS, J. G. et al. **Fundamentals of Wireless Communication**. 1st. ed. [S.l.]: Pearson Education, 2014. ISBN 9780132228380.
- BATOOL, I.; FOUDA, M. M.; FADLULLAH, Z. M. Deep Learning-Based Throughput Prediction in 5G Cellular Networks. In: **2024 International Conference on Smart Applications, Communications and Networking (SmartNets)**. [S.l.]: Institute of Electrical and Electronics Engineers (IEEE), 2024.
- BOUTIBA, K.; BAGAA, M.; KSENTINI, A. Radio link failure prediction in 5G networks. In: **IEEE. 2021 IEEE Global Communications Conference (GLOBECOM)**. [S.l.], 2021. p. 1–6.
- BOX, G. et al. **Time Series Analysis: Forecasting and Control**. Wiley, 2015. (Wiley Series in Probability and Statistics). ISBN 9781118674925. Disponível em: <<https://books.google.com.br/books?id=rNt5CgAAQBAJ>>.
- BREIMAN, L. et al. **Classification and Regression Trees**. [S.l.]: Wadsworth & Brooks/Cole Advanced Books & Software, 1984.
- DICKEY, D. A.; FULLER, W. A. Distribution of the estimators for autoregressive time series with a unit root. **Journal of the American Statistical Association**, Taylor & Francis, v. 74, n. 366a, p. 427–431, 1979.
- ELSHERBINY, H. et al. 4g lte network throughput modelling and prediction. In: **GLOBECOM 2020 - 2020 IEEE Global Communications Conference**. IEEE, 2020. p. 1–6. ISBN 978-1-7281-8298-8. Disponível em: <<https://ieeexplore.ieee.org/document/9322410/>>.

FULLER, W. A. **Introduction to Statistical Time Series**. 2nd. ed. New York, NY: John Wiley & Sons, 1996. (Wiley Series in Probability and Statistics).

GERS, F. A.; SCHMIDHUBER, J.; CUMMINS, F. Learning to forget: Continual prediction with lstm. **Neural computation**, MIT Press, v. 12, n. 10, p. 2451–2471, 2000.

GHOSH, A. et al. 5G Evolution: A View on 5G Cellular Technology beyond 3GPP Release 15. **IEEE Access**, Institute of Electrical and Electronics Engineers Inc., v. 7, p. 127639–127651, 2019. ISSN 21693536. Disponível em: <<https://ieeexplore.ieee.org/document/8826541>>.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. Cambridge, MA: MIT Press, 2016. Disponível em: <<https://www.deeplearningbook.org/>>.

GSMA. **A Economia Móvel na América Latina 2024**. [S.l.], 2024. Accessed: 2024-12-03. Disponível em: <<https://www.gsma.com/solutions-and-impact/connectivity-for-good/mobile-economy/wp-content/uploads/2024/06/A-Economia-Movel-na-America-Latina-2024.pdf>>.

HAMILTON, J. **Time Series Analysis**. [S.l.]: Princeton University Press, 2020. ISBN 9780691218632.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The Elements of Statistical Learning: Data Mining, Inference, and Prediction**. 2. ed. New York: Springer, 2009.

HEWAMALAGE, H.; ACKERMANN, K.; BERGMEIR, C. Forecast evaluation for data scientists: common pitfalls and best practices. **Data Mining and Knowledge Discovery**, Springer, v. 37, p. 788–832, 3 2023. ISSN 1573756X. Disponível em: <<https://link.springer.com/article/10.1007/s10618-022-00894-5>>.

HOCHREITER, S.; SCHMIDHUBER, J. Long short-term memory. **Neural computation**, MIT Press, v. 9, n. 8, p. 1735–1780, 1997.

HOERL, A. E.; KENNARD, R. W. Ridge regression: Biased estimation for nonorthogonal problems. **Technometrics**, v. 12, n. 1, p. 55–67, 1970.

HYNDMAN, R. J. A brief history of forecasting competitions. **International Journal of Forecasting**, Elsevier, v. 36, p. 7–14, 1 2020. ISSN 0169-2070.

HYNDMAN, R. J.; ATHANASOPOULOS, G. **Forecasting: Principles and Practice**. 3rd. ed. Melbourne, Australia: OTexts, 2021. Disponível em: <<https://otexts.com/fpp3/>>.

HYNDMAN, R. J.; KOEHLER, A. B. Another look at measures of forecast accuracy. **International Journal of Forecasting**, v. 22, p. 679–688, 10 2006. ISSN 01692070.

