



UNIVERSIDADE FEDERAL DO AMAZONAS - UFAM
INSTITUTO DE COMPUTAÇÃO - ICOMP
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA - PPGI

Um Método para Predição de Taxa de Transferência em
Redes Móveis a partir de Dados Coletados de
Smartphones

Shermam Tácia da Costa Lima

MANAUS-AM

2025

PODER EXECUTIVO
MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DO AMAZONAS
INSTITUTO DE COMPUTAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM
INFORMÁTICA



Um Método para Predição de Taxa de Transferência em Redes
Móveis a partir de Dados Coletados de Smartphones

Shermam Tácia da Costa Lima

Dissertação apresentada ao Programa de Pós-Graduação em Informática, da Universidade Federal do Amazonas, como parte dos requisitos necessários à obtenção do título de Mestre em Informática, na área de Inteligência Artificial e Ciência de Dados.

Orientador: Raimundo da Silva Barreto, D.Sc.

Coorientador: Vandermi João da Silva, D.Sc.

Este trabalho foi desenvolvido com o apoio do Governo do Estado do Amazonas por meio da Fundação de Amparo à Pesquisa do Estado do Amazonas, com a concessão de bolsa de estudo.

MANAUS-AM

2025

Ficha Catalográfica

Elaborada automaticamente de acordo com os dados fornecidos pelo(a) autor(a).

L732m Lima, Shermam Tácia da Costa
Um Método para Predição de Taxa de Transferência em Redes
Móveis a partir de Dados Coletados de Smartphones / Shermam
Tácia da Costa Lima. - 2025.
123 f. : il., color. ; 31 cm.

Orientador(a): Raimundo da Silva Barreto.
Coorientador(a): Vandermi João da Silva.
Dissertação (mestrado) - Universidade Federal do Amazonas,
Programa de Pós-Graduação em Informática, Manaus, 2025.

1. Aprendizado de Máquina. 2. 4G LTE. 3. Taxa de Transferência.
4. Uplink. 5. Predição. I. Barreto, Raimundo da Silva. II. Silva,
Vandermi João da. III. Universidade Federal do Amazonas.
Programa de Pós-Graduação em Informática. IV. Título



Ministério da Educação
Universidade Federal do Amazonas
Coordenação do Programa de Pós-Graduação em Informática

FOLHA DE APROVAÇÃO

"UM MÉTODO PARA PREDIÇÃO DE TAXA DE TRANSFERÊNCIA EM REDES MÓVEIS A PARTIR DE DADOS COLETADOS DE SMARTPHONES"

SHERMAM TÁCIA DA COSTA LIMA

Dissertação de Mestrado defendida e aprovada pela banca examinadora constituída pelos Professores:

Prof. Dr. Raimundo da Silva Barreto - PRESIDENTE

Prof. Dr. Edjair de Souza Mota - MEMBRO INTERNO

Prof. Dr. Walter Charles Sousa Seiffert Simões - MEMBRO EXTERNO

MANAUS, 13 de junho de 2025.



Documento assinado eletronicamente por **Raimundo da Silva Barreto, Professor do Magistério Superior**, em 26/06/2025, às 08:47, conforme horário oficial de Manaus, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Edjair de Souza Mota, Professor do Magistério Superior**, em 15/07/2025, às 18:58, conforme horário oficial de Manaus, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Walter Charles Sousa Seiffert Simões, Usuário Externo**, em 12/08/2025, às 16:58, conforme horário oficial de Manaus, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).

Dedicatória

*Aos meus pais, **Elieni Lima** e **João Lima** (in memoriam), por serem meu alicerce em todos os momentos. Por cada sacrifício silencioso, cada gesto de amor e cada palavra de incentivo. Esta conquista é também de vocês.*

Agradecimentos

Em primeiro lugar, agradeço a Deus por me conceder força, saúde e sabedoria ao longo dessa jornada acadêmica. Sem Ele, nada disso seria possível.

Agradeço imensamente aos meus pais Elieni Lima e João Lima (in memoriam) e aos meus irmãos, que sempre me apoiaram com amor incondicional, incentivo e orações, mesmo à distância ou no silêncio de cada sacrifício. Obrigada por acreditarem em mim e me ensinarem o valor da perseverança. Agradeço ao meu noivo Bruno Rodrigues por todo empenho, pela dedicação e pelas contribuições fundamentais durante o desenvolvimento desta pesquisa, especialmente nos momentos mais desafiadores desta etapa.

Ao meu orientador, professor Raimundo da Silva Barreto, minha sincera gratidão pela orientação firme e cuidadosa durante todo o mestrado. Obrigada por compartilhar sua experiência, acreditar no potencial do trabalho e, principalmente, pela paciência e incentivo constantes. Estendo também meus agradecimentos ao meu coorientador, professor Vandermi João da Silva, pelo suporte técnico e científico durante todo o desenvolvimento desta pesquisa. E ao professor Carlos Alberto Oliveira de Freitas por todo o apoio técnico durante todo o processo do trabalho.

Agradeço aos membros da banca examinadora, por aceitarem participar deste momento tão importante da minha vida acadêmica. Suas contribuições são valiosas e engrandecem este trabalho.

Sou grata ao projeto SW-PerfI, ao qual este trabalho está vinculado, pelas oportunidades de crescimento científico, pelas reuniões de equipe, pelas viagens e pela experiência de atuar em um ambiente de pesquisa tão colaborativo e inovador. Em especial, agradeço aos colegas de grupo e à equipe que me acompanhou nessa jornada.

Agradeço também aos amigos que estiveram ao meu lado nos momentos mais difí-

ceis e aos que vibraram comigo a cada conquista. Vocês tornaram o percurso mais leve e significativo.

Por fim, agradeço a mim mesma, por não desistir mesmo diante das dificuldades, por cada esforço, cada noite mal dormida e por ter acreditado que seria possível. Muito obrigada a todos que, de alguma forma, fizeram parte dessa caminhada.

Esta pesquisa foi realizada com o apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001. Este trabalho contou com apoio parcial da Fundação de Amparo à Pesquisa do Estado do Amazonas (FAPEAM), por meio do projeto POSGRAD. Nos termos dos Arts. 21 e 22 do Decreto nº 10.521/2020, este trabalho foi parcialmente financiado pela Motorola Mobility Comércio de Produtos Eletrônicos Ltda. e pela Flextronics da Amazônia Ltda., conforme previsto na Lei Federal nº 8.387/1991, por meio do Projeto "SWPERFI – Técnicas de Inteligência Artificial para Análise e Otimização de Desempenho de Software", convênio nº 004/2021, firmado com o ICOMP/UFAM.

Você conseguirá tudo o que quiser!

Resumo

A comunicação móvel em ambientes dinâmicos apresenta desafios à garantia da qualidade dos serviços, sendo a vazão de dados um parâmetro crítico de desempenho. Embora a predição da vazão tenha sido amplamente explorada para o *downlink*, a vazão de *uplink* permanece pouco investigada, apesar do crescimento de aplicações que demandam elevados volumes de envio de dados para a rede. Neste contexto, este trabalho propôs uma abordagem baseada em aprendizado de máquina para a predição da vazão de *uplink* em redes 4G LTE, utilizando métricas da camada física como variáveis preditoras. Foi conduzido um experimento de coleta de dados em ambiente urbano real, abrangendo diferentes horários e fluxos de tráfego, utilizando um smartphone *Android* e o aplicativo *Armadeira*. A base de dados construída permitiu o treinamento e a avaliação de cinco modelos de aprendizado de máquina, sendo o algoritmo floresta aleatória com melhor desempenho, com coeficiente de determinação (R^2) de 0,93 e menor erro absoluto médio. Os resultados confirmaram a viabilidade da abordagem, evidenciando que a análise de parâmetros de qualidade de sinal, como RSRP, RSRQ, RSSI e SNR, possibilita a predição precisa da vazão de *uplink*. A metodologia adotada contribui para a replicabilidade do experimento e fornece subsídios para futuras investigações sobre a otimização de redes móveis. Este trabalho avança o conhecimento técnico-científico na predição de vazão em redes móveis, ao disponibilizar uma base inédita, um método validado em ambiente real e práticas que fortalecem a análise da Qualidade de Serviço (QoS) e da Qualidade de Experiência (QoE) em sistemas de comunicação móvel.

Palavras-chave: Aprendizado de Máquina, 4G LTE, Taxa de Transferência, *Uplink*, Predição, Telefonia Móvel.

Abstract

Mobile communication in dynamic environments poses significant challenges to maintaining service quality, with data throughput being a critical performance parameter. Although throughput prediction in mobile networks has been widely explored for downlink, the uplink direction remains underexplored, despite the increasing demand from applications that require large volumes of data to be uploaded to the network. In this context, this work proposes a machine learning-based approach for predicting uplink throughput in 4G LTE networks, using physical layer metrics as predictive variables. A real-world urban data collection experiment was conducted, covering different times of day and traffic conditions, using an Android smartphone and the Armadeira application. The resulting dataset enabled the training and evaluation of five machine learning models, with the Random Forest algorithm achieving the best performance, reaching a coefficient of determination (R^2) of 0.93 and the lowest mean absolute error. The results confirm the feasibility of the proposed approach, showing that the analysis of signal quality parameters, such as RSRP, RSRQ, RSSI, and SNR, enables accurate uplink throughput prediction. The adopted methodology supports the replicability of the experiment and provides a foundation for future research on mobile network optimization. This work advances technical and scientific knowledge on throughput prediction in mobile networks by providing a novel dataset, a validated method in a real environment, and practices that strengthen the analysis of Quality of Service (QoS) and Quality of Experience (QoE) in mobile communication systems.

Keywords: Machine Learning, 4G LTE, Throughput, Uplink, Prediction, Mobile.

Sumário

1	Introdução	15
1.1	Contexto	15
1.2	Definição do Problema	17
1.3	Motivação	18
1.4	Contribuições	20
1.5	Organização do Trabalho	21
2	Fundamentação Teórica	22
2.1	Long-Term Evolution - LTE	22
2.1.1	Arquitetura LTE	24
2.1.2	Parâmetros de Qualidade de Sinal	27
2.1.3	RSSI (Received Signal Strength Indicator)	29
2.1.4	RSRP (Reference Signal Received Power)	30
2.1.5	RSRQ (Reference Signal Received Quality)	32
2.1.6	SNR (Signal to Noise)	33
2.1.7	Uplink	34
2.1.8	Downlink	35
2.2	Aprendizado de Máquina	36
2.2.1	Tipos de Métodos de Aprendizagem	37
2.2.2	Métodos Preditivos	40
2.2.3	Métricas de Avaliação	45
2.3	Resumo	46

3	Trabalhos Correlatos	47
3.1	Revisão Sistemática da Literatura	47
3.1.1	Protocolo de Revisão Sistemática da Literatura	47
3.1.2	Condução da Revisão Sistemática da Literatura	52
3.2	Trabalhos Correlatos como Resultado Revisão Sistemática da Literatura	55
3.2.1	Ambiente Controlado	55
3.2.2	Experimentos Realizados em Câmara Anecoica	55
3.2.3	Predição de Vazão em Ambiente de Laboratório	57
3.2.4	Métodos de Predição em Dispositivos de Redes Sem Fio Baseados em Aprendizado de Máquina	57
3.2.5	Ambiente Real	58
3.2.6	Aprendizado Supervisionado com Regressão	59
3.2.7	Regressão e Classificação	60
3.2.8	Regressão e Series Temporais	62
3.2.9	Otimização de Taxa de Uplink	68
3.2.10	Uso de KPI para Estimativa de Outros Casos	68
3.3	Comparativo dos Trabalhos Correlatos	69
3.4	Resumo	74
4	Método Proposto	75
4.1	Visão Geral do Método Proposto	75
4.1.1	Coleta de Dados	76
4.1.2	Pré-processamento	78
4.1.3	Geração do Modelo	81
4.2	Resumo	83
5	Resultados e Discussões	84
5.1	Protocolo de Experimento	84
5.1.1	Tecnologias Utilizadas	85
5.1.2	Aplicativo para Coleta de Dados	86
5.1.3	Fluxo de Coleta de Dados	87

5.2	Execução do Experimento	89
5.2.1	Base de Dados	90
5.2.2	Análise Exploratória dos Dados	92
5.2.3	Limitação do Experimento	101
5.3	Aplicação do Modelo de Predição de Uplink	102
5.3.1	Análise de Correlação	102
5.3.2	Discussão dos Resultados da Predição da Vazão de Uplink	105
5.4	Resumo do Capítulo	108
6	Conclusão	109
6.1	Considerações Finais	109
6.2	Trabalhos Futuros	111
6.3	Principais Contribuições	111
6.4	Produção Científica	112
	Referências Bibliográficas	113
A	Descrição dos Dados Coletados	124

Lista de Figuras

2.1	arquitetura de rede LTE	25
2.2	operação de handover em redes LTE	28
2.3	tipos de aprendizagem de máquina	38
2.4	exemplo de aprendizado não supervisionado, clusterização	39
3.1	distribuição do número de artigos aceitos ao longo das etapas da revisão sistemática da literatura	53
3.2	lista dos comparativos dos trabalhos correlatos - parte 1	72
3.3	lista dos comparativos dos trabalhos correlatos - parte 2	73
4.1	visão geral do método proposto	76
5.1	tela home do aplicativo	87
5.2	tela de dados do aplicativo	87
5.3	armazenamento dos dados	88
5.4	arquivo.csv dos dados	88
5.5	fluxo de coleta de dados	89
5.6	rota da coleta de dados da cidade	90
5.7	base de dados para análise exploratória dos dados	92
5.8	quantidade da coleta por data	93
5.9	quantidade da coleta por período	93
5.10	representações gráficas do parâmetro de qualidade de rede RSRP	94
5.11	representações gráficas do parâmetro de qualidade de rede RSRQ	96
5.12	representações gráficas do parâmetro de qualidade de rede SNR	97

5.13	representações gráficas do parâmetro de qualidade de rede RSSI	98
5.14	plotagem das coordenadas de latitude e longitude da base de dados coletada no dia 24.03.2023	99
5.15	plotagem das coordenadas de localização com relação ao RSRP	100
5.16	matriz de correlação de Spearman destacando a relação entre os parâmetros de qualidade da rede e a vazão de uplink em redes 4G LTE	103
5.17	correlação entre parâmetros de qualidade da rede e vazão de <i>uplink</i>	104
5.18	comparação entre valores observados e previstos para diferentes algoritmos de aprendizado de máquina	106

Lista de Tabelas

2.1	tabela de valores de RSSI em redes LTE	30
2.2	tabela de valores de RSRP em redes LTE	31
2.3	tabela de valores de RSRQ em redes LTE	32
2.4	tabela de valores de SNR em redes LTE	33
3.1	metodologia PIO aplicado no contexto da RSL	49
3.2	expressão final da busca aplicada na base Scopus em Setembro de 2024	50
3.3	artigos selecionados com base nos resultados da busca, totalizando 23 artigos, com identificação, título, autores e ano de publicação	54
5.1	ferramentas e recursos utilizados na aplicação do método proposto	85
5.2	parâmetros coletados pelo aplicativo durante o experimento	91
5.3	estatísticas do RSRP	95
5.4	estatísticas do RSRQ	96
5.5	estatísticas do SNR	97
5.6	estatísticas do RSSI	98
5.7	resultados finais da predição da vazão de <i>uplink</i> utilizando diferentes algoritmos de aprendizado de máquina.	107
A.1	lista dos parâmetros coletados pelo aplicativo desenvolvido (Armadeira)	125

Lista de Abreviaturas e Siglas

1G	<i>Primeira Geração</i>
2G	<i>Segunda Geração</i>
3G	<i>Terceira Geração</i>
3GPP	<i>Third Generation Partnership Project</i>
4G	<i>Quarta Geração</i>
5G	<i>Quinta Geração</i>
6G	<i>Sexta Geração</i>
7G	<i>Sétima Geração</i>
AER	<i>Average Error Rate</i>
CDMA	<i>Code Division Multiple Access</i>
CNNs	<i>Convolutional Neural Networks</i>
CQI	<i>Channel Quality Indicator</i>
dB	<i>Decibéis</i>
dBm	<i>Decibéis Miliwatts</i>
DT	<i>Decision Trees</i>
eMBB	<i>Enhanced Mobile Broadband</i>
Enodebs	<i>Evolved NodeB</i>
EPC	<i>Evolved Packet Core</i>
E-UTRAN	<i>Evolved-Universal Terrestrial Radio Access Network</i>
FDD	<i>Frequency Division Duplex</i>
Gbps	<i>Gigabit por segundo</i>
GMDH	<i>Group Method of Data Handling</i>

GSM	<i>Global System for Mobile Communications</i>
HMM	<i>Hidden Markov Model</i>
HSS	<i>Home Subscription Server</i>
IA	<i>Inteligência Artificial</i>
IBGE	<i>Instituto Brasileiro de Geografia e Estatística</i>
IoT	<i>Internet of Things</i>
IP	<i>Internet Protocol</i>
ITS	<i>Intelligent Transportation System</i>
Kbps	<i>kilobits por segundo</i>
KNN	<i>K-Nearest Neighbors</i>
LS	<i>Last Sample</i>
LTE	<i>Long Term Evolution</i>
M2M	<i>Comunicação de Máquina para Máquina</i>
MAE	<i>Mean Absolute Error</i>
MAPE	<i>Mean Absolute Percentage Error</i>
Mbps	<i>Megabits por segundo</i>
MIMO	<i>Multiple Input Multiple Output</i>
ML	<i>Machine Learning</i>
MLR	<i>Multiple Linear Regression</i>
MME	<i>Mobility Management Entity</i>
mMTCs	<i>Massive Machine Type Communications</i>
MOS	<i>Mean Opinion Score</i>
MSE	<i>Mean Squared Error</i>
NRMSE	<i>Normalized Root Mean Square Error</i>
OFDM	<i>Orthogonal Frequency-Division Multiplexing</i>
OFDMA	<i>Orthogonal Frequency Division Multiple Access</i>
PCRF	<i>Policy and Charging Resource Function</i>
P-GW	<i>Packet Data Network Gateway</i>
PIO	<i>População, Intervenção, Resultado</i>
PSNR	<i>Peak Signal-To-Noise Ratio</i>

QoE	<i>Quality of Experience</i>
QoS	<i>Quality of Service</i>
R²	<i>Coefficient of Determination</i>
RB	<i>Resource Block</i>
RF	<i>Random Forest</i>
RNA	<i>Redes Neurais Artificiais</i>
RNN	<i>Recurrent Neural Network</i>
RSL	<i>Revisão Sistemática da Literatura</i>
RSRP	<i>Reference Signal Received Power</i>
RSRQ	<i>Reference Signal Received Quality</i>
RSSI	<i>Received Signal Strength Indicator</i>
RTT	<i>Round Trip Time</i>
SC-FDMA	<i>Single Carrier Frequency Division Multiple Access</i>
S-GW	<i>Serving Gateway</i>
SMS	<i>Short Message Service</i>
SNR	<i>Signal-to-Noise Ratio</i>
SVM	<i>Support Vector Machine</i>
SVR	<i>support vector regression</i>
Tbps	<i>Terabits por segundo</i>
TDD	<i>Time Division Duplex</i>
UE	<i>User Equipment</i>
UIT	<i>União Internacional De Telecomunicações</i>
URLLCs	<i>Ultra-Reliable Low-Latency Communications</i>
V2V	<i>Veículo para Veículo</i>
VoIP	<i>Voice Over Internet Protocol</i>
WCDMA	<i>Wideband Code Division Multiple Access</i>

Capítulo 1

Introdução

Neste capítulo são abordados o contexto e a descrição do problema, além da motivação desta pesquisa. Também são apresentados os objetivos estabelecidos e a organização deste trabalho.