JAMES, G. et al. **An Introduction to Statistical Learning: With Applications in Python**. Cham: Springer, 2023. (Springer Texts in Statistics). ISBN 978-3-031-38746-3. Disponível em: <<https://link.springer.com/book/10.1007/978-3-031-38747-0>>.

JANUSCHOWSKI, T. et al. Criteria for classifying forecasting methods. **International Journal of Forecasting**, Elsevier, v. 36, p. 167–177, 1 2020. ISSN 0169-2070.

JOSEPH, M. **Modern Time Series Forecasting with Python**. Birmingham, England: Packt Publishing, 2022. v. 1.

KALNTIS, M. et al. Through the telco lens: A countrywide empirical study of cellular handovers. **Proceedings of the 2024 ACM on Internet Measurement Conference**, ACM, p. 51–67, 11 2024. Disponível em: <<https://dl.acm.org/doi/10.1145/3646547.3688452>>.

KE, G. et al. Lightgbm: a highly efficient gradient boosting decision tree. In: **Proceedings of the 31st International Conference on Neural Information Processing Systems**. Red Hook, NY, USA: Curran Associates Inc., 2017. (NIPS'17), p. 3149–3157. ISBN 9781510860964.

KUTNER, M. H. et al. **Applied Linear Statistical Models**. 5. ed. Boston: McGraw-Hill, 2005.

MAKRIDAKIS, S.; SPILIOTIS, E.; ASSIMAKOPOULOS, V. The M4 Competition: 100,000 time series and 61 forecasting methods. **International Journal of Forecasting**, Elsevier, v. 36, p. 54–74, 1 2020. ISSN 0169-2070.

_____. M5 accuracy competition: Results, findings, and conclusions. **International Journal of Forecasting**, Elsevier B.V., v. 38, p. 1346–1364, 10 2022. ISSN 01692070.

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

NARAYANAN, A. et al. Lumos5G: Mapping and Predicting Commercial mmWave 5G Throughput. In: **Proceedings of the ACM SIGCOMM Internet Measurement Conference (IMC '20)**. [S.l.]: Association for Computing Machinery (ACM), 2020. p. 176–193. ISBN 9781450381383.

ORESHKIN, B. N. et al. N-BEATS: Neural basis expansion analysis for interpretable time series forecasting. ArXiv:1905.10437. 2019. Disponível em: <<https://arxiv.org/abs/1905.10437v4>>.

_____. N-beats neural network for mid-term electricity load forecasting. **Applied Energy**, Elsevier, v. 293, p. 116918, 2021. ISSN 0306-2619. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0306261921003986>>.

PEARSON, K. Note on regression and inheritance in the case of two parents. **Proceedings of the Royal Society of London**, The Royal Society, v. 58, p. 240–242, 1895.

RACA, D. et al. Beyond throughput, the next generation: A 5G dataset with channel and context metrics. **MMSys 2020 - Proceedings of the 2020 Multimedia Systems Conference**, Association for Computing Machinery, Inc, p. 303–308, 5 2020. Disponível em: <<https://dl.acm.org/doi/10.1145/3339825.3394938>>.

RESHMI, T. R.; AZATH, M. Improved self-healing technique for 5g networks using predictive analysis. **Peer-to-Peer Networking and Applications**, Springer, v. 14, p. 375–391, 1 2021. ISSN 19366450. Disponível em: <<https://link.springer.com/article/10.1007/s12083-020-00926-1>>.