1.1 Contexto

As redes de banda larga móvel desempenham um papel vital em inúmeras operações da sociedade moderna e estão rapidamente se tornando a base fundamental da infraestrutura de comunicações do mundo contemporâneo. Com uma ampla utilização de dispositivos móveis, como, por exemplo, smartphones e tablets, e o acesso a redes de alta capacidade, houve um crescimento exponencial no uso das redes de banda larga móvel (Polak *et al.* 2025; Al-Thaedan *et al.* 2024).

Atualmente é irrefutável a importância dos meios de comunicação no cotidiano dos indivíduos. No Brasil, de acordo com o Instituto Brasileiro de Geografia e Estatística (IBGE), existem mais smartphones do que pessoas, visto que são 242 milhões de smartphones em uso em um país que possui aproximadamente 214 milhões de habitantes (Meirelles 2024).

Com o aumento da frequência do uso dos smartphones, é preciso ressaltar a importância da conectividade das redes sem fio, pois é o meio mais utilizado para acessar a web e fazer as atividades *online*. Tendo como foco a rede móvel, faz-se necessário entender um pouco de sua história e evolução, cujo início ocorreu a partir do surgimento da primeira

geração (1G), através da qual foi possível realizar chamadas em movimento, mas com pouca eficiência e custo elevado. A segunda geração (2G) tornou possível uma melhor locomoção, uma vez que as operadoras passaram a oferecer terminais mais práticos, fáceis de transportar e com preços inferiores, como também a oferta de novos serviços. Já no que se refere à terceira geração (3G), surge a possibilidade de acesso à Internet em ritmo de transmissão próximo aos obtidos em redes de comunicação fixa. Por fim, com a quarta geração (4G), através das redes de comunicação móveis, pode-se alcançar uma velocidade de qualidade semelhante a uma rede fixa. A próxima tecnologia de rede, ainda em implantação no Brasil, é a rede da quinta geração (5G), que visa interconectar bilhões de dispositivos inteligentes de forma autônoma, garantindo segurança, cobertura e confiabilidade (Nikula *et al.* 2025; Batool *et al.* 2024; Sathya *et al.* 2023).

A tecnologia 4G é padronizada como *Long Term Evolution* (LTE) pelo *Third Generation Partnership Project* (3GPP). Essa tecnologia opera em altas taxas de transferência e baixa latência de rede. O desafio da evolução da rede móvel sempre existirá, diante da necessidade de expandir a capacidade da rede e das demandas crescentes dos usuários que esperam maior qualidade no serviço prestado. Destaca-se que é improvável que a internet móvel interrompa seu processo de evolução ou perca relevância no cenário das comunicações (Gautam e Dharmaraja 2023).

Em um mundo globalizado, o crescente tráfego de dados requer uma evolução contínua da tecnologia, principalmente da telefonia móvel. A comunicação é uma área de grande importância para a continuidade e aceleração do mundo atual, e a tecnologia móvel também segue essa tendência, o que é possível verificar na explosão de utilização de telefones celulares, tanto para voz quanto para a crescente troca de dados através da internet móvel (Markappa *et al.* 2024; Kousias *et al.* 2019).

Diante disso, o estudo sobre o tráfego de dados nos dispositivos móveis pode trazer soluções que melhorem a qualidade de transmissão de dados, que evitem interrupções e falhas na conexão, e possibilitem suportar novas aplicações que exigem maiores taxas de transferência. Pode ainda aprimorar a experiência do usuário, proporcionando uma navegação mais fluida e uma maior satisfação com os serviços prestados.

1.2 Definição do Problema

Com o avanço tecnológico, a computação móvel tornou-se foco principal das comunidades científica e empresarial desde a introdução do smartphone que suporta redes móveis 3G e 4G para transmissão de dados. A sociedade agora pode se comunicar a qualquer hora e de qualquer lugar através da facilidade de se conectar à Internet e utilizar os recursos multimídia disponíveis e avançados na rede (Polak *et al.* 2025; Rida 2021).

Mundialmente, vários dispositivos ficam conectados à rede, como, por exemplo, os dispositivos instalados em veículos conectados. Esses dispositivos podem compartilhar informações como situação da estrada, condições climáticas, vaga de estacionamento, pedágios e informações de tráfego. Outra tecnologia é a utilizada nos dispositivos de sensor fitness, que interagem com aplicativos e permitem que os usuários compartilhem dados de condicionamento físico com amigos ou visualizem a análise dos próprios dados (Markappa *et al.* 2024; Amin *et al.* 2022).

Prevê-se um aumento significativo da demanda por serviços e aplicativos em cidades inteligentes, onde tudo está conectado, a qualquer hora e em qualquer lugar. A maioria desses aplicativos depende de redes sem fio para enviar seus dados aos servidores de internet remotos correspondentes. O *Intelligent Transportation System* (ITS) destaca vários serviços que dependem do sistema de comunicação do carro, gerando uma grande quantidade de dados a serem gerenciados por essas redes. Videoconferência veicular, monitoramento de tráfego e upload de dados feitos pela rede celular são alguns serviços que exemplificam esse cenário (Polak *et al.* 2025).

De acordo com uma análise de dados automotivos, espera-se que um carro conectado produza uma média diária de até 30 terabytes de dados, os quais serão enviados aos servidores do ITS. Portanto, diferentemente das aplicações de comunicação Humano-Humano, nas quais a transmissão *downlink* é predominante, os serviços de cidades inteligentes usam principalmente a comunicação *uplink*, invertendo os papéis convencionais. Quando todos os nós conectados estiverem enviando suas informações, uma quantidade significativa de dados deverá ser recebida por seu servidor correspondente, com alta qualidade de serviço (Palaios *et al.* 2023; Oussakel *et al.* 2019b).

Esse grande crescimento da utilização da rede por diversos dispositivos móveis evidencia o impacto dos usuários no sistema e, conseqüentemente, a necessidade de melhorias constantes na infraestrutura de redes das operadoras. Impor qualidade de serviço em tal rede móvel é um desafio, pois os fenômenos de rádio, bem como a mobilidade e a dinâmica do usuário e seus dispositivos, não são controlados (Biswas *et al.* 2024; Oussakel *et al.* 2019b).

As redes móveis evoluem em um ritmo acelerado. A cada nova geração lançada, é oferecida uma maior quantidade de largura de banda e menor latência, possibilitando novos serviços e atendendo às necessidades e expectativas do usuário final. Por exemplo, a indústria de transporte está aprimorando seus veículos com recursos de segurança que incorporam dados altamente precisos de um servidor centralizado. Ao mesmo tempo, os passageiros exigem a capacidade da rede para suas necessidades pessoais, como *streaming* de mídia. No entanto, a largura de banda é um recurso limitado que pode ser facilmente excedido em locais com tráfego intenso, especialmente em centros de transporte (Nikula *et al.* 2025; Schmid *et al.* 2019).

Problemas comuns relacionados ao acesso limitado ao link de dados da rede móvel podem ser solucionados por meio da predição da taxa de transferência, também conhecida como vazão de dados disponível durante a condução. Essa predição pode ser utilizada para ajustar a alocação de largura de banda com base nas prioridades das mensagens ou para pré-carregar dados quando a capacidade estiver disponível (Schmid *et al.* 2019).

Diante disso, surge a oportunidade de estudar sobre a comunicação móvel para verificar a vazão de rede 4G LTE dos dispositivos móveis. O problema considerado neste trabalho pode ser expresso a partir da seguinte questão: é possível construir um modelo para predição de vazão de dados de *uplink* a partir de informações da rede de telefonia móvel obtidas através das camadas inferiores de smartphones?

1.3 Motivação

Nos últimos anos, a predição de vazão na direção *downlink* ganhou destaque, sendo considerada um componente essencial das futuras gerações de redes móveis, conforme indicado por Kousias *et al.* (2019). Dessa forma, a maioria dos estudos em redes 4G LTE tem direcionado

seus esforços à estimativa da taxa de *download*, isto é, o fluxo de dados do servidor para o dispositivo do usuário. Em contraste, há uma escassez de pesquisas voltadas à análise da vazão no sentido *uplink*, que corresponde ao envio de informações do terminal do usuário para os servidores remotos (Chmieliauskas e Paulikas 2025; Yammanuru e Das 2020).

A motivação para estudar a predição de vazão em redes 4G LTE está relacionada a vários fatores, como mudanças no padrão de uso da internet. Nos últimos anos, o uso da internet tem se diversificado com o surgimento de aplicativos e serviços que propiciam uma alta taxa de *uplink*. Algumas aplicações, como as transmissões de vídeo em tempo real, jogos online e aplicativos de compartilhamento de arquivos, geram alto volume de tráfego de dados de *uplink* e exigem uma rede de telecomunicações robusta e eficiente para suportar essas atividades.

Além disso, a tecnologia 4G LTE tem um potencial significativo para impulsionar o crescimento de novas aplicações e serviços inovadores como a *Internet of Things* (IoT), a comunicação de veículo para veículo (V2V) e a comunicação de máquina para máquina (M2M), que dependem fortemente da comunicação de *uplink*. A predição de vazão de *uplink* em redes 4G LTE é, portanto, importante para garantir a viabilidade e a qualidade dessas novas aplicações e serviços (Palaios *et al.* 2023; Gaber *et al.* 2020).

Há um grande potencial em utilizar a predição de vazão para otimizar o desempenho de diversos aplicativos, especialmente os serviços de *streaming* multimídia adaptativos. Estes serviços possuem a capacidade de aprimorar a experiência do usuário, tomando decisões precisas e oportunas sobre a adaptação da qualidade de transmissão. É possível monitorar e ajustar a qualidade da transmissão em tempo real, considerando as condições da rede e do dispositivo do usuário, o que garante uma reprodução de vídeo suave e com qualidade. De forma semelhante, *downloads* em grande escala, como atualizações *Over-the-air*¹ de *firmware* para carros autônomos, podem ser agendadas com eficiência sem congestionar a rede, graças à predição de vazão (Chmieliauskas e Paulikas 2025; Raca *et al.* 2020).

Em aplicações de IoT, a predição de vazão é especialmente crucial, já que os dispositivos geralmente possuem limitações de recursos de energia e rede. Através da predição, é

¹*Over-the-air* refere-se à capacidade de transferir dados, como atualizações de software ou *firmware*, por meio de uma conexão sem fio, sem a necessidade de cabos ou conexões físicas diretas.

possível permitir que esses dispositivos operem de forma mais eficiente, prolongando a vida útil da bateria e reduzindo a necessidade de transmissões desnecessárias ou repetidas. Como resultado, a predição pode ter um impacto significativo na eficiência e no desempenho de várias aplicações em dispositivos móveis e de IoT (Nikula *et al.* 2025; Salahdine *et al.* 2023).

A predição de vazão de *uplink* em redes 4G LTE é importante para garantir a qualidade de serviço e atender às demandas crescentes de tráfego de dados de *uplink*, impulsionando o crescimento de novas aplicações e serviços inovadores na era da tecnologia móvel avançada.

Portanto, o estudo da predição de vazão de *uplink* busca desenvolver técnicas e modelos eficazes para estimativas sob diferentes condições de rede e de dispositivo. A capacidade de antecipar com precisão a taxa de *uplink* pode auxiliar os provedores de serviço a otimizar o desempenho da infraestrutura e aprimorar a experiência do usuário, especialmente em cenários de alto tráfego. Dadas as características dinâmicas dessas redes, estimar a vazão em ambientes móveis constitui uma tarefa complexa e repleta de desafios significativos (Markappa *et al.* 2024; Wei *et al.* 2018).

1.4 Contribuições

Esta dissertação apresenta como contribuição principal o desenvolvimento de um modelo para a predição da vazão de *uplink* em redes de telefonia móvel, utilizando dados coletados por smartphones. Além disso, realiza-se uma análise comparativa com trabalhos relevantes da literatura, a fim de contextualizar os resultados obtidos.

As contribuições específicas deste trabalho podem ser assim destacadas:

1. Coleta de dados referentes a parâmetros da camada física de redes móveis, por meio de dispositivo portátil;
2. Identificação dos parâmetros que apresentam correlação significativa com a vazão de *uplink*;
3. Construção de um modelo preditivo para prever a vazão de *uplink*;

4. Avaliação experimental da solução proposta, com análise de sua eficácia e comparação com abordagens existentes.

1.5 Organização do Trabalho

A introdução deste trabalho, apresentada no [Capítulo 1](#), traz a contextualização do tema, a definição do problema, as motivações da pesquisa, os objetivos da dissertação e, por fim, a estrutura adotada.

O [Capítulo 2](#) contempla a fundamentação teórica, abordando os principais conceitos relacionados à comunicação móvel, à evolução das gerações de redes celulares, à tecnologia LTE e às técnicas de aprendizado de máquina, com o intuito de fornecer os conhecimentos essenciais para o desenvolvimento do trabalho.

O [Capítulo 3](#) apresenta os trabalhos correlatos identificados por meio de uma revisão sistemática da literatura. Os estudos analisados são classificados em duas abordagens: experimentos conduzidos em ambientes reais e aqueles realizados em ambientes controlados, permitindo identificar lacunas e contribuições existentes na área.

O [Capítulo 4](#) descreve o método empregado nesta dissertação, abrangendo a visão geral da abordagem, o processo de coleta de dados, as etapas de pré-processamento e a construção dos modelos preditivos baseados em métricas da camada física da rede. O [Capítulo 5](#) apresenta os experimentos realizados, incluindo o protocolo de execução, as ferramentas utilizadas, o ambiente de coleta, bem como os resultados obtidos por meio da aplicação e avaliação dos algoritmos de aprendizado de máquina.

Por fim, o [Capítulo 6](#) discute as conclusões finais da dissertação, com ênfase nas contribuições alcançadas, na eficácia do método desenvolvido e nas possibilidades de pesquisas futuras na área de predição de vazão em redes móveis.

Capítulo 2

Fundamentação Teórica

Este capítulo apresenta os conceitos fundamentais abordados neste trabalho, a fim de proporcionar um embasamento teórico que permita a compreensão de termos da arquitetura LTE, aprendizado de máquina e alguns algoritmos para predição.

2.1 Long-Term Evolution - LTE

Comumente conhecido como rede de quarta geração (4G), é um padrão especificado pelo 3GPP para interface de redes celulares. O 3GPP é formado por empresas e laboratórios de pesquisas que são responsáveis não apenas pela especificação, mas também por identificar, reportar e sugerir soluções para a tecnologia de rede de smartphone, com objetivo de buscar fazer frente às necessidades futuras (Ismail *et al.* 2024; Salahdine *et al.* 2023; Carvalho *et al.* 2019).

O objetivo do 4G LTE foi melhorar a velocidade, a capacidade e a eficiência da rede de telefone celular em relação à sua antecessora, a tecnologia 3G. O 4G LTE permite que os usuários de telefones celulares acessem a Internet com maior rapidez e estabilidade, além de possibilitar o uso de aplicativos e serviços que exigem maior largura de banda e menor latência, como videochamadas, *streaming* de vídeo e jogos online em tempo real. Com o 4G LTE, as operadoras de telefonia móvel podem atender a um número maior de usuários simultaneamente, oferecendo serviços mais rápidos e eficientes (Hajlaoui *et al.* 2019).

A comutação de pacotes do 4G é uma técnica de transmissão de dados que divide as

informações em pequenos pacotes e os envia pela rede de forma independente. Cada pacote é roteado de forma otimizada através da rede, para garantir que chegue ao destino o mais rápido possível. Isso difere da comutação de circuitos usados em tecnologias anteriores, que exigia a criação de um caminho físico dedicado entre os dispositivos que se comunicavam, mesmo que a comunicação não fosse contínua (Ismail *et al.* 2024; Miyim e Wakili 2019).

Com a comutação de pacotes, o 4G consegue oferecer uma rede mais eficiente e flexível, já que os recursos da rede são alocados conforme a demanda em tempo real. Além disso, a comutação de pacotes permite que o mesmo canal de comunicação seja usado por vários usuários simultaneamente, maximizando a eficiência da rede e garantindo que todos os usuários tenham acesso rápido e confiável à Internet móvel (Hajlaoui *et al.* 2019; Krasniqi *et al.* 2018). Abaixo estão listadas algumas características da rede 4G LTE:

- Tecnologia *Multiple Input Multiple Output* (MIMO) é uma técnica usada para melhorar a qualidade e velocidade do sinal sem fio. Ela permite que os dispositivos móveis se comuniquem com várias antenas ao mesmo tempo, usando várias vias de comunicação para transmitir e receber dados (George *et al.* 2015).
- O uso de modulação *Orthogonal Frequency-Division Multiplexing* (OFDM) para melhorar a eficiência do espectro de frequência é dotado de novas tecnologias de codificação para aumentar a eficiência da transmissão de dados. Estas melhorias permitem que o 4G ofereça velocidades mais rápidas e uma melhor qualidade de conexão em relação às tecnologias anteriores (Hajlaoui *et al.* 2019; Sezgin *et al.* 2018).
- Compatibilidade com outros padrões de redes, como o *Global System for Mobile Communications* (GSM) e o *Universal Mobile Telecommunications System* (UMTS), tecnologias utilizadas por muitas redes de telefonia móvel em todo o mundo. A compatibilidade com o GSM permitiu que os usuários de telefones celulares pudessem fazer chamadas e enviar mensagens de texto em áreas onde não havia cobertura 4G (Hajlaoui *et al.* 2019).
- Possibilidade de configuração da largura de banda, permitindo que as operadoras possam alocar diferentes larguras de banda para diferentes usuários, dependendo de suas

necessidades. Isso pode ajudar a melhorar a eficiência do espectro, bem como fornecer um melhor desempenho para os usuários [Sezgin et al. \(2018\)](#)

- A capacidade de operar nos modos *Frequency Division Duplex* (FDD) e *Time Division Duplex* (TDD). No modo FDD, as frequências de transmissão e recepção são separadas, permitindo a transmissão e recepção simultânea de dados. No modo TDD, a transmissão e recepção de dados ocorrem em intervalos de tempo alternados. Isso permite que a largura de banda seja alocada de forma dinâmica entre transmissão e recepção, proporcionando maior flexibilidade na utilização do espectro de frequência. A capacidade de operar nos modos FDD e TDD permite que o 4G seja mais adaptável às diferentes necessidades das operadoras em diferentes regiões e ambientes ([George et al. 2015](#)).

2.1.1 Arquitetura LTE

A arquitetura de rede utilizada no LTE é considerada mais simples do que as utilizadas em outras redes 3GPP, pois os pacotes são processados e gerenciados no núcleo *Evolved Packet Core* (EPC), tornando-os mais rápidos e, com isso, melhorando a taxa de transmissão e o tempo de latência ([Ismail et al. 2024](#); [Cao et al. 2014](#)).

A arquitetura LTE é dividida em quatro domínios principais de alto nível sendo: *User Equipment* (UE), *Evolved-Universal Terrestrial Radio Access Network* (E-UTRAN), *Evolved Packet Core* (EPC) e o *service domain* ([Miyim e Wakili 2019](#)), como mostra a [Figura 2.1](#).

O UE é o dispositivo móvel pelo qual o usuário se conecta na estação rádio base para comunicação fim-a-fim. Podem ser diversos dispositivos como, por exemplo, aparelhos celulares, notebooks, tablets e outros ([Gautam e Dharmaraja 2023](#)).

O E-UTRAN é a composição de toda estrutura da interface aérea para interligação terrestre da rede ao lado das *evolved NodeB* (eNodeBs), responsável pela transmissão ao nível de hardware pelos quais estão interconectados através das interfaces X2 ([Gautam e Dharmaraja 2023](#); [Souza et al. 2022](#)).

A interface X2 é uma interface de comunicação que conecta duas estações-base de E-UTRAN ou eNodeBs. Ela é usada para suportar funções de *handover* inter-eNodeB, ou seja, para transferir a conexão do usuário de uma estação-base para outra durante o movimento do

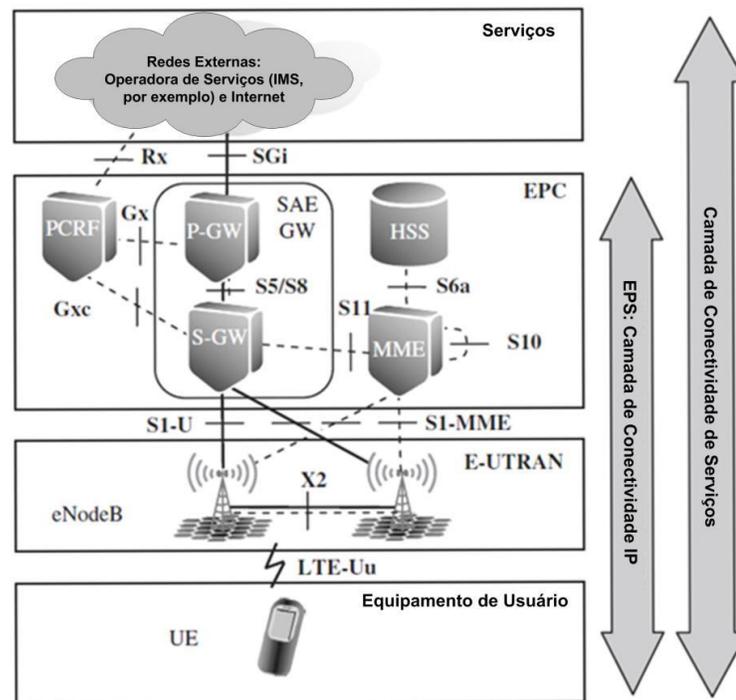


Figura 2.1: arquitetura de rede LTE
 Fonte: adaptado de [Holma e Toskala \(2011\)](#).

usuário. A interface X2 permite que as estações-base troquem informações importantes, como informações de medição, informações de configuração, informações de identidade e controle de recursos de rádio. Isso ajuda a garantir uma conexão de alta qualidade e confiabilidade para os usuários em movimento ([Gautam e Dharmaraja 2023](#)).

Um eNodeB é uma estação rádio base que controla todas as funções relacionadas ao rádio na parte fixa do sistema. As estações-base, como eNodeB, são normalmente distribuídas por toda a área de cobertura da rede, com cada eNodeB próximo à antena de rádio real. Funcionalmente, o eNodeB atua como uma ponte de camada 2 entre o UE e o EPC, é o ponto de terminação de todos os protocolos de rádio para o UE e retransmite dados entre a conexão de rádio e a conexão baseada em *Internet Protocol* (IP) correspondente ao EPC ([Prajwala et al. 2018](#); [Baxhaku et al. 2018](#)).

O núcleo da rede que gerencia toda parte de troca de sinalização e controle de fluxo de dados é chamado de EPC. É uma rede centralizada no núcleo da rede de telefonia móvel, responsável por gerenciar o tráfego de dados dos dispositivos móveis. Ele fornece conectividade entre o dispositivo móvel e os serviços externos, como a Internet ([Hajlaoui](#)

et al. 2019; Prajwala *et al.* 2018).

Policy and Charging Resource Function (PCRF): é um elemento central da rede de comunicação de dados do sistema LTE. É responsável por controlar a política de *Quality of Service* (QoS) e gerenciar as funções de cobrança na rede. Ele interage com o *Serving Gateway* (S-GW) para controlar o fluxo de tráfego na rede, definindo prioridades e limites de uso de dados para cada usuário. O PCRF também gerencia o registro do usuário e as informações de cobrança a ele relacionadas, garantindo a conformidade com as políticas definidas pelos provedores de serviços (Gautam e Dharmaraja 2023; Hajlaoui *et al.* 2019).

Packet Data Network Gateway (P-GW): é um componente-chave da rede LTE, que atua como ponto de entrada e saída da rede de pacotes. O P-GW é responsável pela recepção móvel de endereços IP aos dispositivos, bem como pelo gerenciamento de sessões de rede e de políticas de cobrança. Além disso, é responsável por realizar funções de policiamento e filtragem de tráfego, garantindo que os pacotes de dados sejam entregues com segurança e eficiência (Mauro e Liotta 2020; Sezgin *et al.* 2018).

Serving Gateway (S-GW): atua como ponto de ancoragem para o tráfego de pacotes de dados e é responsável pelo encaminhamento do tráfego de dados entre o eNodeB e o P-GW. Além disso, o S-GW também é responsável pela comutação de pacotes de dados entre as células da rede LTE. Executa funções de gerenciamento de mobilidade, como gerenciamento de *handover*, login de usuários e provisionamento de QoS para diferentes tipos de tráfego (Yammanuru e Das 2020).

Home Subscription Server (HSS): é um servidor que reúne dados dos usuários, atuando como um banco de informações que define quais serviços cada assinante pode acessar. Essa base concentra os dados de perfil, incluindo autenticação, autorização e detalhes de cobrança. Ele gerencia o processo de login, a atribuição de endereços IP e o controle de acesso dos usuários conectados à rede LTE. Além disso, o HSS mantém registros sobre os serviços contratados, políticas tarifárias e recursos de rede, contribuindo para a garantia da qualidade e da segurança do sistema. (Mauro e Liotta 2020; Baxhaku *et al.* 2018).

Mobility Management Entity (MME): sua função principal é gerenciar a mobilidade do usuário em toda a rede, incluindo a autenticação e a seleção do ponto de acesso. Também é responsável por estabelecer sessões de dados e liberá-los quando não forem mais necessários,

além de manter informações de localização do usuário e gerenciar as configurações de segurança. O MME é geralmente associado ao S-GW, que é o elemento que gerencia o tráfego de dados da rede móvel. Juntos, o MME e o S-GW são responsáveis por garantir que os usuários possam se mover livremente entre as células da rede sem continuidade ou perda de conexão (Yammanuru e Das 2020; Sezgin *et al.* 2018).

A última camada é o *service domain*. Entre as funcionalidades disponíveis incluem: serviços de voz, dados e multimídia fornecidos aos usuários finais. Com a arquitetura do LTE, os serviços são fornecidos de forma mais eficiente e escalonável, permitindo maior largura de banda e melhor qualidade de serviço para os usuários finais. Como exemplo disto, os serviços de voz sobre IP, *Voice over Internet Protocol* (VoIP): permitem que os usuários façam chamadas de voz pela rede de dados em vez da rede de voz tradicional. Isso permite chamadas de voz mais baratas e com melhor qualidade de som. Outro exemplo é o streaming de vídeo: a largura de banda mais alta e menor latência da rede 4G LTE permitem transmitir vídeos de alta qualidade em dispositivos móveis (Miyim e Wakili 2019; Kassim *et al.* 2017).

2.1.2 Parâmetros de Qualidade de Sinal

O parâmetro de qualidade de sinal em redes LTE é importante porque afeta diretamente a experiência do usuário ao utilizar serviços de dados móveis. Quando o sinal é fraco ou de baixa qualidade, o dispositivo móvel pode ter dificuldade em se conectar à rede, transmitir ou receber dados, ou pode, ainda, experimentar uma conexão lenta e instável.

A qualidade do sinal também pode afetar a capacidade de um dispositivo de realizar chamadas de voz ou vídeo com clareza e sem interrupções. Portanto, para garantir uma experiência de usuário satisfatória, é importante que a qualidade do sinal seja mantida em um nível alto e consistente.

Os parâmetros de qualidade de sinal, como *Received Signal Strength Indicator* (RSSI), *Reference Signal Received Power* (RSRP), *Reference Signal Received Quality* (RSRQ), *Signal-to-Noise Ratio* (SNR) e *Throughput*, ajudam a monitorar a qualidade da rede LTE e identificar possíveis problemas como áreas com baixa cobertura ou interferências de sinal. As operadoras podem usar esses parâmetros para ajustar e otimizar suas redes LTE para fornecer melhor cobertura e qualidade de sinal aos usuários. Isso pode incluir a instalação de

novas antenas, a adição de mais frequências ou a melhoria da cobertura em áreas com baixa qualidade de sinal. Esses parâmetros são essenciais para as operações de *handover* (Polak *et al.* 2025; Aprillia *et al.* 2020).

A Figura 2.2 aborda o funcionamento do *handover*, que é o processo de transferir uma chamada ou sessão de dados de uma célula para outra quando o dispositivo móvel se desloca para uma área de cobertura diferente. Seu objetivo é manter a qualidade da conexão e minimizar a sobrecarga do serviço durante o deslocamento entre células. Nesse contexto, o dispositivo móvel mede a qualidade do sinal e seleciona a célula de destino com base nos critérios definidos pela operadora. (Krasniqi *et al.* 2018).

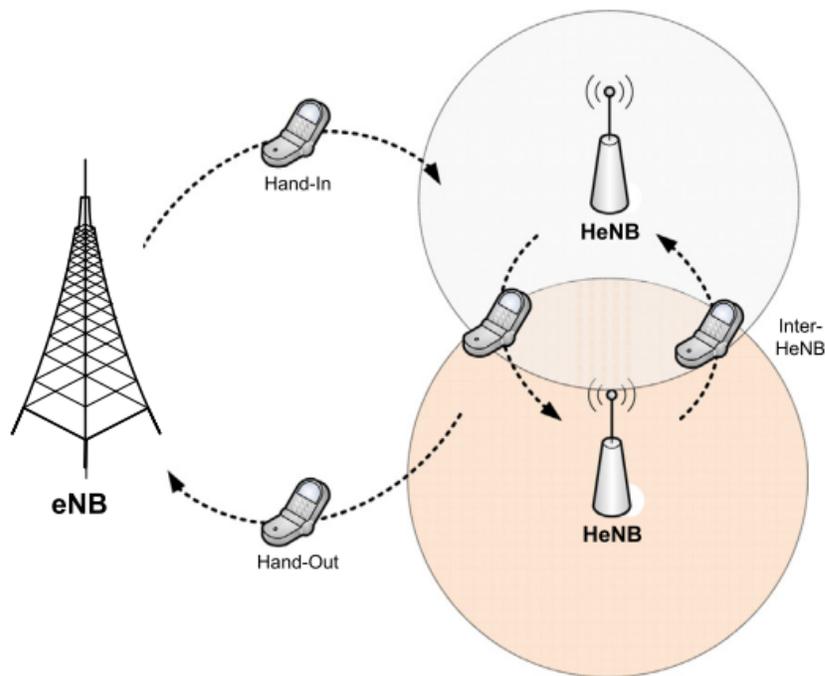


Figura 2.2: operação de handover em redes LTE

Fonte: Gódor *et al.* (2015).

O dispositivo móvel sempre está conectado a um eNodeB (estação base LTE) e realiza medições periódicas do sinal das células vizinhas. Essas medições são usadas para determinar se há outras células disponíveis com um sinal mais forte ou de melhor qualidade do que a célula atual. Se uma célula vizinha for identificada como mais adequada, o dispositivo móvel solicita um *handover* à estação base atual. Segundo Gódor *et al.* (2015), esse procedimento envolve três etapas: detecção e medição; decisão; e execução.

- Detecção e medição: o dispositivo móvel realiza medições da qualidade do sinal das células vizinhas enquanto está conectado à célula atual. Se a qualidade do sinal de uma célula vizinha for considerada melhor do que a atual, o dispositivo móvel envia uma solicitação de *handover* para a estação base.
- Decisão: a estação base avalia a solicitação de *handover* do dispositivo móvel, toma uma decisão sobre sua necessidade e, em caso afirmativo, qual a célula de destino adequada. A estação base leva em consideração fatores como a qualidade do sinal, a capacidade da célula de destino e a carga de tráfego.
- Execução: a estação base envia uma mensagem de *handover* para o dispositivo móvel instruindo-o a mudar para a célula de destino. O dispositivo interrompe a conexão com a célula atual e se conecta à célula de destino, mantendo assim a continuidade da comunicação.

2.1.3 RSSI (Received Signal Strength Indicator)

O RSSI é um parâmetro crucial para avaliar a qualidade do sinal em redes de comunicação sem fio, incluindo a rede LTE. Ao analisar o RSSI, os engenheiros de rede podem identificar problemas de sinal, otimizar o posicionamento das células e melhorar a experiência geral do usuário (Polak *et al.* 2025; Afroz *et al.* 2015).

O RSSI é um indicador que mede a potência do sinal de radiofrequência recebido por um dispositivo móvel de uma estação base. A medida é expressa em decibéis em relação ao *miliwatt* (dBm), valores mais altos correspondem a sinais mais fortes. Um sinal mais forte geralmente se traduz em uma melhor qualidade de conexão e um desempenho mais alto da rede (Aprillia *et al.* 2020).

O RSSI desempenha um papel fundamental na avaliação da qualidade do sinal e na otimização das redes LTE, pois ajuda a identificar áreas com cobertura protegida. Os engenheiros de rede podem localizar áreas com baixa intensidade de sinal e implementar melhorias para garantir uma cobertura mais uniforme e confiável. Ele auxilia também na seleção de células, pois os dispositivos móveis usam o RSSI para determinar qual célula oferece o melhor sinal e se conectar a ela. Isso é essencial para garantir a melhor qualidade de conexão

possível para os usuários e, por fim, facilita o *handover*. Quando um dispositivo móvel se desloca entre as células, o RSSI ajuda a determinar qual célula oferece o sinal mais forte e a realizar uma transição suave entre as células (Krasniqi *et al.* 2018). Conforme discutido por Ferreira *et al.* (2021), a Tabela 2.1 apresenta os valores de RSSI em dBm correspondentes a diferentes níveis de sinal.

Tabela 2.1: tabela de valores de RSSI em redes LTE

RSSI (dBm)	Nível de sinal
> - 60	Muito Forte
-60 a -80	Forte
-80 a -95	Médio
-95 a -110	Fraco
< -110	Muito fraco

Fonte: Ferreira *et al.* (2021).

2.1.4 RSRP (Reference Signal Received Power)

RSRP é o nível médio de potência do sinal de referência recebido pela antena do dispositivo móvel de uma determinada célula LTE. É usado como um indicador da qualidade do sinal, que afeta a velocidade de dados e a qualidade de voz. Valores mais altos de RSRP indicam que o dispositivo está mais próximo da estação base, portanto, pode ter uma melhor qualidade de sinal e taxa de *throughput* de dados. Vale ressaltar que o RSRP é medido em dBm (Polak *et al.* 2025).

Os valores típicos de RSRP em redes LTE podem variar de -140 dBm (fraco) a -44 dBm (forte), conforme apresentado na Tabela 2.2. É importante lembrar que os valores de RSRP podem ser afetados por fatores como distância da estação base, interferência de sinal, obstruções no caminho do sinal e condições climáticas, e podem variar conforme o local e horário do dia (Putra *et al.* 2021).

O cálculo do RSRP em redes LTE é baseado na medição da potência do sinal de referência recebido pela antena do dispositivo móvel em relação a uma determinada célula de rede. Esse sinal de referência é transmitido pela estação base em um intervalo de tempo e

Tabela 2.2: tabela de valores de RSRP em redes LTE

RSRP (dBm)	Nível de sinal
-80 a -44	Excelente
-90 a -80	Bom
-100 a -90	Regular
-110 a -100	Ruim
-140 -110	Muito Ruim

Fonte: adaptado de [Putra et al. \(2021\)](#).

frequência conhecidos para permitir que o dispositivo móvel o detecte e o compare com outros sinais recebidos e importantes para o *handover* ([Hadian et al. 2024](#); [Putra et al. 2021](#)).

A [Equação 2.1](#) descreve o cálculo do RSRP conforme [Putra et al. \(2021\)](#) e [Aprillia et al. \(2020\)](#):

$$\text{RSRP}(\text{dBm}) = \text{RSSI}(\text{dBm}) - 10 \times \log(12 \times N) \quad (2.1)$$

Onde, N representa o número de *Resources Bloks* disponíveis no sistema em função da largura de banda utilizada. O *Resource Block* (RB)¹ é uma pequena unidade de recurso da estrutura de quadros definida para o padrão LTE. Este recurso é utilizado para trafegar informações de dados e controle em redes.

O parâmetro RSRP trabalha em conjunto com o parâmetro RSSI, medido em dBm, sendo este uma métrica mais tradicional, utilizada para exibir a força do sinal no GSM, pois integra toda a energia de rádio frequência dentro da banda passante do canal. Em outras palavras, no LTE, para a medição do RSSI na largura de banda, são consideradas todas as rádios de frequência ativas ([Rolim 2014](#)).

¹*Resource Block* (RB) refere-se à unidade mínima de recursos de tempo e frequência alocada para a transmissão de dados em redes móveis, como LTE e 5G. Cada RB é composto por um conjunto de subportadoras em uma determinada faixa de frequência e uma duração específica no domínio do tempo. Sua função é garantir a eficiência e a flexibilidade na utilização do espectro, permitindo que o sistema ajuste dinamicamente a alocação de recursos conforme as necessidades de cada usuário e as condições da rede.

2.1.5 RSRQ (Reference Signal Received Quality)

O RSRQ é outro parâmetro importante para medir a qualidade do sinal em redes LTE. Mede a qualidade do sinal recebido em relação ao ruído, sendo calculado como a razão entre a potência do sinal de referência recebido e a densidade espectral de ruído na célula. Também fornece informações sobre a qualidade do sinal de referência e a interferência na célula. É, ainda, um indicador da capacidade da rede de transmitir dados (Hadian *et al.* 2024; Ferreira *et al.* 2021).

A Tabela 2.3 apresenta dados do RSRQ sendo relatado em dB e os valores típicos podem variar de -20 dB -3 dB. Valores mais altos de RSRQ indicam uma melhor qualidade de sinal, pois isso significa que a potência do sinal de referência recebido é mais forte em relação ao ruído na célula. No entanto, valores muito baixos podem indicar uma interferência excessiva na célula, o que pode levar a uma degradação da qualidade do sinal (Putra *et al.* 2021).