- RODGERS, J. L.; NICEWANDER, W. A. Thirteen ways to look at the correlation coefficient. **The American Statistician**, Taylor & Francis, v. 42, n. 1, p. 59–66, 1988.
- RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning representations by back-propagating errors. **Nature**, v. 323, n. 6088, p. 533–536, 1986.
- SANTOS, G. L. et al. When 5g meets deep learning: A systematic review. **Algorithms**, Multidisciplinary Digital Publishing Institute, v. 13, p. 208, aug 2020. ISSN 1999-4893. Disponível em: <<https://www.mdpi.com/1999-4893/13/9/208>>.
- SCHUSTER, M.; PALIWAL, K. K. Bidirectional recurrent neural networks. **IEEE Transactions on Signal Processing**, v. 45, n. 11, p. 2673–2681, 1997.
- SEBER, G. A. F.; LEE, A. J. **Linear Regression Analysis**. 2. ed. Hoboken: Wiley, 2003.
- SHARMA, A.; PANDIT, S.; TALLURI, S. R. Throughput prediction of fifth-generation cellular system using hybrid feature selection and enhanced sequential decision tree machine learning algorithm. **Wireless Networks**, Springer, v. 31, p. 3025–3042, 2025.
- SUTSKEVER, I.; VINYALS, O.; LE, Q. V. Sequence to sequence learning with neural networks. In: **Advances in Neural Information Processing Systems (NeurIPS)**. Curran Associates, Inc., 2014. v. 27. Disponível em: <<https://papers.nips.cc/paper/2014/hash/a14ac55a4f27472c5d894ec1c3c743d2-Abstract.html>>.
- TIBSHIRANI, R. Regression shrinkage and selection via the lasso. **Journal of the Royal Statistical Society: Series B (Methodological)**, v. 58, n. 1, p. 267–288, 1996.
- YEASER, K. M. A.; HASSAN, K. M. A. 5G NR V2X Throughput Prediction Using Deep Hybrid Learning. In: KALAM, A.; MEKHILEF, S.; WILLIAMSON, S. S. (Ed.). **Innovations in Electrical and Electronics Engineering: Proceedings of the 5th ICIEEL 2024**. Singapore: Springer Nature Singapore, 2025. p. 685–693. ISBN 978-981-97-9112-5.
- YINGJIE, Z.; ABOLGHASEMI, M. **Local vs. Global Models for Hierarchical Forecasting**. 2024. Disponível em: <<https://arxiv.org/abs/2411.06394v1>>.

Apêndices

APÊNDICE A

Estatísticas de desempenho dos modelos por cenário

Este apêndice reúne, em formato tabular, os valores numéricos que sustentam os *box-plots* apresentados no Capítulo 5. Para cada modelo, em cada configuração experimental de atividade e mobilidade (cenários agregados, estático, condução, transmissão de vídeo e download de arquivo), são apresentadas estatísticas descritivas dos erros RMSE e MAE (mínimo, primeiro quartil, mediana, terceiro quartil e máximo). As tabelas são organizadas de modo a facilitar a consulta e a assegurar a transparência e a reprodutibilidade das análises.

Tabela 16 – Resumo estatístico dos erros de previsão (RMSE e MAE) para múltiplos modelos considerando todos os cenários incorporados.

Modelo	RMSE					MAE				
	Min	Q1	Mediana	Q3	Máx	Min	Q1	Mediana	Q3	Máx
LightGBM [Global não covariado]	0,512	1,537	3,400	4,516	93,547	0,407	0,874	1,360	2,413	70,794
LightGBM [Global covariado]	0,531	1,502	3,457	4,623	89,713	0,426	0,916	1,385	2,424	69,812
AutoARIMA [Local univariado]	0,311	1,580	3,489	5,483	108,879	0,151	0,797	1,400	2,780	94,327
Transformer [Global não covariado]	0,637	1,678	3,530	4,874	98,443	0,509	1,096	1,688	3,190	75,454
Naive Aggregate [Global não covariado]	0,001	1,439	3,558	4,684	106,373	0,001	0,655	1,277	1,790	82,978
Linear Regression [Global covariado]	0,330	1,726	3,575	4,897	97,949	0,282	0,979	1,620	3,222	73,742
Random Forest [Global não covariado]	0,605	1,805	3,587	4,846	92,667	0,521	0,855	1,459	2,656	70,460
Random Forest [Global covariado]	0,719	1,613	3,591	4,947	90,317	0,557	1,031	1,548	3,022	68,849
Block RNN [Global não covariado]	0,673	1,679	3,599	4,851	97,068	0,527	1,029	1,666	3,146	75,070
N-BEATS [Global covariado]	0,180	1,685	3,603	5,205	107,062	0,159	0,709	1,158	2,418	81,376
Block RNN [Global covariado]	0,515	1,687	3,628	4,934	91,773	0,460	1,057	1,644	3,478	71,119
N-BEATS [Global não covariado]	0,218	1,829	3,639	5,272	103,850	0,212	0,983	1,325	2,452	76,939
Transformer [Global covariado]	0,711	1,646	3,685	5,511	99,019	0,561	1,141	1,781	3,921	85,594
Linear Regression [Global não covariado]	0,439	1,726	3,699	4,788	97,772	0,356	1,031	1,688	2,825	73,827
Naive Mean [Local univariado]	0,665	1,662	3,777	6,326	96,785	0,550	0,835	1,491	4,708	83,510
Exponential Smoothing [Local univariado]	0,001	1,891	4,011	6,062	122,763	0,001	0,739	1,413	2,949	98,667
Naive Seasonal [Global não covariado]	0,001	1,907	4,137	5,900	128,076	0,000	0,778	1,367	2,133	106,022
Naive Drift [Global não covariado]	0,002	3,252	7,049	10,707	217,441	0,001	1,400	2,462	3,622	184,954

Fonte: Autoria própria.