Tabela 2.3: tabela de valores de RSRQ em redes LTE

RSRQ (dB)	Qualidade do sinal
-10 a -3	Excelente
-12 a -10	Bom
-14 a -12	Regular
-17 a -14	Ruim
-20 a -17	Muito Ruim

Fonte: adaptado de Putra *et al.* (2021).

RSRQ é usado para determinar a capacidade da célula de transmitir dados e para identificar áreas com baixa qualidade de sinal e interferência. E os seus valores são usados em conjunto com outros parâmetros de qualidade de sinal, como o RSRP, para determinar a melhor célula de destino durante o *handover*. Isso ajuda a garantir que os dispositivos móveis estejam conectados à célula mais apropriada e proporcionem uma melhor experiência ao usuário (Rolim 2014).

A Equação 2.2 descreve o RSRQ em redes LTE, conforme mencionado por Putra *et al.* (2021) e Aprillia *et al.* (2020):

$$RSRQ = N \times RSRP/RSSI \quad (2.2)$$

Em que, N é o número de RB, blocos de recursos.

2.1.6 SNR (Signal to Noise)

O SNR é um parâmetro importante que mede a relação entre a potência do sinal recebido e a potência do ruído em uma rede de comunicações. Em outras palavras, é a relação entre o sinal que queremos receber e o ruído de fundo na rede. O SNR é um indicador crítico da qualidade do sinal em redes de comunicação de redes móveis. A [Equação 2.3](#) para o cálculo do SNR é descrita por [Putra et al. \(2021\)](#):

$$SNR = S/(I + N) \quad (2.3)$$

Na [Equação 2.3](#), S é a potência do sinal recebido, I é a interferência causada por outros sinais transmitidos na mesma faixa de frequência e N é o ruído térmico aditivo.

O SNR é calculado como a relação entre a potência do sinal recebido e a potência do ruído de fundo, em dB. Quanto maior o valor do SNR, melhor será a qualidade do sinal, conforme apresentado na [Tabela 2.4](#). O SNR é importante porque determina a capacidade da rede de transmitir e receber dados com uma alta taxa de transferência, com uma menor taxa de erros e com uma melhor qualidade de voz e imagem ([Hadian et al. 2024](#)).

Tabela 2.4: tabela de valores de SNR em redes LTE

SNR (dB)	Qualidade do sinal
10 a 30	Excelente
3 a 10	Bom
-3 a 3	Regular
-20 a -3	Ruim

Fonte: adaptado de [Putra et al. \(2021\)](#).

Em redes, o SNR é particularmente importante para redes móveis, porque afeta a qualidade da chamada e a velocidade de transferência de dados. Um valor baixo de SNR

indica uma baixa qualidade de sinal, podendo causar quedas de chamadas, atrasos na transferência de dados e uma experiência de usuário ruim. Por outro lado, um alto valor de SNR pode melhorar a qualidade da chamada e a velocidade de download de dados, proporcionando uma experiência de usuário mais agradável.

As operadoras usam o SNR como um indicador de qualidade do sinal, juntamente com outros parâmetros como RSRP e RSRQ em redes LTE. O monitoramento contínuo do SNR e outros parâmetros de qualidade do sinal é essencial para garantir uma experiência do usuário de alta qualidade em redes móveis e para ajustar as redes conforme as necessidades dos usuários e das áreas de cobertura.

2.1.7 Uplink

Uplink é um termo usado no contexto das redes de comunicação para se referir à transmissão de dados do dispositivo do usuário, como um smartphone ou tablet, para a rede de comunicação, como uma estação base ou torre de celular. Essa conexão é crucial para uma comunicação eficaz e eficiente, pois permite o envio de informações como mensagens de texto, chamadas telefônicas, imagens e vídeos para outros dispositivos. O *uplink* é um componente essencial das redes de comunicação modernas, sendo complementar ao *downlink*, que é a transmissão de dados da rede para o dispositivo do usuário.

O *uplink* é importante porque permite que os dispositivos dos usuários se comuniquem com a rede e, em última instância, com outros dispositivos conectados à rede. Isso possibilita a troca de informações e a realização de tarefas em tempo real, como fazer chamadas, enviar e-mails e navegar na web. A qualidade do *uplink* tem um impacto direto na velocidade e eficiência das comunicações, bem como na satisfação do usuário com a experiência do serviço.

Na arquitetura LTE, o *uplink* funciona por meio de um processo conhecido como *Single Carrier Frequency Division Multiple Access* (SC-FDMA). Essa técnica de modulação é usada para alocar recursos de frequência e tempo para os dispositivos dos usuários, garantindo uma distribuição eficiente da largura de banda disponível. O SC-FDMA minimiza a interferência entre os dispositivos e melhora a qualidade geral do *uplink*, permitindo uma comunicação mais rápida e confiável (Oussakel *et al.* 2019b; Zhou e Guo 2013; Holma e Toskala

2011).

Conforme mencionado por [Dahlman et al. \(2011\)](#), a tecnologia SC-FDMA foi escolhida para o *uplink* no LTE devido às suas características. Esta tecnologia ajuda a reduzir o pico de potência transmitido pelos dispositivos móveis durante a transmissão de dados. Isso significa que a quantidade máxima de energia necessária para transmitir os dados é menor em comparação com outras técnicas de modulação. Isso torna a transmissão de dados mais eficiente em termos de energia, pois ajuda a prolongar a vida útil da bateria e reduzir o consumo total de energia ([Ben Hcine e Bouallegue 2015](#); [Sezgin et al. 2018](#)). Além disso, a redução do pico de potência durante a transmissão também ajuda a reduzir a interferência em outras comunicações próximas. Isso é importante porque, em uma rede de comunicação sem fio, várias comunicações podem estar ocorrendo simultaneamente, e a interferência pode causar problemas de qualidade de sinal ou até mesmo falhas na comunicação.

2.1.8 Downlink

O *downlink* se refere à transmissão de dados da rede de comunicação, como uma estação base ou torre de celular, para o dispositivo do usuário, como um smartphone ou tablet ([Dahlman et al. 2011](#)). Essa conexão desempenha um papel crucial nas comunicações modernas, permitindo que os usuários recebam informações de outros dispositivos e da rede. O *downlink* é um componente essencial das redes de comunicação atuais e opera em conjunto com o *uplink*, responsável pela transmissão de dados do dispositivo do usuário para a rede.

A importância do *downlink* reside na capacidade de fornecer informações e permitir comunicações entre dispositivos e a rede. O *downlink* possibilita o acesso a um conteúdo online, como sites, aplicativos e serviços de *streaming*, bem como a comunicação com outras pessoas por meio de chamadas, mensagens e videochamadas ([Dahlman et al. 2011](#)). A qualidade do *downlink* afeta diretamente a velocidade e eficiência das comunicações e a satisfação do usuário com a experiência do serviço.

Na arquitetura LTE, o *downlink* é gerenciado por meio da tecnologia OFDMA ([Oussakel et al. 2019b](#); [Dahlman et al. 2011](#)). Essa técnica de modulação permite a alocação eficiente de recursos de frequência e tempo aos dispositivos dos usuários, minimizando a interferência entre os dispositivos e melhorando a qualidade geral do *downlink*. O OFDMA

oferece alta eficiência espectral e suporta diferentes esquemas de modulação como QPSK, 16QAM e 64QAM, que podem ser adaptados dinamicamente às condições do canal e aos requisitos de taxa de dados (Holma e Toskala 2011; Sezgin *et al.* 2018).

2.2 Aprendizado de Máquina

O aprendizado de máquina, ou *machine learning* em inglês, é uma subárea da Inteligência Artificial (IA) que visa desenvolver algoritmos e sistemas capazes de aprender a partir de dados e melhorar seu desempenho. Este ramo da IA tem ganhado destaque nas últimas décadas devido ao aumento exponencial do volume de dados disponíveis e ao avanço das capacidades computacionais (Goodfellow *et al.* 2016).

A importância do aprendizado de máquina reside em sua capacidade de lidar com tarefas complexas e volumosas de dados que seriam inviáveis ou ineficientes para serem realizadas por seres humanos. O aprendizado de máquina tem sido aplicado com sucesso em diversas áreas, como medicina, finanças, marketing, transporte, energia e pesquisa científica (Jordan e Mitchell 2015).

Na medicina, por exemplo, algoritmos de aprendizado de máquina têm sido utilizados para diagnóstico de doenças, análise de imagens médicas e desenvolvimento de tratamentos personalizados (Topol 2019; Stam *et al.* 2023; Mamalakis *et al.* 2023). No setor financeiro, têm sido empregados em detecção de fraudes, otimização de carteiras de investimento e análise de risco de crédito (Vats e Samdani 2019; Chen e Guestrin 2016). Já no campo do marketing, é posto em prática para segmentação de clientes, análise de sentimentos e otimização de campanhas publicitárias (Sharma *et al.* 2022; Bhaskaraputra *et al.* 2022).

Além disso, o aprendizado de máquina tem um papel crucial no desenvolvimento de sistemas de IA mais avançados e autônomos como assistentes virtuais, carros autônomos e robôs colaborativos (Perla *et al.* 2022). Por isso, seu domínio é fundamental para impulsionar a inovação e melhorar a qualidade de vida das pessoas em diversas áreas (Mnih *et al.* 2015).

O aprendizado de máquina é um conceito bem definido na comunidade científica e tem aplicações em diversas áreas do conhecimento. Muitos dados são capturados diariamente ao redor do mundo. Com uma quantidade massiva de dados é importante para os pesquisado-

res e a indústria descobrir como extrair informações. Com uma quantidade pequena de dados o usuário consegue manipular e extrair informação, porém quando a quantidade de dados aumenta a análise torna-se cada vez mais difícil para os humanos. Diante disso, utilizar uma máquina ou computador para processar e extrair informações desses dados é mais adequado pela velocidade dos computadores em realizar essa tarefa (Sutton e Barto 2018).

Esse processo pode ser entendido como a capacidade de fazer o computador modificar ou adaptar suas ações de maneira a torná-las mais precisas. Para ser possível o aprendizado faz-se necessário que seja dado como entrada um conjunto de dados ou *dataset* para que o computador possa aprender.

Bhaskaraputra *et al.* (2022) definem aprendizado de máquina como computadores programados para aprender com a experiência. Para fazer isso, eles empregam um princípio chamado indução, no qual conclusões gerais são tiradas de um conjunto específico de exemplos. Assim, um algoritmo de aprendizado de máquina aprende a generalizar a partir de dados que representam instâncias de um problema, uma função ou hipótese que soluciona a demanda.

2.2.1 Tipos de Métodos de Aprendizagem

Uma abordagem de aprendizado de máquina é observar um modelo computacional com um conjunto de padrões (instâncias) e tentar definir uma solução geral para um problema. Embora seja uma ferramenta com uma definição geral curta, existem vários tipos de métodos. Conforme mostrado na Figura 2.3, o aprendizado é dividido em supervisionado, não supervisionado, semi-supervisionado e aprendizado por reforço (Geron 2019).

O aprendizado de máquina supervisionado é uma das principais abordagens, na qual os algoritmos são treinados com um conjunto de dados rotulados, ou seja, cada exemplo no conjunto de dados é associado a um rótulo ou resposta correta (Sudarsan *et al.* 2025). O objetivo principal do aprendizado supervisionado é criar um modelo capaz de generalizar a partir dos exemplos rotulados e fazer previsões precisas para novos dados não rotulados.

Essa abordagem é amplamente utilizada em tarefas de classificação e regressão. Na classificação, o objetivo é prever uma classe ou categoria para uma dada entrada, enquanto na regressão o objetivo é prever um valor contínuo (Bishop 2011). Um algoritmo de aprendi-

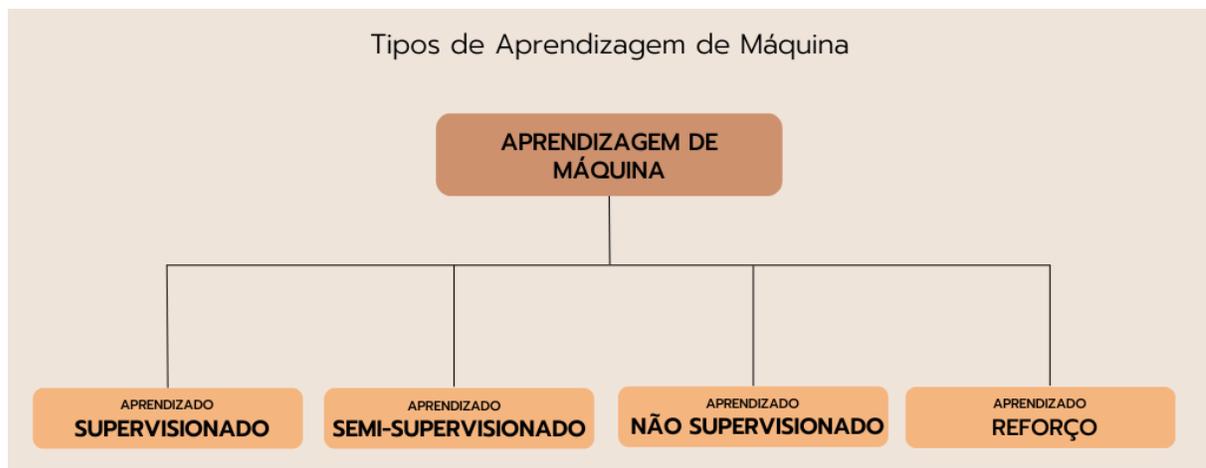


Figura 2.3: tipos de aprendizagem de máquina
Fonte: elaboração própria (2025).

zado supervisionado pode ser utilizado para diagnosticar doenças a partir de exames médicos. Nesse caso, o conjunto de treinamento consiste em exemplos de exames previamente rotulados, indicando se o paciente possui ou não a doença. O algoritmo utiliza esses dados para aprender um modelo capaz de classificar novos exames de pacientes ainda não rotulados (Topol 2019).

No aprendizado supervisionado, para cada exemplo apresentado ao algoritmo de aprendizado, a resposta necessária precisa ser fornecida (ou seja, no caso de um problema de classificação de imagens, um rótulo que informa a qual classe o exemplo pertence, por exemplo, como distinguir imagens de gatos e cães). Cada exemplo é descrito por um conjunto de valores (atributos) e rótulos de classe associados. O objetivo do algoritmo é construir um classificador que possa determinar corretamente a classe de novos exemplos não rotulados. Este problema é conhecido como classificação de rótulos de classes discretas e regressão de valores contínuos. Este método de aprendizagem é o mais utilizado (Prasanth *et al.* 2025; Vats e Samdani 2019).

Os não supervisionados são divididos em: agrupamento, resumo e associação. No agrupamento, os dados são agrupados conforme a similaridade. Em resumo, o objetivo é encontrar uma descrição simples e compacta para os dados. Ao associar, um esquema deve ser encontrado para a associação entre os atributos de dados (Vats e Samdani 2019).

No aprendizado não supervisionado, são fornecidos exemplos de algoritmos sem ró-

tulos. Existem clusterização, detecção de anomalias e de novidades, visualização e redução da dimensionalidade e aprendizado de regras por associação. Em relação a clusterização, o algoritmo agrupa exemplos com base na similaridade de atributos. O algoritmo analisa os exemplos dados e tenta determinar se alguns deles podem ser agrupados de alguma forma, compondo um *cluster*. Uma vez que os agrupamentos tenham sido identificados, muitas vezes é necessária uma análise para determinar o que cada agrupamento significa no contexto do problema que está sendo analisado (Kelleher *et al.* 2015)

Por exemplo, uma pessoa tem uma grande quantidade de informações sobre os visitantes do seu blog e executa um algoritmo de agrupamento com o objetivo de identificar grupos de visitantes semelhantes aos da Figura 2.4. Em nenhum momento são fornecidas ao algoritmo informações sobre a qual grupo o visitante pertence, o algoritmo encontrará essas relações sem a intervenção do usuário. Por exemplo, ele pode descobrir que 40% dos visitantes são homens que gostam de quadrinhos e geralmente leem seu blog à noite, enquanto 20% são jovens fãs de ficção científica que visitam seu blog nos fins de semana. Se você utilizar um algoritmo de agrupamento hierárquico, também será possível subdividir cada grupo em subgrupos menores. Isso pode ajudá-lo a adaptar suas postagens para atender melhor a cada um desses grupos (Geron 2019).

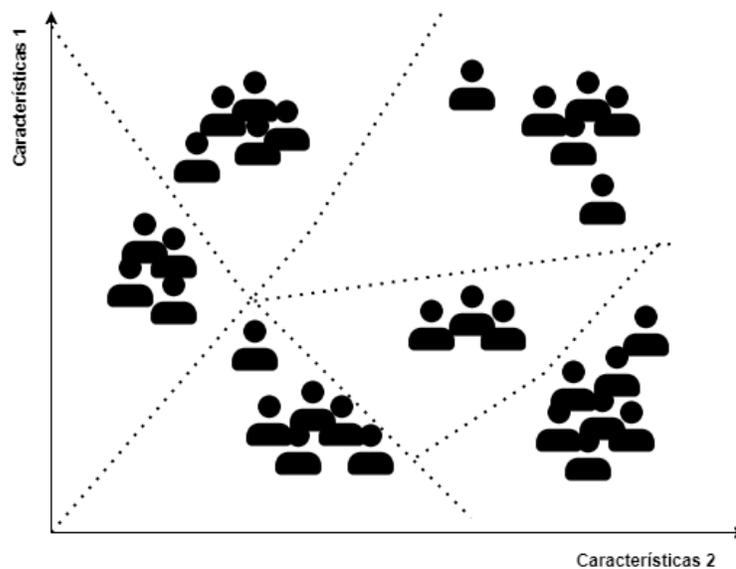


Figura 2.4: exemplo de aprendizado não supervisionado, clusterização
Fonte: adaptado de Geron (2019).

O aprendizado semi-supervisionado é uma abordagem de aprendizado de máquina que combina características dos métodos supervisionados e não supervisionados, utilizando conjuntos de dados parcialmente rotulados para treinamento. Nessa abordagem, os algoritmos aproveitam as informações fornecidas pelos exemplos rotulados e pelos exemplos não rotulados, permitindo a criação de modelos mais precisos e generalizáveis, especialmente quando a rotulação completa dos dados é cara ou difícil de obter (Sutton e Barto 2018).

O aprendizado por reforço é uma abordagem em que um agente aprende a tomar decisões sequenciais interagindo com um ambiente e buscando maximizar uma recompensa acumulada ao longo do tempo (Sutton e Barto 2018). Nesse paradigma, o agente não recebe rótulos ou exemplos de soluções corretas, mas sim recompensas ou punições com base em suas ações. O objetivo é aprender uma política ótima, que é uma função que mapeia estados do ambiente às ações que maximizam a recompensa esperada a longo prazo (Geron 2019).

O aprendizado por reforço pode ser ilustrado com um exemplo simples: suponha que queremos ensinar um robô a caminhar em uma determinada direção. Inicialmente, o robô não sabe como se movimentar corretamente. O aprendizado por reforço consiste em dar ao robô *feedback* positivo ou negativo, dependendo da sua ação. Se o robô der um passo correto, ele recebe uma recompensa (*feedback* positivo); se ele der um passo errado, ele recebe uma punição (*feedback* negativo).

Com o tempo, o robô aprende a tomar as ações corretas que levam a uma recompensa mais alta, que, neste caso, é caminhar na direção correta. O objetivo é aprender uma política ótima, que é uma função que mapeia estados do ambiente às ações que maximizam a recompensa esperada a longo prazo.

2.2.2 Métodos Preditivos

Em tarefas de predição, o objetivo é encontrar uma função de hipótese a partir dos dados de treinamento. A finalidade desse método é marcar ou atribuir uma "classe" a qualquer instância de esquema. Existem muitos algoritmos de predição ou classificação propostos na literatura, alguns deles são:

Floresta Aleatória (RF)

O algoritmo floresta aleatória, do inglês *random forest* (RF) tem sido amplamente utilizado em problemas de regressão devido à sua alta precisão e capacidade de lidar com dados de alta dimensionalidade. Como afirmado por [Sudarsan et al. \(2025\)](#), a floresta aleatória é uma técnica de aprendizado de máquina popular para regressão e classificação, devido à sua precisão e facilidade de uso.

O algoritmo RF é baseado em árvores de decisão que opera criando inúmeras árvores de decisão aleatórias, cada uma treinada em um subconjunto aleatório dos dados de treinamento. Cada árvore produz uma predição e a saída final é a média dessas predições. A [Equação 2.4](#) do algoritmo para regressão é dada por:

$$y = \frac{1}{B} \sum_{i=1}^B f_i(x)$$

onde y é a saída prevista,
 f_i é a previsão produzida pela i -ésima árvore,
 B é o número total de árvores no modelo e
 x é o conjunto de características.

(2.4)

Uma das vantagens do algoritmo é que ele é capaz de lidar com dados desbalanceados e correlacionados. Além disso, ele pode identificar as características mais importantes para a predição, o que pode ser útil para entender o problema em questão e para selecionar as melhores características para o modelo ([Jordan e Mitchell 2015](#)).

K Vizinhos Mais Próximos (KNN)

O algoritmo de k vizinhos mais próximos, do inglês *k-nearest neighbors* (KNN) é um método de aprendizado de máquina usado para problemas de classificação e regressão ([Altman 1992](#)). É um algoritmo não paramétrico que não faz suposições sobre a distribuição dos dados e, em vez disso, depende da proximidade entre os pontos de dados.

No caso de problemas de regressão, o objetivo do KNN é prever um valor numérico para uma variável de destino com base nas características das variáveis independentes. O algoritmo calcula a distância entre o novo ponto de dados e todos os pontos de dados no

conjunto de treinamento ([Mohri et al. 2018](#)). Ele seleciona os k pontos mais próximos, com base na distância euclidiana, onde k é um número inteiro positivo pré-definido pelo usuário.

Em seguida, o KNN usa esses k pontos mais próximos para prever o valor da variável de destino para o novo ponto de dados. Isso é feito calculando a média dos valores da variável de destino dos k vizinhos mais próximos ([Altman 1992](#)). A [Equação 2.5](#) do KNN para regressão:

$$y = \frac{1}{k} \sum_{i=1}^k y_i$$

onde y é a variável de destino para o novo ponto de dados,
 k é o número de vizinhos mais próximos selecionados e
 y_i é o valor da variável de destino para o i -ésimo vizinho mais próximo.

(2.5)

Para entender melhor como funciona o KNN, vamos considerar um exemplo. Digamos que você esteja tentando prever o preço de uma casa com base em sua área e número de quartos. Você tem um conjunto de treinamento com dados de preços de casas para outras áreas e números de quartos. Quando um novo ponto de dados é apresentado, o KNN calcula a distância entre os valores de área e número de quartos do novo ponto de dados e de todos os pontos de dados no conjunto de treinamento. Em seguida, ele seleciona os k vizinhos mais próximos com base na distância euclidiana e calcula a média de seus preços. Este valor médio é a predição do KNN para o preço da casa para o novo ponto de dados ([Jordan e Mitchell 2015](#)).

Máquina de Vetores de Suporte (SVM)

A máquina de vetor de suporte, do inglês *support vector machine* (SVM) é uma técnica baseada em métodos estatísticos. O SVM é uma opção atraente por sua capacidade de lidar com grandes volumes de dados, e por ser uma técnica bastante robusta e flexível ([Bayram et al. 2013](#)).

A formulação para a regressão com SVM envolve a minimização da soma do erro de treinamento e da complexidade do modelo. A [Equação 2.6](#) de regressão para SVM é dada por:

$$f(x) = w \cdot x + b$$

onde x é o vetor de entrada,
 w é o vetor de peso e
 b é o termo de polarização (bias).

(2.6)

A escolha da função de kernel é importante para a regressão com SVM, assim como na classificação. Algumas funções de kernel comuns são o linear, o polinomial e o *Radial Basis Function*.

O algoritmo de otimização de SVM encontra os valores ótimos de w e b minimizando uma função de custo que é a soma do erro de treinamento e uma penalidade para limitar a complexidade do modelo e evitar o *overfitting*². A penalidade é controlada pelo parâmetro C , ajustado por validação cruzada (Bayram *et al.* 2013).

SVM tem sido amplamente utilizado em diversas áreas de aplicação, como, por exemplo, previsão de preços de ações, previsão de tempo, previsão de demanda, etc. Essa técnica demonstra um excelente desempenho em tarefas de regressão, sendo capaz de lidar com grandes volumes de dados e de generalizar bem para novos exemplos (Jordan e Mitchell 2015).

Regressão Linear (LR)

Regressão linear, do inglês *linear regression* (LR) é uma técnica estatística amplamente empregada na análise de dados, com o objetivo de modelar a relação entre uma variável dependente e uma ou mais variáveis independentes. Quando há apenas uma variável preditora, o modelo é denominado regressão linear simples, e quando há múltiplas variáveis preditoras, utiliza-se o modelo de regressão linear múltipla (James *et al.* 2013). A Equação 2.7 de regressão linear simples pode ser escrita como:

² *Overfitting* ocorre quando um modelo de aprendizado de máquina se ajusta excessivamente aos dados de treinamento, levando a um desempenho ruim ao lidar com novos dados.

$$y = \beta_0 + \beta_1 x + \epsilon \quad (2.7)$$

onde y é a variável dependente,
 x é a variável independente,
 β_0 é o intercepto,
 β_1 é o coeficiente de regressão e
 ϵ é o erro residual.

O intercepto representa o valor esperado da variável dependente quando a variável independente é zero, enquanto o coeficiente de regressão mede a mudança esperada na variável dependente quando a variável independente aumenta em uma unidade.

A equação de regressão linear simples serve para prever valores de uma variável dependente com base em valores conhecidos de uma variável independente. Por exemplo, se estivermos estudando o efeito da quantidade de tempo que os alunos estudam em suas notas de teste, poderíamos usar a equação de regressão linear simples para prever a nota de um aluno com base no tempo que ele estudou ([James et al. 2013](#)).

A regressão linear também pode ser utilizada para análise de dados de múltiplas variáveis, ou seja, a regressão linear múltipla. Neste caso, a [Equação 2.8](#) de regressão é escrita como:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \epsilon \quad (2.8)$$

Onde x_1, x_2, \dots, x_n são as variáveis independentes e $\beta_1, \beta_2, \dots, \beta_n$ são os coeficientes de regressão correspondentes. A regressão linear múltipla permite analisar a relação entre várias variáveis independentes e uma variável dependente ([Geron 2019](#)).

Rede Neural (NN)

As redes neurais, do inglês *neural network* (NN), são modelos computacionais inspirados na estrutura e no funcionamento do cérebro humano. São compostas por camadas de neurônios interconectados que processam informações em paralelo. A capacidade de aprender padrões complexos a partir dos dados torna as NNs especialmente eficazes em tarefas de regressão e classificação ([Sudarsan et al. 2025](#)).

Uma rede neural típica para regressão é composta por uma camada de entrada, uma ou mais camadas ocultas e uma camada de saída. Cada neurônio aplica uma função de ativação não linear sobre uma combinação linear dos valores de entrada. A [Equação 2.9](#) representa o funcionamento de um neurônio em uma camada é dada por:

$$y = f \left(\sum_{i=1}^n w_i x_i + b \right)$$

onde y é a saída do neurônio,
 x_i são as entradas,
 w_i são os pesos das conexões,
 b é o termo de polarização (bias) e
 f é a função de ativação.

(2.9)

Durante o treinamento, os pesos e os bias são ajustados por meio do algoritmo de retropropagação, em conjunto com otimizadores como o *Stochastic Gradient Descent* (SGD) ou o *Adam*, com o objetivo de minimizar uma função de custo, geralmente o erro quadrático médio (MSE) em tarefas de regressão.

NN são particularmente úteis em contextos com grande volume de dados e relações não lineares entre as variáveis, sendo aplicadas com sucesso em diversas áreas, como reconhecimento de padrões, séries temporais, previsão de demanda e, neste trabalho, na previsão de vazão em redes móveis.

2.2.3 Métricas de Avaliação

A avaliação do desempenho preditivo de modelos é comumente realizada por meio de diversas métricas de erro que quantificam a discrepância entre os valores previstos e os observados. Entre as métricas mais utilizadas destacam-se: erro quadrático médio (MSE), erro absoluto médio (MAE), raiz do erro quadrático médio (RMSE) e coeficiente de determinação (R^2). Essas métricas permitem avaliar, sob diferentes perspectivas, a acurácia e a capacidade explicativa dos modelos. A seguir, apresentam-se suas definições formais e respectivas expressões matemáticas.

Erro Quadrático Médio (MSE): mede a média dos quadrados das diferenças entre os

valores observados (Y_i) e previstos (\hat{Y}_i), penalizando mais severamente grandes desvios, conforme definido na [Equação 2.10](#):

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (2.10)$$

Erro Absoluto Médio (MAE): calcula a média das diferenças absolutas entre os valores observados e previstos, proporcionando uma interpretação direta dos erros, conforme definido na [Equação 2.11](#):

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |Y_i - \hat{Y}_i| \quad (2.11)$$

Raiz do Erro Quadrático Médio (RMSE): representa a raiz quadrada do MSE, fornecendo a magnitude média dos erros na mesma unidade dos valores observados, conforme definido na [Equação 2.12](#):

$$\text{RMSE} = \sqrt{\text{MSE}} \quad (2.12)$$

Coefficiente de Determinação (R^2): quantifica a proporção da variabilidade dos dados explicada pelo modelo, sendo que valores próximos a 1 indicam melhor ajuste, conforme definido na [Equação 2.13](#):

$$R^2 = 1 - \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \quad (2.13)$$

2.3 Resumo

Este capítulo descreve os conceitos fundamentais necessários para a compreensão dos tópicos abordados neste trabalho. Desde sobre tecnologia LTE, sua arquitetura e sua operação em camadas, fornecendo uma explicação interna de cada camada. Além disso, alguns parâmetros de avaliação de desempenho da rede como RSSI, RSRP, RSRQ e SNR foram apresentados e, por fim, foram discutidos o *downlink* e o *uplink* e suas características. Para finalizar, a seção [2.2](#) apresentou sobre o aprendizado de máquina, sua definição e seus tipos de métodos, como supervisionado e não supervisionado, além de discutir alguns algoritmos preditivos.

Capítulo 3

Trabalhos Correlatos

Neste capítulo são apresentadas diversas publicações científicas que tratam do tema central deste trabalho, e estudos realizados por diferentes autores. Esses trabalhos foram obtidos por meio de uma revisão sistemática da literatura, que será descrita em detalhes. Também é fornecido um comparativo entre esses estudos e o trabalho proposto.

3.1 Revisão Sistemática da Literatura

A Revisão Sistemática da Literatura (RSL) é uma técnica de pesquisa que visa reunir, avaliar e sintetizar todas as evidências disponíveis sobre uma determinada questão de pesquisa. Seu objetivo é fornecer uma visão geral e crítica do conhecimento existente sobre um tema específico, identificando lacunas na literatura e áreas que precisam de mais investigação. É uma ferramenta essencial para a tomada de decisões baseadas em evidências e para o avanço do conhecimento em uma determinada área (Kitchenham e Brereton 2013).

3.1.1 Protocolo de Revisão Sistemática da Literatura

O protocolo de revisão sistemática elaborado neste estudo seguiu as diretrizes apresentadas por Mafra e Travassos (2006) e Kitchenham *et al.* (2009), que são referências importantes nesta área. Na etapa de planejamento, foram definidas a questão de pesquisa, os critérios de inclusão e exclusão, as fontes de pesquisa e as estratégias de busca. A busca e seleção de estudos foram conduzidas de forma sistemática, com a aplicação dos critérios de inclusão

e exclusão definidos previamente. A extração de dados foi realizada de forma padronizada, com a coleta de informações relevantes dos estudos selecionados. A síntese dos resultados foi realizada por meio de uma análise temática. O protocolo elaborado neste estudo permitiu a condução de uma revisão sistemática rigorosa e transparente, que contribuiu para a identificação e síntese das evidências disponíveis sobre o tema de interesse.

Questão de Pesquisa

Para garantir que a pesquisa atingisse o objetivo, foi necessário ter clareza sobre as questões que precisavam ser respondidas. O processo de definição das questões é um passo fundamental para garantir que a pesquisa esteja focada e direcionada.

No caso específico deste estudo, foram definidas duas Questões de Pesquisa (QP) visando fornecer uma estrutura para investigar a vazão em rede de telefonia móvel celular. Essas questões foram elaboradas com base nos objetivos específicos da pesquisa e buscam identificar informações relevantes para o estudo em questão.

- **QP1:** *"Quais fatores impactam o desempenho da rede de smartphone?"*;
- **QP2:** *"Quais técnicas e métodos são usados para fazer a predição da vazão?"*;

Base de Dados e Idioma

A escolha da base de dados Scopus para a realização desta revisão sistemática foi baseada em sua ampla cobertura de periódicos, eventos e fontes de informação, recursos avançados de busca e análise de citações, além da sua grande abrangência e diversidade de conteúdo científico em diversas áreas do conhecimento. Esses recursos ajudam a garantir a seleção de artigos relevantes e a avaliação da qualidade e validade da revisão sistemática, tornando-a mais confiável e rigorosa.

A língua escolhida para a realização desta revisão sistemática foi o inglês, considerado a língua franca da ciência e das publicações científicas. A escolha da língua inglesa se deve ao fato de que a maioria dos artigos científicos relevantes para esta pesquisa são publicados em inglês, assim como a maioria das bases de dados científicas. Além disso, a escolha do inglês permite a inclusão de uma ampla gama de artigos relevantes para a revisão sistemática,

umentando a abrangência da pesquisa e a identificação de soluções mais eficientes para a transmissão de dados em redes de telefonia móvel celular.

String de Busca

Na definição das strings de busca foi aplicada uma metodologia que consiste em usar uma conjunção de disjunções de três blocos. Esses blocos contêm palavras e sinônimos associados, respectivamente, à população, à intervenção e aos resultados desejados da revisão sistemática. Os itens para cada categoria da metodologia PIO (População, Intervenção e Resultado) são especificados na [Tabela 3.1](#).

Tabela 3.1: metodologia PIO aplicado no contexto da RSL

PIO	Descrição
População	Android Smartphones.
Intervenção	A partir da identificação da vazão
Resultado	Técnicas, métricas e análises relacionado a vazão em smartphone Android

Fonte: elaboração própria (2025).

O PIO utilizado nesta revisão sistemática foi desenvolvido para investigar a vazão em smartphones Android. A população selecionada foram os smartphones com sistema operacional Android, que são amplamente utilizados em todo o mundo. A intervenção considerada foi a identificação de vazão em redes de telefonia móvel celular para avaliar a eficiência da transmissão de dados nesses dispositivos.

O resultado esperado desta revisão sistemática é a identificação de técnicas, métricas e análises relacionadas à vazão em smartphones Android. Essas informações são importantes para entender as possíveis limitações e oportunidades de melhoria na transmissão de dados em redes de telefonia móvel celular. Com a aplicação do PIO, é possível identificar estudos relevantes e coletar informações que contribuem para o desenvolvimento de soluções mais eficientes para a transmissão de dados. As palavras-chave utilizadas para indexação e busca dos artigos foram selecionadas com base nos objetivos da revisão sistemática e são apresentadas na [Tabela 3.2](#). Foram utilizadas palavras relacionadas a dispositivos móveis, vazão de

dados, coleta de dados e técnicas de predição, além do aprendizado de máquina e modelos preditivos.

Tabela 3.2: expressão final da busca aplicada na base Scopus em Setembro de 2024

Indexing Base	String
Scopus	(“device” OR “smartphone” OR “mobile” OR “cellular”) AND (“uplink” OR “throughput prediction”) AND (“metric” OR “model” OR “technique” OR “Data Collection”) AND (“machine learning” OR “predictive models” OR “prediction”)

Fonte: elaboração própria (2025).

As palavras-chave incluíram "*device*", "*smartphone*", "*mobile*" e "*cellular*" para identificar artigos relacionados a dispositivos móveis. Além disso, foram incluídas palavras-chave relacionadas à vazão, como "*uplink*" e "*throughput prediction*" para identificar estudos que abordam a predição de vazão de dados em redes.

Para identificar estudos que abordam técnicas de coleta de dados e análise foram incluídas palavras-chave como "*metric*", "*model*", "*technique*" e "*data collection*". Essas palavras-chave permitem identificar estudos que desenvolvem técnicas de coleta e análise de dados relacionados à vazão.

Por fim, foram incluídas palavras-chave relacionadas a *machine learning* e modelos preditivos, como "*machine learning*", "*predictive models*" e "*prediction*". Essas palavras-chave permitem identificar estudos que utilizam técnicas de aprendizado de máquina e modelos preditivos para prever a vazão em redes.

A combinação dessas palavras-chave, que incluem palavras relacionadas a dispositivos móveis, vazão, coleta de dados, técnicas de predição e *machine learning*, permite identificar estudos relevantes para a revisão sistemática e contribuir para o desenvolvimento de soluções mais eficientes para a transmissão de dados.

Critério de Exclusão e Inclusão

O critério de exclusão utilizado nesta RSL teve como objetivo desconsiderar publicações que não eram pertinentes ao objetivo da pesquisa, que não puderam ser acessadas ou que não

abordavam a vazão em redes de telefonia móvel celular. Foram desconsideradas ainda publicações que nitidamente tratavam de outros assuntos não pertinentes à revisão sistemática.

Para o Critério de Exclusão foram desconsideradas as publicações que:

- (CE1) Nitidamente tratassem de outros assuntos não pertinentes ao objetivo desta revisão sistemática;
- (CE2) Que não pudessem ser baixados ou acessados;
- (CE3) Não abordassem vazão de dados.