Tabela 17 – Resumo estatístico dos erros de previsão (RMSE e MAE) para múltiplos modelos no cenário estático.

Modelo	RMSE					MAE				
	Min	Q1	Mediana	Q3	Máx	Min	Q1	Mediana	Q3	Máx
AutoARIMA [Local univariado]	0,750	2,774	4,277	59,105	108,188	0,365	1,135	1,490	36,481	94,327
Naive Mean [Local univariado]	0,830	2,781	4,370	65,269	96,785	0,658	1,122	1,491	46,103	83,510
LightGBM [Global não covariado]	0,763	2,777	4,388	54,756	82,270	0,604	1,205	2,285	31,695	69,708
LightGBM [Global covariado]	0,849	2,774	4,422	54,892	83,708	0,639	1,181	2,176	33,069	69,812
Random Forest [Global covariado]	0,953	2,852	4,547	55,204	81,052	0,705	1,187	1,893	34,434	68,849
Naive Aggregate [Global não covariado]	0,639	2,906	4,591	60,753	92,944	0,291	0,989	1,490	35,355	79,866
Transformer [Global não covariado]	0,944	2,820	4,621	56,545	88,492	0,683	1,451	3,127	33,625	73,683
Block RNN [Global não covariado]	0,903	2,822	4,626	56,884	86,239	0,706	1,439	3,022	34,083	70,683
Block RNN [Global covariado]	1,037	2,804	4,670	55,992	83,845	0,659	1,364	2,918	35,225	70,909
Random Forest [Global não covariado]	0,825	2,922	4,709	54,994	81,239	0,610	1,148	1,863	32,315	67,253
Linear Regression [Global não covariado]	0,811	2,903	4,709	57,136	87,888	0,627	1,450	2,765	32,154	73,603
N-BEATS [Global não covariado]	1,004	2,889	4,733	60,333	94,519	0,420	1,131	1,928	34,765	75,477
Exponential Smoothing [Local univariado]	0,625	2,826	4,753	75,224	119,712	0,296	1,034	1,479	45,291	98,667
Linear Regression [Global covariado]	0,848	2,885	4,792	56,786	87,933	0,503	1,268	3,370	32,227	73,499
N-BEATS [Global covariado]	1,237	2,817	4,809	60,134	93,627	0,554	0,946	1,698	34,287	75,404
Transformer [Global covariado]	1,067	2,844	5,075	59,393	97,119	0,709	1,516	3,815	36,491	85,594
Naive Seasonal [Global não covariado]	0,803	3,380	5,621	78,612	128,076	0,291	1,082	1,505	48,202	106,022
Naive Drift [Global não covariado]	1,517	4,937	10,066	132,942	213,422	0,434	2,101	2,974	84,162	184,954

Fonte: Autoria própria.

Tabela 18 – Resumo estatístico dos erros de previsão (RMSE e MAE) para múltiplos modelos no cenário de condução.

Modelo	RMSE					MAE				
	Min	Q1	Mediana	Q3	Máx	Min	Q1	Mediana	Q3	Máx
Naive Aggregate [Global não covariado]	0,001	1,263	2,824	4,229	106,373	0,001	0,577	0,989	1,638	82,978
LightGBM [Global covariado]	0,531	1,232	2,849	4,021	89,713	0,426	0,848	1,311	1,859	67,168
AutoARIMA [Local univariado]	0,311	1,454	2,875	4,414	108,879	0,151	0,780	1,200	2,012	69,557
Random Forest [Global covariado]	0,719	1,576	2,966	4,031	90,317	0,557	1,028	1,453	2,011	68,684
LightGBM [Global não covariado]	0,512	1,234	2,974	3,940	93,547	0,407	0,871	1,257	1,883	70,794
N-BEATS [Global covariado]	0,180	1,632	3,014	4,336	107,062	0,159	0,606	1,045	1,825	81,376
Linear Regression [Global covariado]	0,330	1,721	3,017	4,155	97,949	0,282	0,942	1,525	2,001	73,742
Transformer [Global não covariado]	0,637	1,497	3,049	4,081	98,443	0,509	1,089	1,513	2,368	75,454
Transformer [Global covariado]	0,711	1,536	3,072	4,366	99,019	0,561	1,081	1,641	2,582	77,363
N-BEATS [Global não covariado]	0,218	1,743	3,088	4,196	103,850	0,212	0,923	1,262	1,804	76,939
Linear Regression [Global não covariado]	0,439	1,324	3,102	4,320	97,772	0,356	1,011	1,562	1,986	73,827
Random Forest [Global não covariado]	0,605	1,776	3,118	4,320	92,667	0,521	0,848	1,327	1,976	70,460
Block RNN [Global não covariado]	0,673	1,417	3,133	4,116	97,068	0,527	0,998	1,598	2,343	75,070
Block RNN [Global covariado]	0,515	1,524	3,156	4,118	91,773	0,460	0,999	1,485	2,625	71,119
Naive Seasonal [Global não covariado]	0,001	1,626	3,494	5,183	124,234	0,000	0,680	1,079	1,712	95,799
Naive Mean [Local univariado]	0,665	1,537	3,505	5,713	95,133	0,550	0,828	1,477	2,945	67,061
Exponential Smoothing [Local univariado]	0,001	1,810	3,539	5,263	122,763	0,001	0,715	1,358	2,058	94,305
Naive Drift [Global não covariado]	0,002	2,728	5,957	9,495	217,441	0,001	1,234	2,086	3,340	176,705

Fonte: Autoria própria.

Tabela 19 – Resumo estatístico dos erros de previsão (RMSE e MAE) para múltiplos modelos no cenário de transmissão de vídeo.

Modelo	RMSE					MAE				
	Min	Q1	Mediana	Q3	Máx	Min	Q1	Mediana	Q3	Máx
LightGBM [Global não covariado]	0,512	2,639	3,780	7,142	93,547	0,407	1,188	1,826	3,749	70,794
LightGBM [Global covariado]	0,531	2,618	3,789	7,414	89,713	0,426	1,178	1,817	3,843	69,812
Block RNN [Global não covariado]	0,698	2,760	3,841	7,315	97,068	0,527	1,480	2,077	4,743	75,070
Transformer [Global não covariado]	0,713	2,712	3,855	7,460	98,443	0,539	1,432	2,000	5,020	75,454
AutoARIMA [Local univariado]	0,311	2,535	3,856	7,384	108,879	0,151	1,115	1,593	4,356	94,327
Random Forest [Global covariado]	0,806	2,703	3,861	7,423	90,317	0,616	1,316	1,818	3,668	68,849
Linear Regression [Global não covariado]	0,439	2,794	3,908	7,343	97,772	0,356	1,549	1,835	4,190	73,827
Linear Regression [Global covariado]	0,330	2,546	3,919	7,360	97,949	0,282	1,467	1,760	3,839	73,742
Random Forest [Global não covariado]	0,605	2,808	3,995	7,131	92,667	0,543	1,196	1,773	3,569	70,460
Naive Aggregate [Global não covariado]	0,001	2,529	4,073	7,712	106,373	0,001	1,020	1,563	3,643	82,978
Block RNN [Global covariado]	0,515	2,784	4,082	8,324	91,773	0,460	1,321	2,198	5,677	71,119
N-BEATS [Global não covariado]	0,218	2,813	4,102	8,099	103,850	0,212	1,135	1,558	6,944	76,939
N-BEATS [Global covariado]	0,180	2,694	4,124	7,459	107,062	0,159	0,946	1,413	4,095	81,376
Transformer [Global covariado]	0,743	2,749	4,192	8,264	99,019	0,584	1,503	1,955	6,642	85,594
Naive Mean [Local univariado]	0,724	2,898	4,468	9,896	96,785	0,550	1,350	2,133	9,093	83,510
Exponential Smoothing [Local univariado]	0,001	3,046	4,516	8,484	122,763	0,001	1,182	1,594	4,217	98,667
Naive Seasonal [Global não covariado]	0,001	3,092	5,080	8,851	128,076	0,000	1,066	1,587	4,355	106,022
Naive Drift [Global não covariado]	0,002	5,001	8,612	16,513	217,441	0,001	2,053	3,075	8,221	184,954

Fonte: Autoria própria.

Tabela 20 – Resumo estatístico dos erros de previsão (RMSE e MAE) para múltiplos modelos no cenário de download de arquivo.