Por outro lado, os critérios de inclusão foram definidos para considerar publicações que abordam métricas usadas para medir a vazão, citam técnicas ou métodos para fazer a predição de vazão, e descrevem os fatores que impactam o desempenho da rede do smartphone. Esses critérios permitiram a seleção de publicações relevantes para a análise e discussão na revisão sistemática. Para o Critério de Inclusão (CI) foram consideradas ainda as publicações que:

- (CI1) Abordassem sobre as métricas usadas para medir a vazão.
- (CI2) Citassem as técnicas ou métodos usados para fazer a predição de vazão;
- (CI3) Descrevessem os fatores que impactam o desempenho da rede do smartphone.

Critério de Seleção

O procedimento de seleção e critério adotado nesta revisão sistemática foi dividido em duas etapas:

1^a Etapa: foram lidos os resumos e palavras-chave das publicações selecionadas conforme a busca nas fontes. Se a publicação atendesse pelo menos a um dos critérios de inclusão, ela seria selecionada para a segunda etapa. E caso atendesse a algum critério de exclusão, seria eliminada.

2^a Etapa: as publicações selecionadas na primeira etapa foram lidas na íntegra e somente aquelas que mantiveram os critérios de inclusão foram selecionadas para análise e discussão. Esse procedimento de seleção permitiu a identificação de artigos relevantes para a revisão sistemática e aumentou a confiabilidade e rigor da seleção dos artigos.

Extração dos Dados

Para a realização da revisão sistemática, o procedimento de extração de dados foi um passo essencial para responder às questões de pesquisa. Nesse processo, foram extraídas informações relevantes das publicações selecionadas e registradas em tabelas para posterior análise e discussão. As informações extraídas das publicações incluem dados para identificação do artigo, tais como autor(es), título, base, ano de publicação, e informações para responder às questões de pesquisa, as técnicas e métricas de medição de vazão em redes de telefonia móvel celular, e os fatores que impactam o desempenho da rede do smartphone. A extração desses dados permitiu uma análise mais aprofundada dos resultados e contribuiu para a identificação de padrões e tendências relevantes na área de pesquisa.

3.1.2 Condução da Revisão Sistemática da Literatura

Para realizar esta RSL, foram considerados artigos publicados entre os anos de 2018 a 2025 e, além disso, apenas artigos em língua inglesa foram selecionados. Essa seleção temporal e linguística permitiu que a revisão se baseasse em estudos recentes e atualizados na área de pesquisa e incluísse uma ampla gama de publicações relevantes. Na etapa da condução, foi aplicado o protocolo estabelecido previamente e realizada a busca na base de dados. Inicialmente, foram identificadas 168 publicações relevantes para a pesquisa, conforme demonstrado na [Figura 3.1](#). Em seguida, as publicações foram submetidas ao primeiro filtro, no qual foram lidos o resumo e as palavras-chave. Com base nesse critério, apenas 41 publicações foram selecionadas para uma análise mais aprofundada.

Das 41 publicações selecionadas, todas foram lidas na íntegra e submetidas ao segundo filtro definido no protocolo. Esse segundo filtro teve como objetivo validar a relevância e a pertinência das publicações. Ao final, apenas 23 publicações foram selecionadas para a análise e discussão conforme os critérios estabelecidos. A aplicação rigorosa dos critérios de inclusão e exclusão permitiu a seleção de publicações relevantes para a pesquisa, aumentando a confiabilidade e validade da revisão sistemática. A [Tabela 3.3](#) apresenta os 23 artigos selecionados após o segundo processo de filtragem e é composta por colunas que destacam informações importantes, como o ID de cada artigo, título, autores e ano de publicação,

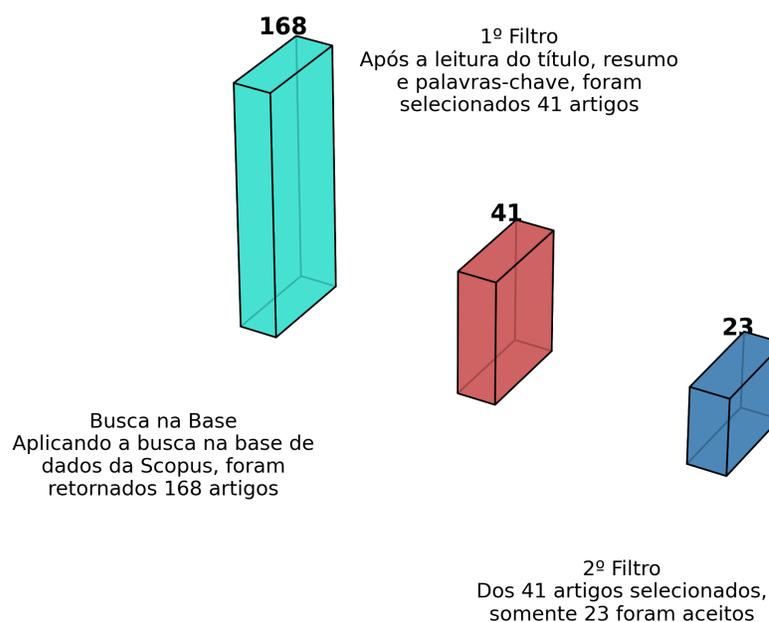


Figura 3.1: distribuição do número de artigos aceitos ao longo das etapas da revisão sistemática da literatura

Fonte: elaboração própria (2025).

fornecendo uma visão geral rápida dos artigos selecionados.

Com base na [Tabela 3.3](#), resultado da string de busca na base Scopus, observa-se, na análise temporal dos artigos selecionados na RSL, uma concentração expressiva de publicações em 2019 e 2020, com seis artigos em cada ano. Esse pico revela o crescente interesse da comunidade científica pelo tema da predição de vazão em redes móveis. Em 2024, verifica-se nova elevação no volume de publicações, totalizando cinco trabalhos, sugerindo uma retomada do interesse, possivelmente impulsionada pela evolução e consolidação das tecnologias de redes móveis. Por outro lado, os anos de 2018, 2021, 2022 e 2023 apresentaram frequência menor, com um a dois artigos por ano. Esses dados indicam que o tema permanece relevante e em constante desenvolvimento, acompanhando a evolução das tecnologias de comunicação móvel.

Ao final do processo de filtragem, foram selecionados 23 trabalhos que serviram como base para a pesquisa de trabalhos correlatos. A seleção desses artigos foi realizada de forma criteriosa, levando em consideração a relevância e a qualidade metodológica, de modo a assegurar uma fundamentação sólida para o desenvolvimento deste estudo.

Tabela 3.3: artigos selecionados com base nos resultados da busca, totalizando 23 artigos, com identificação, título, autores e ano de publicação

ID	Title	Authors	Year
A01	A Machine Learning Framework for Predicting Downlink Throughput in 4G-LTE/5G Cellular Networks	Al-Thaedan et al.	2024
A02	Deep Learning-Based Throughput Prediction in 5G Cellular Networks	Batool et al.	2024
A03	A Kalman Filter Based Low Complexity Throughput Prediction Algorithm for 5G Cellular Networks	Biswas et al.	2024
A04	Comparative Analysis of LSTM and GRU for Uplink Data Rate Prediction in 5G Networks	Markappa et al.	2024
A05	Predictive QoS for Cellular-Connected UAV Communications	Ann Varghese et al.	2024
A06	Downlink throughput prediction using machine learning models on 4G-LTE networks	Al-Thaedan et al.	2023
A07	Machine Learning for QoS Prediction in Vehicular Communication: Challenges and Solution Approaches	Alexandros Palaios et al.	2023
A08	Downlink Throughput Prediction in LTE Cellular Networks Using Time Series Forecasting	Ali Mostafa et al.	2022
A09	Machine-Learning-Based Uplink Throughput Prediction from Physical Layer Measurements	Eyceyurt et al.	2022
A10	Throughput Prediction Using Machine Learning in LTE and 5G Networks	Minovski et al.	2021
A11	Throughput Prediction using Machine Learning in LTE and 5G Networks	Elsherbiny et al.	2020
A12	Lumos5G: Mapping and Predicting Commercial mmWave 5G Throughput	Narayanan et al.	2020
A13	Mobile Internet Throughput Prediction using Machine Learning Techniques	Jha, H. and Vijayarajan, V.	2020
A14	4G Uplink Power Control Tuning Approach Using Unsupervised Machine Learning	Gaber et al.	2020
A15	On Leveraging Machine and Deep Learning for Throughput Prediction in Cellular Networks: Design, Performance, and Challenges	Raca et al.	2020
A16	GMDH modelling for mobile user throughput forecasting	Lawal, I.A.	2020
A17	Experimental Estimation of LTE-A Performance	Oussakel et al.	2019
A18	Cellular uplink bandwidth prediction based on radio measurements	Oussakel et al.	2019
A19	Analysis and estimation of video QoE in wireless cellular networks using machine learning	Minovski et al.	2019
A20	Fast and Efficient Cross Band Channel Prediction Using Machine Learning	Bakshi et al.	2019
A21	A deep learning approach for location independent throughput prediction	Schmid et al.	2019
A22	Estimating downlink throughput from end-user measurements in mobile broadband networks	Kousias et al.	2019
A23	Throughput Prediction Using Recurrent Neural Network Model	Wei et al.	2018

Fonte: elaboração própria (2025).

3.2 Trabalhos Correlatos como Resultado Revisão Sistemática da Literatura

Esta seção apresenta os trabalhos correlatos identificados a partir dos resultados da revisão sistemática da literatura. Os estudos foram classificados em duas categorias principais: aqueles que realizaram experimentos em ambientes controlados e os que conduziram suas investigações em ambientes reais.

3.2.1 Ambiente Controlado

Esta seção trata dos experimentos controlados e suas diferentes abordagens. Os experimentos em ambiente controlado são realizados em um ambiente cuidadosamente preparado para o experimento, permitindo maior controle sobre as variáveis de estudo. Embora essa abordagem seja útil em muitas situações, ela pode não refletir a realidade do uso cotidiano, o que pode levar a resultados que não correspondem às expectativas, porém são considerados essenciais para pesquisas preliminares.

3.2.2 Experimentos Realizados em Câmara Anecoica

Os experimentos realizados em laboratórios, em câmaras anecoicas, são conduzidos em ambientes completamente isolados de quaisquer outros fenômenos de rádio, o que garante que não haja degradação ou perda de sinal devido à falta de recursos de transmissão. Dessa forma, é possível assegurar que a capacidade máxima da rede seja preservada durante a realização dos experimentos (Oussakel *et al.* 2019b).

O propósito da pesquisa de Oussakel *et al.* (2019b) foi investigar o desempenho da tecnologia LTE-*Advanced* em um ambiente controlado e observar o impacto dos fenômenos de rádio na variação da largura de banda. Para isso, os autores implantaram um *testbed*¹ de LTE-*Advanced* em um laboratório com o intuito de coletar dados de desempenho em diferentes cenários. Os autores utilizaram um software de gerenciamento de tráfego para controlar a carga de tráfego gerada pelos usuários e para monitorar as métricas de desempenho.

¹ *Testbed* é um espaço controlado e geralmente simulado que permite aos pesquisadores, desenvolvedores e engenheiros testar e validar novas ideias, tecnologias ou conceitos.

O estudo utilizou um *testbed* 4G para analisar a predição de vazão em transmissões de *uplink*, considerando métricas das camadas inferiores. Os algoritmos de aprendizado de máquina supervisionado aplicados foram regressão linear e floresta aleatória (RF), permitindo a predição da largura de banda recebida no *uplink* em diferentes ambientes de fenômenos de rádio. Durante a investigação, foi realizada uma análise detalhada do impacto dos problemas de rádio na predição de largura de banda. Os resultados obtidos evidenciaram uma alta precisão da predição, com um erro médio na faixa de 6% a 12% para todos os cenários avaliados.

Os resultados do estudo indicam que o desempenho da rede LTE-*Advanced* é afetado significativamente pela carga de tráfego e pelo número de usuários simultâneos. Uma das limitações do estudo é que os resultados foram obtidos em um ambiente controlado e podem não ser generalizáveis para ambientes reais, onde a rede pode ser afetada por interferências e outros fatores externos. No entanto, os resultados do estudo são relevantes para o desenvolvimento de técnicas de gerenciamento de rede e para a melhoria do desempenho da rede LTE-*Advanced* em ambientes controlados.

O artigo de [Oussakel et al. \(2019a\)](#) propõe o uso de métricas de rádio como recursos-chave para prever a largura de banda que as conexões consideradas podem aproveitar nas próximas centenas de milissegundos. Como o fenômeno do rádio e a mobilidade do usuário não podem ser controlados, a solução proposta pelos autores é estimar a qualidade das transmissões em um futuro próximo para um usuário conectado. Para isso, métricas de rádio são usadas para prever a largura de banda. Com o uso de algoritmos de aprendizado de máquina supervisionado, como RF e SVM, o estudo visa prever a largura de banda recebida no *uplink* e até mesmo a quantidade de dados recebidos no lado do servidor durante um período específico, em uma escala muito baixa de 100 milissegundos.

Os resultados da pesquisa mostraram que, embora a predição da largura de banda para *milissegundo* com base em medições de rádio seja mais precisa, as previsões de largura de banda de *uplink* também são possíveis e úteis em serviços específicos de usuários. Os autores observaram que a precisão das previsões de largura de banda de *uplink* baseadas em medições de rádio é menor do que a de *downlink*, e isso pode ser atribuído à influência do canal de rádio nas transmissões de *uplink*. O estudo também enfatizou a importância das

métricas de rádio como uma fonte de informação valiosa para a predição de largura de banda em redes móveis, especialmente para as transmissões de *uplink*, que são mais sensíveis a essas métricas.

3.2.3 Predição de Vazão em Ambiente de Laboratório

O artigo de [Wei et al. \(2018\)](#) propõe um novo método para prever a vazão em redes de comunicação sem fio usando um modelo de *Neural Network Recorrente* (RNN). O objetivo do estudo é melhorar a qualidade de experiência do usuário em aplicações de transmissão de vídeo, como YouTube e Netflix. Para avaliar o método proposto, os autores coletaram dados usando um aplicativo desenvolvido para celular Android, usando tecnologia LTE. Os experimentos foram realizados em um ambiente estático de laboratório durante a manhã, tarde e noite. A metodologia proposta pelos autores envolve o uso de um modelo RNN para prever a vazão futura com base em dados históricos, como a vazão e a indicação da força do sinal recebido. Os autores compararam a precisão de predição do método proposto com outros métodos convencionais como o *Hidden Markov Model* (HMM) e o *Last Sample* (LS). A precisão do método proposto foi avaliada usando métricas como o *Average Error Rate* (AER) e o *Normalized Root Mean Square Error* (NRMSE). Os resultados mostraram que o método proposto usando RNN superou os métodos convencionais em termos de precisão de predição, reduzindo o erro de predição em até 29,39% em relação aos métodos LS e HMM. As principais contribuições do artigo incluem a proposta de um novo método para predição de vazão em redes de comunicação sem fio, com base em um modelo RNN, bem como a demonstração de sua eficácia em comparação com métodos convencionais.

3.2.4 Métodos de Predição em Dispositivos de Redes Sem Fio Baseados em Aprendizado de Máquina

O objetivo principal do artigo de [Baxhaku et al. \(2018\)](#) é propor uma solução para predição de canais em sistemas de comunicação sem fio que utilizam diferentes bandas de frequência para transmissão e recepção. Especificamente, o estudo se concentra na predição do canal de *downlink* com base nas informações do canal de *uplink*. A metodologia proposta envolve

o uso de uma rede neural treinada em um modelo padrão de canal para gerar estimativas grosseiras das variáveis do canal e, em seguida, utilizar um framework de otimização de antena única para obter estimativas mais precisas, que podem ser usadas para predição do canal de *downlink*. O trabalho foi implementado em rádios definidos por software e os resultados foram avaliados por meio de experimentos e simulações.

O estudo utiliza técnicas de aprendizado de máquina, especificamente redes neurais, para gerar predições de canais mais eficientes e precisas em comparação com as abordagens convencionais. Os resultados mostram que a solução proposta reduz a complexidade e o tempo de execução em pelo menos uma ordem de magnitude em comparação com as abordagens atuais, mantendo a qualidade de predição semelhante. As principais contribuições do estudo são a aplicação de aprendizado de máquina para predição de canais em sistemas de comunicação sem fio e a proposta de um método mais eficiente e flexível para predição de canais de *downlink*. No entanto, uma limitação do trabalho é a aplicabilidade limitada em cenários de múltiplas antenas. Em relação à predição de vazão de *uplink* em redes de telefonia móvel, este trabalho pode ser relevante, pois a predição de canal é um componente importante na predição de desempenho da rede.

3.2.5 Ambiente Real

Experimentos de ambiente real são conduzidos em situações e contextos normais de uso, em ambientes do mundo real. Dessa forma, é possível obter dados que refletem a experiência real dos usuários, sem interferências artificiais ou limitações do ambiente de laboratório. Ao realizar experimentos em ambientes do mundo real, os pesquisadores têm a oportunidade de coletar dados de dispositivos móveis em operação real, o que pode levar a uma maior variabilidade e complexidade dos resultados obtidos. No entanto, essa abordagem também apresenta desafios em termos de controle experimental e análise de dados, já que há uma série de fatores externos que podem influenciar os resultados.

3.2.6 Aprendizado Supervisionado com Regressão

O artigo de [Eyceyurt *et al.* \(2022\)](#) apresenta uma abordagem promissora para prever a vazão de dados de *uplink* em redes de comunicação sem fio. Uma das principais contribuições do estudo é o uso de aprendizado de máquina para prever a vazão de *uplink* com base em medidas físicas da rede. O modelo de regressão proposto utiliza características como a força do sinal para prever a vazão de dados de *uplink*. Os resultados dos experimentos mostram que a abordagem proposta apresenta um desempenho significativo.

Entre as principais vantagens deste estudo está a sua abordagem inovadora para prever a vazão de *uplink* em redes de comunicação sem fio. O uso de aprendizado de máquina usando as características físicas pode ajudar a melhorar a qualidade de serviço em aplicações de rede que exigem alta vazão de dados. Além disso, os resultados dos experimentos mostram que a abordagem proposta é eficaz em um ambiente de rede real: utilizou-se como principal algoritmo o KNN e obteve 0,92 de (R^2) e 3,76 de erro quadrático médio (MSE).

No entanto, algumas limitações deste trabalho relacionado devem ser destacadas. A abordagem proposta pode exigir medidas físicas adicionais e pode não ser adequada para todos os tipos de dispositivos ou aplicações de rede. O artigo de [Eyceyurt *et al.* \(2022\)](#), embora apresente algumas limitações, traz contribuições relevantes para o avanço da pesquisa em predição de vazão em redes de comunicação sem fio.

O trabalho de [Minovski *et al.* \(2021\)](#) desenvolveu e validou um modelo de aprendizado de máquina para prever a vazão disponível em redes, tanto a 4G LTE quanto a 5G não autônomo². O objetivo do estudo foi coletar as condições de vazão em diferentes cenários de rede, incluindo áreas urbanas, suburbanas e rurais, bem como testes em grandes multidões.

Os resultados do experimento mostraram que o modelo proposto obteve uma precisão de 93% na predição de vazão em redes LTE e 84% em redes 5G não autônomas, com erros quadráticos médios de 0,06 e 0,17, respectivamente. Esses resultados indicam que o modelo é altamente preciso e pode ser aplicado para melhorar a qualidade de serviço em redes de comunicação sem fio. Além disso, o modelo é baseado em métricas de camada inferior que descrevem o ambiente de rádio, o que pode fornecer informações valiosas para os provedores

²5G não autônomo é uma forma de implementação do 5G que utiliza a infraestrutura de rede 4G LTE já existente para fornecer conectividade inicial aos dispositivos 5G, antes de aproveitarem totalmente as vantagens do 5G em termos de velocidade, latência e capacidade de rede.

de serviços de rede em relação ao desempenho da rede em diferentes locais e condições.

O trabalho de Al-Thaedan *et al.* (2024) foi desenvolvido um modelo preditivo utilizando algoritmos de aprendizado de máquina para estimar a vazão de downlink em redes celulares 4G-LTE e 5G. O estudo utilizou dados reais coletados de operadoras, abrangendo diferentes horários do dia, incluindo horários de pico e normais, com as coletas realizadas por meio de um escâner de RF acoplado a um carro em movimento.

As variáveis utilizadas para treinar os modelos incluem parâmetros de qualidade de sinal como RSRP, RSSI, SINR e RSRQ. A seleção dessas características foi realizada com base em coeficientes de correlação, sendo o RSRP o parâmetro mais correlacionado com a vazão. O modelo proposto alcançou precisão entre 89% e 96% no coeficiente de determinação (R^2), evidenciando a eficácia da abordagem. O artigo ainda destaca a intenção de, em trabalhos futuros, aplicar técnicas similares para a predição da vazão de uplink com uso de aprendizado profundo, especialmente voltado para aplicações em veículos autônomos.

O trabalho de Al-Thaedan *et al.* (2023) propôs um modelo preditivo para estimar a vazão de *downlink* em redes 4G-LTE por meio de algoritmos de aprendizado de máquina. Foram utilizados dados reais coletados de 377 células de três operadoras distintas na cidade de Najaf, Iraque, durante testes de direção (*drive tests*) em diferentes horários do dia. Os dados incluíram métricas como RSRP, RSRQ, SINR, RSSI e coordenadas GPS. Os autores avaliaram o desempenho de quatro modelos: SVR, Regressão Linear, KNN e *Decision Tree Regression* (DTR), sendo este último o que apresentou os melhores resultados, com coeficiente de determinação (R^2) de até 0.99 e RMSE de até 1.484 Mbps. A pesquisa demonstrou que os modelos DTR e KNN são mais eficazes para previsão de vazão, evidenciando sua aplicabilidade para planejamento e gerenciamento de redes móveis.

3.2.7 Regressão e Classificação

O trabalho de Narayanan *et al.* (2020) visa mapear e prever a vazão da rede 4G e 5G *mmWave*³ em um ambiente comercial, usando técnicas de ML. O estudo coletou dados de campo em um ambiente comercial com várias torres de células 5G *mmWave* em diferentes

³5G *mmWave* é uma versão do 5G que utiliza faixas de frequência entre 30 e 300 *gigahertz*, oferecendo largura de banda e velocidades de transferência superiores às das frequências mais baixas, como no 4G.

localizações, usando um dispositivo móvel de teste. A coleta de dados incluiu medições de sinais de rádio, bem como informações contextuais, como a localização do dispositivo e a direção da antena. Com base nesses dados, os autores desenvolveram um modelo de aprendizado de máquina para prever a vazão da rede 5G *mmWave*.

A coleta de dados durou em torno de 6 meses e a velocidade média do automóvel foi de 45 km. Os resultados mostraram que o modelo de aprendizado de máquina desenvolvido foi capaz de prever a vazão com alta precisão, com um *Mean Absolute Error* (MAE) de 6,5 Mbps e uma precisão de 91,4%. Além disso, os autores utilizaram esses dados para mapear a cobertura da rede 5G *mmWave* em um ambiente comercial, o que pode ser útil para planejamento e otimização da rede.

As principais contribuições deste estudo são a coleta de dados de campo em um ambiente comercial para análise de vazão da rede 5G *mmWave*, a construção de um modelo de aprendizado de máquina para predição de vazão e o mapeamento da cobertura da rede 5G *mmWave* em um ambiente comercial. Destacando que foi realizado aprendizado de máquina supervisionado, tanto classificação quanto regressão.

O trabalho de [Palaios et al. \(2023\)](#) propõe uma abordagem detalhada para predição de vazão em redes veiculares, com foco tanto em *uplink* quanto *downlink*, utilizando aprendizado de máquina supervisionado. A pesquisa foi conduzida em uma rede LTE controlada, com coleta de dados em mais de 3000 km percorridos por quatro veículos equipados com dispositivos de medição (DME) e *consumer-grade* (CE). O estudo avaliou diversos cenários, como ambientes urbanos, rurais e rodoviários. Foram analisadas diferentes estratégias de divisão dos dados (como aleatória, por tempo, por execução de experimento e por *folds*) e de conjuntos de variáveis de entrada. O algoritmo *Gradient Boosted Trees* (GB) obteve o melhor desempenho em predições de vazão, com MAE de 2,46 Mbps para *downlink* e 1,08 Mbps para *uplink*. O estudo destaca a importância de fatores como a variabilidade do ambiente de rádio, a frequência de amostragem, a presença de *concept drift*⁴ e o uso de técnicas explicáveis (XAI)⁵ para entender os modelos. Os resultados demonstram que dispositivos de baixo custo também podem ser utilizados com boa precisão para tarefas de predição de QoS.

⁴ *Concept drift* é a mudança nos padrões ou distribuições dos dados ao longo do tempo, reduzindo a precisão de modelos treinados. Em redes móveis, resulta de variações no ambiente de rádio, tráfego ou infraestrutura.

⁵ XAI refere-se as técnicas que visam tornar os modelos de ML mais transparentes e interpretáveis.

3.2.8 Regressão e Series Temporais

O objetivo central do artigo de [Elsherbiny et al. \(2020\)](#) é analisar e comparar dois métodos para análise de vazão de rede: aprendizado de máquina clássico e predição de séries temporais. O estudo utilizou dados coletados em redes 4G LTE em um ambiente de transporte público em Kingston, Ontário, visando modelar e prever a vazão de rede.

Os autores realizaram uma análise comparativa de alguns modelos de aprendizado de máquina, incluindo RF para regressão. Além disso, modelos de predição de séries temporais foram utilizados para prever valores futuros de vazão. Os diferentes métodos de modelagem foram comparados usando diversas métricas de avaliação, a fim de determinar o melhor modelo para predição de vazão.

Os resultados indicaram que a abordagem de RF para regressão obteve a maior precisão com 0,78 de R^2 , devido à sua capacidade de generalização e redução da variância. No entanto, a precisão da abordagem de séries temporais foi inferior em comparação com as outras abordagens de aprendizado de máquina.

As principais contribuições deste estudo são a análise comparativa de diferentes modelos de aprendizado de máquina e predição usando séries temporais para estimativa de vazão em redes 4G LTE, bem como a identificação da abordagem de RF como a mais precisa para esse fim. No entanto, uma limitação deste estudo é que ele foi realizado em um ambiente específico de transporte público em Kingston, Ontário, o que pode afetar a generalização dos resultados para outros ambientes e cenários de rede.

Já o intuito do trabalho de [Jha e Vijayarajan \(2020\)](#) foi explorar técnicas que possam ser utilizadas para prever a vazão da internet móvel de um usuário por meio da utilização de vários parâmetros básicos facilmente acessíveis. Essa predição ajudaria o servidor a prever a vazão do usuário com antecedência e servir dados ou conteúdo a uma taxa de bits adequada, o que levaria a uma menor ocorrência de *buffering*⁶ excesso de carregamento e, conseqüentemente, a uma melhor *Quality of Experience* (QoE).

Os autores utilizaram um conjunto de dados coletados pelo *Mobile and Internet*

⁶*buffering* é o processo de armazenar temporariamente dados em um buffer (memória intermediária) durante a transmissão de mídia para evitar interrupções na reprodução causadas por variações na vazão ou atrasos na transmissão. Ele permite uma reprodução mais suave e contínua de conteúdo em tempo real, como *streaming* de vídeo ou áudio.

Systems Laboratory da *University College Cork*. O conjunto de dados contém 20 classes, cada uma representando uma parte fundamental na identificação da conexão 4G LTE, como SNR e RSRQ, e dados mais centrados no usuário, como tempo e localização. A partir dessas classes, os autores escolheram apenas algumas que poderiam ser facilmente obtidas do lado do cliente para vários dispositivos, permitindo fácil escalabilidade e sem invasão da privacidade do usuário. Na execução do experimento, os autores realizaram uma divisão do conjunto de dados de treinamento, a fim de criar um conjunto de testes que pudesse ser utilizado para verificar a precisão dos resultados obtidos.

Como resultado, foi constatado que a solução para este problema se mostrou mais eficiente quando implementada como um modelo de regressão, em contraposição a um modelo de série temporal. O modelo ARIMA não foi muito preciso e é mais adequado para um conjunto de dados de série temporal. Os modelos de árvore de decisão e *Naïve Bayes* apresentaram resultados similares em termos de precisão e puderam lidar com vários saltos na vazão, prevendo valores mais próximos do real. Com base nos dados de treinamento de cada usuário individual, esses modelos teriam diferentes parâmetros referentes a cada usuário específico.

No artigo não foi abordado qual modelo apresentou melhores resultados nem quais métodos de avaliação foram utilizados. Além disso, não foram mencionados detalhes sobre a base de dados como sua organização, método de coleta, velocidade média e outras características relevantes que poderiam ter sido consideradas.

A principal contribuição do trabalho é o uso de técnicas de aprendizado de máquina para prever a vazão da internet móvel de um usuário com base em vários parâmetros básicos facilmente acessíveis. O que pode levar a uma melhor qualidade de experiência do usuário e à otimização do uso de dados móveis. Na experimentação, o modelo de árvore de decisão permaneceu como a abordagem mais promissora.

O estudo de [Mostafa et al. \(2022\)](#) propôs um sistema híbrido de predição de vazão *downlink* para usuários finais em redes LTE, integrando técnicas de aprendizado de máquina e aprendizado profundo. A abordagem utiliza um modelo sazonal ARIMA para prever três KPIs críticos da rede (tráfego de *downlink*, número de PRBs e utilização de PRBs de *downlink*), que são, posteriormente, utilizados como entrada para uma Rede Neural de Regressão

(RNN), com o objetivo de estimar a vazão final do usuário. Os dados utilizados abrangeram o período de maio de 2021 a janeiro de 2022, com os resultados avaliados por métricas como RMSE, MAE e MSE. A média do RMSE obtida foi de 2,3 Mbps, representando aproximadamente 10% da média real de vazão dos usuários (22,7 Mbps). Os autores destacam que essa abordagem melhora significativamente a acurácia dos métodos estatísticos tradicionais aplicados ao planejamento de redes celulares. A RNN apresentou melhor desempenho em relação ao ARIMA, sendo considerada a principal contribuição do estudo para a predição de vazão em ambientes LTE.

O estudo de [Batool et al. \(2024\)](#) apresenta uma abordagem baseada em aprendizado profundo para predição da vazão em redes 5G, utilizando os modelos *Long Short-Term Memory* (LSTM), *Bidirectional LSTM* (BiLSTM) e Redes Neurais Artificiais (ANN). O objetivo é enfrentar os desafios associados à natureza dinâmica das redes móveis e otimizar a alocação de recursos de rede, assegurando a qualidade de serviço. Os autores utilizaram uma base de dados de produção gerada a partir de padrões de mobilidade estática e veicular, com aplicações de *streaming* de vídeo e download de arquivos. Os dados foram coletados via o aplicativo Android G-Net e incluem métricas como RSRP, RSRQ, RSSI, SINR e CQI. O modelo BiLSTM demonstrou alta capacidade de adaptação e previsão, obtendo bons resultados com base em métricas como MSE e R^2 . Como contribuição, o estudo destaca a eficácia dos modelos recorrentes para previsão de vazão em redes 5G, reforçando seu potencial para manutenção adaptativa da rede e gestão proativa de congestionamentos.

O artigo de [Biswas et al. \(2024\)](#) propõe uma solução simples e de baixa complexidade computacional para a predição de vazão em redes 5G. A abordagem se baseia em uma regressão linear múltipla dos parâmetros atuais da rede e da vazão atual, integrada a um filtro de Kalman⁷ que ajusta as estimativas futuras com base na variância do erro de medição e de predição, aspectos geralmente ignorados por trabalhos anteriores. Os experimentos foram realizados com sete conjuntos de dados públicos de redes 5G, avaliando diferentes janelas de predição. Os resultados mostraram que o método proposto superou algoritmos tradicio-

⁷O *filtro de Kalman* é um algoritmo recursivo utilizado para estimar o estado de sistemas dinâmicos a partir de medições sujeitas a ruído. Ele combina predições baseadas em modelos com observações reais para produzir estimativas mais precisas e confiáveis. É amplamente empregado em áreas como navegação, robótica e telecomunicações, ajustando continuamente as estimativas conforme novas medições são incorporadas.

nais de aprendizado de máquina e aprendizado profundo em termos de precisão, tempo de inferência e re-treinamento. Além disso, ao ser aplicado a cenários de *streaming* adaptativo com controle preditivo de taxa de bits, o modelo demonstrou melhorar significativamente a qualidade de experiência dos usuários em transmissões de vídeo sob demanda e ao vivo.

O trabalho de Lawal (2020) busca apresentar uma abordagem alternativa de vazão do usuário em uma rede para dispositivos móveis, utilizando o algoritmo *Group Method of Data Handling* (GMDH). O estudo mediu a vazão durante três semanas e usou esses dados para treinar um modelo baseado em GMDH para prever a vazão para a quarta semana. A estratégia de modelagem adotada emprega um único modelo para prever a vazão para toda a semana. Desta forma Lawal (2020) obteve sucesso, em seu estudo, alcançando um erro médio percentual de 1,87%.

Os dados foram coletados a partir de estações de base da operadora. Os dados são caracterizados por uma sazonalidade de 24 horas e também exibem alta variabilidade. Para a coleta de dados, o artigo não especifica o método exato usado, mas sugere que os dados foram coletados a partir de estações de base da operadora. Não há informações adicionais sobre a frequência de amostragem ou sobre como os dados foram pré-processados antes do treinamento do modelo.

Os resultados experimentais obtidos pelo modelo baseado em GMDH demonstraram um desempenho muito bom, alcançando um *Mean Absolute Percentage Error* (MAPE) de 1,87%, o que é melhor do que o modelo *Long Short-Term Memory* (LSTM), baseado em RNN. Além disso, a capacidade do algoritmo GMDH de selecionar apenas as variáveis de entrada eficazes durante o treinamento do modelo reduz a dimensionalidade dos dados de treinamento em até 43%, permitindo o desenvolvimento de um modelo mais simples e fácil de ser interpretado para predição de vazão.

As principais contribuições do trabalho de Lawal (2020) incluem a apresentação de uma abordagem alternativa para a predição de vazão do usuário móvel, utilizando o algoritmo GMDH. Além disso, o estudo demonstra que o modelo baseado em GMDH pode alcançar uma performance comparável ao modelo LSTM baseado em RNN, com a vantagem de selecionar apenas as variáveis de entrada eficazes durante o treinamento do modelo, reduzindo assim a complexidade do modelo e tornando-o mais fácil de ser interpretado. No entanto, uma das

limitações do estudo é o período de coleta de dados, limitado a três semanas, o que pode limitar a capacidade do modelo de prever a vazão em períodos mais longos.

O trabalho de Schmid *et al.* (2019) tem como objetivo apresentar uma abordagem baseada em aprendizado profundo para prever a vazão de redes móveis de forma independente de localização. Para isso, os autores coletaram um grande conjunto de dados de parâmetros de rede e medições correspondentes de vazão em várias localizações. Em seguida, eles realizaram o pré-processamento dos dados e utilizaram *Convolutional Neural Networks* (CNNs) para aprender a relação entre esses parâmetros de rede e a vazão, com foco especial em modelos de RNN e LSTM, obtendo um MAE de 305,98 Kbps.

Para a coleta de dados foi utilizado o módulo MC7304 LTE, da *Sierra Wireless*, durante um período de 4 dias. Os principais parâmetros de rede coletados foram RSRQ, RSSI, *Round Trip Time* (RTT)⁸ e vazão. Os resultados obtidos mostram que o modelo proposto pode prever com precisão a vazão em diferentes localizações, o que é importante para garantir a comunicação confiável e segura entre veículos e infraestrutura em ambientes de condução autônoma.

As principais contribuições do artigo incluem a demonstração da eficácia de técnicas de aprendizado profundo para previsão de vazão de redes móveis e a proposta de uma abordagem independente de localização para esse problema. Em relação à dissertação, o trabalho tem relevância para a previsão de vazão de uplink dos dados em redes móveis, uma vez que apresenta uma abordagem baseada em aprendizado profundo para esse problema, com foco em *downlink*.

O trabalho de Kousias *et al.* (2019) tem como objetivo propor uma solução de aprendizado de máquina supervisionada para estimar a vazão de *downlink* em redes de banda larga móvel, com alta precisão e redução significativa no consumo de dados. O método de coleta de dados envolve a exploração de metodologias de teste controlado e de *crowdsourcing*⁹, utilizando a plataforma RTR-NetTest e sua versão empacotada MONROE-NetTest. Os dados foram coletados de diversos recursos, incluindo a categoria do dispositivo LTE, o provedor de

⁸O RTT é a medida do tempo, em milissegundos, que um pacote de dados leva para ir do dispositivo de origem ao destino e retornar.

⁹*Crowdsourcing* é uma abordagem que envolve a participação de várias pessoas para realizar tarefas, coletar dados ou fornecer feedback, geralmente de forma distribuída e colaborativa.

serviços móveis, o RSRP, o RSRQ, a latitude e a longitude do dispositivo, a rede de acesso da operadora e outros recursos disponíveis na Android API. Além disso, foram realizados testes de velocidade de upload e download para coletar informações de vazão.

Foram utilizados três algoritmos de ML: Regressão linear múltipla (MLR), *Support Vector Regression* (SVR) e RF. Para selecionar as características mais importantes para o modelo de predição, foi empregado o processo de seleção de recursos. O *Forward Selection*, um algoritmo iterativo que utiliza o ajuste do modelo como critério de seleção, foi utilizado para selecionar as características mais importantes em três clusters diferentes, com base no custo de coleta dos dados.

O estudo apresentou um modelo de aprendizado de máquina supervisionado que reduz o consumo de dados e oferece estimativas precisas de vazão *downlink* em redes de banda larga móvel. Os resultados mostraram que o modelo reduziu o consumo de dados em 39,7% e apresentou um erro percentual médio absoluto de 5,55%. Além disso, foi observado que a inclusão de informações de *downlink* não aumentou significativamente a precisão das estimativas de vazão. A contribuição do trabalho está na proposta de um modelo eficiente e econômico na coleta e seleção de dados.