Modelo	RMSE					MAE				
	Min	Q1	Mediana	Q3	Máx	Min	Q1	Mediana	Q3	Máx
LightGBM [Global não covariado]	0,622	0,995	1,172	2,915	4,255	0,453	0,676	0,840	1,204	1,883
Naive Aggregate [Global não covariado]	0,517	0,973	1,172	3,058	4,463	0,209	0,509	0,689	0,879	1,738
Naive Mean [Local univariado]	0,665	1,008	1,188	2,661	4,234	0,576	0,742	0,810	0,940	1,924
AutoARIMA [Local univariado]	0,619	1,021	1,191	2,760	4,235	0,458	0,631	0,785	0,889	1,923
LightGBM [Global covariado]	0,631	1,072	1,211	2,899	4,270	0,484	0,746	0,951	1,157	2,054
Exponential Smoothing [Local univariado]	0,625	1,096	1,236	3,233	5,731	0,187	0,543	0,733	0,906	1,853
Random Forest [Global não covariado]	0,681	1,006	1,273	2,953	4,895	0,521	0,760	0,848	1,041	1,977
Transformer [Global não covariado]	0,637	1,071	1,291	3,057	4,276	0,509	0,868	1,053	1,504	2,614
Linear Regression [Global covariado]	0,637	1,008	1,292	3,081	4,818	0,517	0,711	0,987	1,398	3,274
Block RNN [Global não covariado]	0,673	1,028	1,293	3,040	4,252	0,568	0,867	0,992	1,516	2,535
Linear Regression [Global não covariado]	0,641	1,012	1,308	3,102	4,505	0,512	0,756	0,990	1,442	2,367
Transformer [Global covariado]	0,711	1,096	1,358	3,072	4,488	0,561	0,930	1,081	1,547	2,944
Random Forest [Global covariado]	0,719	1,133	1,416	2,819	4,313	0,557	0,783	1,028	1,376	1,969
Block RNN [Global covariado]	0,704	1,158	1,484	2,980	4,527	0,553	0,819	1,110	1,419	2,784
Naive Seasonal [Global não covariado]	0,782	1,115	1,509	3,722	5,350	0,252	0,560	0,794	0,934	1,595
N-BEATS [Global covariado]	0,619	1,194	1,611	2,942	4,449	0,256	0,675	0,749	0,947	1,561
N-BEATS [Global não covariado]	0,759	1,261	1,696	3,042	4,642	0,368	0,754	0,995	1,266	1,804
Naive Drift [Global não covariado]	1,337	1,767	2,604	5,957	9,865	0,429	1,001	1,437	1,967	3,485

Fonte: Autoria própria.

Tabela 21 – Resumo estatístico dos erros de previsão (RMSE e MAE) para múltiplos modelos no cenário de transmissão de vídeo em condição estática.

Modelo	RMSE					MAE				
	Min	Q1	Mediana	Q3	Máx	Min	Q1	Mediana	Q3	Máx
Linear Regression [Global não covariado]	2,314	4,319	10,637	65,280	87,888	0,627	2,366	6,100	40,500	73,603
Linear Regression [Global covariado]	2,309	4,380	10,734	64,770	87,933	0,503	2,875	5,858	40,390	73,499
Naive Aggregate [Global não covariado]	2,315	4,264	11,074	68,455	92,944	0,291	1,364	6,045	44,782	79,866
LightGBM [Global covariado]	2,292	4,077	11,191	61,143	83,708	0,639	1,950	6,078	39,567	69,812
Random Forest [Global covariado]	2,387	4,233	11,303	61,755	81,052	0,939	1,781	7,087	41,330	68,849
LightGBM [Global não covariado]	2,293	4,022	11,371	59,772	82,270	0,641	1,905	6,453	39,417	69,708
Block RNN [Global covariado]	2,303	4,567	11,873	63,701	83,845	0,824	2,795	6,791	42,434	70,909
Transformer [Global não covariado]	2,310	4,173	12,006	63,713	88,492	0,802	2,735	6,831	42,915	73,683
Transformer [Global covariado]	2,317	4,695	12,272	68,219	97,119	0,846	3,264	6,965	44,336	85,594
Block RNN [Global não covariado]	2,309	4,164	12,369	62,160	86,239	0,790	2,657	8,010	43,251	70,683
N-BEATS [Global não covariado]	2,320	4,416	12,832	70,885	94,519	0,420	1,650	8,514	43,332	75,477
Random Forest [Global não covariado]	2,400	4,448	12,857	60,183	81,239	0,777	1,644	7,843	40,765	67,253
N-BEATS [Global covariado]	2,378	4,463	13,262	74,497	93,627	0,554	1,261	7,013	41,148	75,404
AutoARIMA [Local univariado]	2,300	4,015	13,420	66,608	108,188	0,365	1,425	9,532	47,579	94,327
Exponential Smoothing [Local univariado]	2,376	4,396	14,093	89,341	119,712	0,296	1,409	7,474	53,951	98,667
Naive Mean [Local univariado]	2,314	4,110	14,104	82,242	96,785	0,795	1,412	11,118	62,330	83,510
Naive Seasonal [Global não covariado]	2,500	5,281	15,355	88,501	128,076	0,291	1,373	8,324	59,116	106,022
Naive Drift [Global não covariado]	3,304	8,765	26,271	162,964	213,422	0,434	2,621	13,756	104,255	184,954

Fonte: Autoria própria.

Tabela 22 – Resumo estatístico dos erros de previsão (RMSE e MAE) para múltiplos modelos no cenário de download de arquivo em condição estática.

Modelo	RMSE					MAE				
	Min	Q1	Mediana	Q3	Máx	Min	Q1	Mediana	Q3	Máx
Naive Mean [Local univariado]	0,830	1,008	1,085	1,188	1,443	0,658	0,722	0,746	0,834	0,940
Linear Regression [Global covariado]	0,848	1,008	1,089	1,292	1,711	0,692	0,711	0,739	1,007	1,420
AutoARIMA [Local univariado]	0,750	1,021	1,090	1,191	1,444	0,495	0,610	0,737	0,834	0,940
Naive Aggregate [Global não covariado]	0,639	0,973	1,093	1,167	1,462	0,346	0,509	0,689	0,741	0,926
Transformer [Global não covariado]	0,944	1,071	1,098	1,179	1,608	0,683	0,761	0,868	0,884	1,313
Block RNN [Global não covariado]	0,903	1,026	1,098	1,236	1,619	0,706	0,753	0,830	0,905	1,285
Linear Regression [Global não covariado]	0,811	1,012	1,100	1,214	1,598	0,714	0,717	0,772	0,879	1,280
Exponential Smoothing [Local univariado]	0,625	1,097	1,104	1,236	1,833	0,337	0,569	0,733	0,746	1,019
LightGBM [Global covariado]	0,849	1,072	1,113	1,162	1,525	0,688	0,710	0,746	0,775	1,157
LightGBM [Global não covariado]	0,763	0,995	1,116	1,140	1,492	0,604	0,642	0,676	0,740	1,068
Random Forest [Global covariado]	0,953	1,056	1,133	1,215	1,588	0,705	0,767	0,785	0,852	1,078
Block RNN [Global covariado]	1,037	1,151	1,158	1,298	1,702	0,659	0,716	0,999	1,055	1,419
Random Forest [Global não covariado]	0,825	1,006	1,160	1,214	1,509	0,610	0,634	0,768	0,795	1,009
Transformer [Global covariado]	1,067	1,092	1,168	1,185	1,680	0,709	0,874	0,951	1,009	1,395
Naive Seasonal [Global não covariado]	0,803	1,115	1,364	1,509	1,920	0,376	0,623	0,794	0,907	1,111
N-BEATS [Global covariado]	1,237	1,357	1,378	1,407	1,696	0,675	0,698	0,712	0,720	0,996
N-BEATS [Global não covariado]	1,004	1,343	1,518	1,518	2,022	0,702	0,754	0,916	0,968	1,408
Naive Drift [Global não covariado]	1,517	1,767	2,081	2,604	3,432	0,753	1,052	1,437	1,651	2,212

Fonte: Autoria própria.

Tabela 23 – Resumo estatístico dos erros de previsão (RMSE e MAE) para múltiplos modelos no cenário de transmissão de vídeo em condição de condução.

Modelo	RMSE					MAE				
	Min	Q1	Mediana	Q3	Máx	Min	Q1	Mediana	Q3	Máx
Linear Regression [Global covariado]	0,330	1,817	3,061	4,235	97,949	0,282	1,361	1,655	2,098	73,742
LightGBM [Global não covariado]	0,512	2,004	3,481	4,139	93,547	0,407	1,022	1,373	2,376	70,794
LightGBM [Global covariado]	0,531	1,760	3,484	4,616	89,713	0,426	0,968	1,392	2,416	67,168
Transformer [Global não covariado]	0,713	2,034	3,496	4,912	98,443	0,539	1,288	1,683	2,862	75,454
AutoARIMA [Local univariado]	0,311	1,946	3,514	5,588	108,879	0,151	0,851	1,520	2,740	69,557
Naive Aggregate [Global não covariado]	0,001	1,790	3,542	4,246	106,373	0,001	0,736	1,507	1,740	82,978
Block RNN [Global covariado]	0,515	1,985	3,583	4,915	91,773	0,460	0,999	1,755	3,265	71,119
Block RNN [Global não covariado]	0,698	2,245	3,590	4,938	97,068	0,527	1,257	1,665	3,116	75,070
Random Forest [Global não covariado]	0,605	2,088	3,608	4,350	92,667	0,543	1,023	1,531	2,649	70,460
N-BEATS [Global covariado]	0,180	1,814	3,617	5,461	107,062	0,159	0,674	1,288	2,411	81,376
Transformer [Global covariado]	0,743	2,250	3,643	5,475	99,019	0,584	1,247	1,742	3,124	77,363
Random Forest [Global covariado]	0,806	2,067	3,653	4,938	90,317	0,616	1,140	1,566	3,012	68,684
N-BEATS [Global não covariado]	0,218	2,625	3,680	5,825	103,850	0,212	1,043	1,310	2,429	76,939
Linear Regression [Global não covariado]	0,439	2,175	3,686	4,759	97,772	0,356	1,131	1,728	2,532	73,827
Naive Mean [Local univariado]	0,724	2,575	3,844	6,574	95,133	0,550	1,110	1,941	5,040	67,061
Naive Seasonal [Global não covariado]	0,001	2,472	3,862	5,494	124,234	0,000	0,797	1,582	2,126	95,799
Exponential Smoothing [Local univariado]	0,001	2,451	4,309	6,084	122,763	0,001	0,934	1,581	2,881	94,305
Naive Drift [Global não covariado]	0,002	3,814	6,901	10,551	217,441	0,001	1,562	2,851	3,595	176,705

Fonte: Autoria própria.

Tabela 24 – Resumo estatístico dos erros de previsão (RMSE e MAE) para múltiplos modelos no cenário de download de arquivo em condição de condução.

Modelo	RMSE					MAE				
	Min	Q1	Mediana	Q3	Máx	Min	Q1	Mediana	Q3	Máx
Naive Mean [Local univariado]	0,665	1,007	1,740	3,046	4,234	0,576	0,759	0,817	0,977	1,924
AutoARIMA [Local univariado]	0,619	1,011	1,748	3,055	4,235	0,458	0,650	0,795	0,913	1,923
LightGBM [Global não covariado]	0,622	1,031	1,790	3,201	4,255	0,453	0,758	0,893	1,294	1,883
LightGBM [Global covariado]	0,631	1,082	1,806	3,197	4,270	0,484	0,767	0,959	1,334	2,054
Random Forest [Global não covariado]	0,681	1,072	1,844	3,374	4,895	0,521	0,761	0,860	1,365	1,977
Block RNN [Global não covariado]	0,673	1,141	1,845	3,258	4,252	0,568	0,927	1,069	1,564	2,535
Transformer [Global não covariado]	0,637	1,143	1,933	3,307	4,276	0,509	0,971	1,114	1,568	2,614
Random Forest [Global covariado]	0,719	1,201	1,933	3,223	4,313	0,557	0,802	1,090	1,437	1,969
Block RNN [Global covariado]	0,704	1,200	1,943	3,264	4,527	0,553	0,998	1,261	1,535	2,784
Linear Regression [Global não covariado]	0,641	1,102	1,955	3,282	4,505	0,512	0,824	1,081	1,465	2,367
Transformer [Global covariado]	0,711	1,156	1,963	3,353	4,488	0,561	1,007	1,185	1,700	2,944
Naive Aggregate [Global não covariado]	0,517	0,926	1,974	3,424	4,463	0,209	0,512	0,709	0,902	1,738
Linear Regression [Global covariado]	0,637	1,076	1,980	3,362	4,818	0,517	0,794	1,002	1,416	3,274
N-BEATS [Global covariado]	0,619	1,187	2,123	3,177	4,449	0,256	0,546	0,848	0,956	1,561
Exponential Smoothing [Local univariado]	0,625	0,998	2,172	3,523	5,731	0,187	0,508	0,745	0,944	1,853
N-BEATS [Global não covariado]	0,759	1,237	2,200	3,245	4,642	0,368	0,829	1,094	1,288	1,804
Naive Seasonal [Global não covariado]	0,782	1,175	2,434	3,821	5,350	0,252	0,558	0,803	0,943	1,595
Naive Drift [Global não covariado]	1,337	1,820	3,541	6,330	9,865	0,429	0,984	1,509	1,973	3,485

Fonte: Autoria própria.