O trabalho de [Markappa et al. \(2024\)](#) investigou a predição da taxa de dados de *uplink* em redes 5G não autônomas, com foco em aplicações para sistemas de transporte inteligente. Para isso, foram utilizados dados reais coletados com um dispositivo embarcado (Klas Telecom TRX-R6) contendo um modem 5G, em movimento por diferentes regiões da cidade de Cork. Os autores propuseram o uso de variantes de redes neurais recorrentes (RNN), especificamente os modelos LSTM e GRU, que demonstraram desempenho promissor na predição com múltiplas combinações de lags e épocas de treinamento. As métricas utilizadas foram RMSE e RMSLE. Os resultados indicaram que o modelo GRU apresentou desempenho superior ao LSTM na maioria dos cenários testados, especialmente em contextos com dados contendo valores nulos ou zero nos parâmetros de rede 5G, o que é comum em ambientes dinâmicos. A pesquisa reforça o potencial de modelos baseados em aprendizado profundo para antecipar variações na qualidade do sinal de *uplink* em redes móveis.

3.2.9 Otimização de Taxa de Uplink

Os autores [Gaber *et al.* \(2020\)](#) visam apresentar uma abordagem para aprimorar o controle de potência de *uplink* em redes 4G utilizando técnicas de aprendizado de máquina não supervisionado. O estudo busca melhorar a qualidade do serviço para os usuários finais, reduzindo os custos operacionais das operadoras.

Para atingir esse objetivo, foi proposto um modelo de clusterização baseado em *K-means* para agrupar os usuários em diferentes categorias de consumo. Em seguida, um modelo de aprendizado por reforço foi desenvolvido para ajustar dinamicamente a potência de transmissão de cada usuário, buscando maximizar a qualidade do serviço e minimizar o consumo.

Os resultados obtidos no estudo de [Gaber *et al.* \(2020\)](#) mostram que a abordagem proposta pode ajustar a potência de transmissão dos usuários de forma mais eficiente do que abordagens tradicionais baseadas em algoritmos de busca exaustiva. A abordagem de clusterização foi eficaz em agrupar usuários com comportamentos semelhantes de consumo de *uplink*, permitindo que o modelo de aprendizado de máquina não supervisionado fosse aplicado com sucesso em tempo real.

Embora o objetivo do estudo não esteja diretamente relacionado à predição de vazão de *uplink*, a abordagem proposta pode ser vista como um passo importante para melhorar a eficiência e a qualidade do serviço em redes 4G. Uma melhor gestão da potência de transmissão dos usuários pode impactar diretamente na vazão de *uplink*, permitindo uma utilização mais eficiente do espectro e dos recursos da rede. Portanto, o estudo [Gaber *et al.* \(2020\)](#) apresenta contribuições relevantes para o campo da predição de vazão em redes móveis.

3.2.10 Uso de KPI para Estimativa de Outros Casos

O foco do estudo de [Minovski *et al.* \(2019\)](#) foi utilizar o aprendizado de máquina para analisar e estimar a qualidade da experiência de vídeo em redes celulares sem fio. Para isso, os autores propõem uma abordagem alternativa que utiliza RF para estimar a qualidade de vídeo com base nas métricas de qualidade da rede como RSSI, RSRP, RSRQ e *Carrier to Interference plus Noise Ratio* (CINR), além de realizar uma análise de causa raiz para descrever a decisão

das estimativas.

Os resultados mostraram que a abordagem proposta conseguiu estimar a qualidade de vídeo com alta precisão, considerando várias métricas de qualidade como *Mean Opinion Score* (MOS), *Peak Signal-to-Noise Ratio* (PSNR), atraso de quadro, saltos de quadro e desfoque. Isso pode servir para melhorar a qualidade dos serviços de vídeo. Como contribuição, o estudo propõe uma abordagem alternativa para estimar a qualidade de experiência do usuário com vídeo, utilizando técnicas de aprendizado de máquina e métricas de qualidade de rede.

Relacionando com a predição de vazão de *uplink*, pode-se dizer que a abordagem proposta pelo estudo pode ser complementar a outras técnicas de predição de vazão, como algoritmos de aprendizado de máquina e ferramentas de monitoramento de rede. A utilização de métricas de qualidade de rede pode fornecer informações adicionais para a predição de vazão, permitindo uma melhor adaptação da qualidade de experiência de vídeo em redes celulares sem fio. Além disso, como o estudo aponta, é possível explorar e testar métricas adicionais de rádio/rede e treinar o modelo em casos mais singulares para melhorar ainda mais a precisão.

3.3 Comparativo dos Trabalhos Correlatos

Esta seção apresenta um comparativo dos trabalhos correlatos abordados nesta dissertação. As [Figura 3.2](#) e [Figura 3.3](#) mostram o comparativo dos 23 trabalhos identificados na RSL, destacando os autores do artigo e o ano de publicação, além das informações sobre a coleta de dados como a velocidade média mantida pelo automóvel, o ambiente em que o experimento foi realizado, o dispositivo utilizado para a coleta de dados, o período de coleta e o aplicativo utilizado. Além disso, é abordada a descrição do experimento especificado, destacando pontos importantes para serem reproduzidos. Essas informações são importantes para entender as abordagens adotadas pelos autores em relação à coleta e análise de dados em experimentos de rede para dispositivos móveis, proporcionando *insights* valiosos dados para a elaboração de futuros trabalhos nessa área.

No contexto do aprendizado de máquina, as figuras destacam o algoritmo que obteve o melhor desempenho. As métricas utilizadas para avaliar o desempenho também foram

destacadas. Além disso, foi analisada a característica do trabalho em relação à sua abordagem de *uplink* ou *downlink* e a que geração da rede estava direcionada. Essas informações são extremamente importantes. Tanto para entender as abordagens adotadas pelos autores na análise de desempenho de redes móveis quanto como as técnicas de aprendizado de máquina podem ser aplicadas de forma eficaz para melhorar a qualidade e desempenho da rede.

Na seção 3.2, foram destacados os objetivos e as contribuições de cada artigo estudado. Em resumo, os artigos foram classificados em dois tipos: aqueles que realizaram experimentos em ambiente real e aqueles que realizaram experimentos em laboratórios. Este estudo utilizou a coleta de dados em ambiente real. A coleta de dados de rede de dispositivos móveis em ambiente real é essencial para compreender e analisar o comportamento e as características desses dispositivos em situações reais de uso. Assim, é possível uma investigação mais precisa e completa do desempenho e comportamento dos usuários, considerando fatores como mobilidade, interferência e carga da rede. Essas informações são cruciais para entender a performance real dos dispositivos móveis em cenários reais e para desenvolver técnicas eficazes de otimização e melhoria das redes móveis.

Os artigos analisados utilizaram principalmente smartphones da marca Samsung, enquanto este trabalho utilizou dispositivos da marca Motorola. Quanto ao período de coleta de dados, alguns artigos não especificaram o período, enquanto este estudo realizou a coleta de dados realizada em diferentes dias e horários, a fim de garantir uma amostra variada e representativa. Alguns artigos utilizaram aplicativos de coleta de dados, enquanto outros criaram suas próprias ferramentas. Neste trabalho, foi desenvolvida uma ferramenta própria para coleta de dados de parâmetros de rede, o que possibilita a coleta de dados específicos e necessários para a análise.

Com relação aos algoritmos utilizados, os artigos estudados concentraram-se principalmente em algoritmos de regressão como RF, RNN e SVM. Para classificação, foi utilizado o *K-means*, e alguns trabalhos também utilizaram algoritmos de séries temporais. No presente estudo, foram utilizados cinco algoritmos para a predição de vazão de *uplink* em redes 4G LTE. A escolha dos algoritmos se baseia em sua eficácia comprovada em estudos anteriores e em sua capacidade de lidar com a complexidade dos dados coletados em ambiente real. Este trabalho utilizará diversas métricas de avaliação de desempenho para avaliar os algoritmos

utilizados, permitindo uma comparação mais abrangente dos resultados obtidos.

Este estudo busca trazer contribuições para a predição de vazão de *uplink* em redes de dispositivos móveis, utilizando abordagens de aprendizado de máquina supervisionado e métricas da camada física da rede. Espera-se que os resultados deste trabalho possam auxiliar na melhoria da qualidade de serviço e na otimização do uso de recursos de rede, proporcionando uma melhor experiência do usuário. Além disso, este estudo pode incentivar a pesquisa em outras áreas de aplicação de aprendizado de máquina em redes móveis.

ID	Artigo	Velocidade Média	Ambiente	dispositivo	Período Coleta	Aplicativo para Coleta	Descrição Especificado do Experimento	Principal Algoritmo	Avaliação de Desempenho	Variáveis Utilizadas	Uplink	downlink	Rede
A01	Al-Thaedan et al. (2024)	10 km	Real	Scanner de RF	Não informa	Scanner de RF foi acoplado a um carro	Sim	Random Forest	R ² ; MSE; MAE	RSRP, RSRQ, RSSI, SNR	Não	Sim	4G-LTE e 5G
A02	Batool, Fouda, Fadlullah (2024)	Não informa	Real	Android	Não informa	Base de Dados Outros	Não informa	BiLSTM	MSE e R ²	RSRP, RSRQ, RSSI, SINR, CQI	Não	Sim	5G
A03	Biswas, Chakraborty e Palit (2024)	Não informa	Real	Android	Não informa	Base de Dados Outros	Não informa	KFTP (Kalman Filter + MLR)	R ² ; MAE	RSRP, SINR	Não	Sim	5G
A04	Markappa et al. (2024)	Não informa	Real	Klas Telecom TRX-R6 com modem Telit FN980	Não informa	UDP via iPerf	Não informa	GRU	RMSE e RMSLE	RSRP, RSRQ, RSSI, NR-RSRP, NR-RSRQ, NR-RSSI	Sim	Não	4G e 5G
A05	Varghese et al. (2024)	Não informa	Real	Drone com conexão celular	Não informa	Ferramentas personalizadas com suporte da Rohde & Schwarz e MONROE-Nettest	Sim	Radom Forest	MAE; MSE; RMSE; MAPE	NR RSRP, RSRQ, SNR, UL Throughput, UL Modulation, PCI, altitude, latitude, longitude, velocidade	Sim	Não	5G
A06	Al-Thaedan et al. (2023)	Não informa	Real	Via Driving test	Não informa	Não informa	Não informa	Decision Tree Regression	R; RMSE	RSRP, RSRQ, RSSI, SINR, localização GPS	Não	Sim	4G LTE
A07	Palaios et al. (2023)	Não informa	Real	Equipamentos DME e CE (consumer-grade e equipamentos dedicados em veículos)	Aproximadamente e 1 semana	MobileInsight, G-Net Track Pro, iPerf	Sim	Gradient Boosted Trees (GB)	R ² ; MAE; MedA; RMSE; MAPE	RSRP, RSSI, SINR, Cell Load, RTT, Distância da célula, Velocidade do veículo	Sim	Sim	4G
A08	Mostafa et al. (2022)	Não Informa	Real	Não informado	3 meses	Ferramenta própria	Sim	Random Forest	R2; RMSE; MAE	RSRP, RSRQ, CQI, SNR	Não	Sim	4G
A09	Eyceyurt et al. (2022)	40 km/h	Real	Samsung Note 10	14 a 25 de setembro de 2019	RantCell Test Analytics	Não	RNN	R2; RMSE	RSRP, RSRQ e SNR	Sim	Não	4G
A10	Minovski et al. (2021)	63 km/h	Real	Samsung S20	Não Informa	Ferramenta própria, Teme Pocket para registrar e software TWAMP	Sim	XGBoost	R2;MSE	RSSI, RSRP, RSRQ e SINR	Sim	Sim	4G e 5G
A11	Elsherbiny et al. (2020)	Não Informa	Real	Não Informa	Não Informa	Não Informa	Não	Random Forest	R2; RMSE;	RSRP, RSSI, SNR, RSRQ, data/hora, longitude e latitude.	Não	Sim	4G
A12	Narayanan et al. (2020)	45 km/h	Real	Samsung Galaxy S10 5G	6 meses	Ferramenta própria e compilação com o iPerf 3.7	Sim	Seq2seq	RMSE; MAE	(L+M) + Radio Type + LTE Signal Strength + 5G Signal Strength + Horizontal Handoff + Vertical Handof	Não	Sim	4G e 5G
A13	Jha e Vijayarajan (2020)	Não Informa	Real	Não Informa	Não Informa	Base de Dados Outros	Não	Regressão Linear e Arvore de Descição	Não Informa	timestamp, localização, speed, download bitrate, Coordinate Cluster	Não	Sim	4G

Figura 3.2: lista dos comparativos dos trabalhos correlatos - parte 1
 Fonte: elaboração própria (2025).

ID	Artigo	Velocidade Média	Ambiente	dispositivo	Período Coleta	Aplicativo para Coleta	Descrição Especificado do Experimento	Principal Algoritmo	Avaliação de Desempenho	Variáveis Utilizadas	Uplink	downlink	Rede
A14	Gaber et al. (2020)	Não Informa	Real	Não Informa	1 mês	Não Informa	Não	K-Means	Não Informa	Pusch sinr, 64QAM, UL Traffic, Uplink Interference	Sim	Não	4G
A15	Raca et al. (2020)	Não Informa	Real	Samsung J5	Não Informa	Não Informa	Sim	LSTM	R2; RMSE; MAPE;	SNR; CQI; RSRQ, Velocidade do Dispositivo, Throughput	Não	Sim	4G
A16	Lawal (2020)	Não Informa	Real	Não Informa	4 semanas	Estações Base da Operadora	Não	GMDH	RMSE; MAE; MAPE	Não Informa	Não	Sim	3G
A17	Oussakel et al. (2019)	Não Possui (Estático)	Controlado	Não Informa	Não Informa	IPERF3	Sim	Random Forest	RMSE; MAE; MAPE	Timing Advance (TA); SNR; received UL-CQI; PUCCH received power and noise power; PUCCH threshold.	Sim	Não	4G
A18	Oussakel, Owezarski, Berthou (2019)	Não Possui (Estático)	Controlado	Samsung Galaxy J3 2017	Não Informa	IPerf; Programa LabVIEW.	Sim	Random Forest	Não Informa	Received Interference Power; Uplink Received Signal Strength Indicator; SNR; Rx_power; The received power; RSSI	Sim	Não	4G
A19	Minovski et al. (2019)	40 km/h	Real	Samsung Galaxy S9	Não Informa	Aplicativo TWAMP modificado; TEMS Pocket;	Sim	Random Forrest	R2; RMSE; MSE	RSSI, RSRP, RSRQ e CINR;	sim (Foco em previsão de canal)	sim (Foco em previsão de canal)	4G
A20	Bakshi et al. (2019)	Não Possui (Estático)	Controlado	Não Informa	Não Informa	Não Informa	Não	Rede Neural Recorrente (RNN)	Não Informa	Não Informa	Não	sim (foi utilizado, mas ele não foi o foco)	4G
A21	Schmid et al. (2019)	50 km/h	Real	Módulo MC7304 LTE da Sierra Wireless	13 à 16 de novembro de 2017 (4 dias)	TCPAnalyzer	Sim	LSTM RNN	MAE; MSE; MRE; MSRE	RSRQ, RSSI, Tempo de Ida e Volta (RTT) e Taxa de Transferência	Não	Sim	4G
A22	Kousias et al. (2019)	Não Informa	Real	Sansung S8	18 de Janeiro e 18 de Outubro (10 meses)	RTR-NetTest e MONROE-NetTest	Sim	SVR	R2; RMSE; MAE	categoria de dispositivo LTE, operadora, RSSI, RSRP, RSRQ, latitude, longitude, nome da operadora móvel virtual (MVNO), informações temporais e estatísticas de latência.	Não	Sim	4G
A23	Wei et al. (2018)	Não Possui (Estático)	Controlado	Não Informa	Não Informa	Ferramenta Própria	Não	Rede Neural Recorrente (RNN)	ARE; NRMSR	throughput; RSSI; Qualidade da Comunicação	Não	Sim	4G
Autor	Lima et al. (2023)	40Km/h	Real	Motorola Moto G10	03 de março à 24 de março de 2023 (22 dias)	Ferramenta Própria	Sim	Random Forest	R2; RMSE; MAR; MSE; MAPE; ARE; MRE; MSRE; NRMSE	RSRP, RSRQ, SNR, RSSI, CQI, localização, timestamp	Sim	Não	4G LTE

Figura 3.3: lista dos comparativos dos trabalhos correlatos - parte 2
 Fonte: elaboração própria (2025).

3.4 Resumo

Este capítulo apresentou os trabalhos correlatos identificados por meio da revisão sistemática da literatura. A [Seção 3.1](#) descreveu os principais estudos existentes na área relacionados à predição de vazão, detalhando o protocolo e a condução da revisão. Na [Seção 3.2](#), foram listados 23 artigos analisados nesta pesquisa, classificados conforme o ambiente experimental utilizado como controlado ou real.

Por fim, a [Seção 3.3](#) apresentou o comparativo entre os trabalhos analisados e esta dissertação. A principal distinção esteve na abordagem investigativa: enquanto este estudo concentrou-se na predição da vazão de *uplink* em ambiente real, com coletas realizadas em diferentes dias e horários, a maioria dos trabalhos correlatos limitou-se a ambientes controlados ou a experimentos com condições mais restritas, como o uso de um único cenário, dispositivo ou configuração de rede. O método experimental adotado utilizou uma rota fixa, percorrida em horários distintos ao longo do período de coleta.

Para a predição da vazão de *uplink*, foram utilizados os algoritmos floresta aleatória, regressão linear, k vizinhos mais próximos, máquina de vetores de suporte e rede neural. A avaliação de desempenho foi realizada por meio de métricas como o erro médio absoluto (MAE), o erro quadrático médio (MSE) e a raiz do erro quadrático médio (RMSE), com o objetivo de proporcionar uma análise comparativa abrangente e servir de base para pesquisas futuras.

Foi possível identificar, entre os trabalhos analisados, práticas recorrentes como a condução de experimentos em ambientes laboratoriais, os quais, embora controlados, não refletem fielmente as condições reais de uso. Por outro lado, os estudos realizados em ambientes reais enfrentaram desafios na coleta de dados, estando sujeitos a interferências externas como condições climáticas. Ademais, muitos desses estudos utilizaram um número reduzido de parâmetros de rede, o que pode ter limitado a precisão das predições. A inclusão de parâmetros adicionais foi considerada uma oportunidade para aprimorar a acurácia dos modelos de predição de vazão.

Capítulo 4

Método Proposto

*Este capítulo apresenta o método proposto, detalhando as etapas conduzidas para a predição da vazão de *uplink*, desde o processo de coleta dos dados até a aplicação dos modelos de aprendizado de máquina.*

4.1 Visão Geral do Método Proposto

A visão geral apresenta as etapas necessárias para prever a vazão de *uplink*, conforme ilustrado na [Figura 4.1](#). O método proposto é composto por três etapas principais. Inicialmente, realiza-se a coleta de dados da rede 4G LTE, etapa que constitui o alicerce para os procedimentos subsequentes. A relevância e a qualidade dos dados coletados são cruciais para o sucesso da abordagem.

A segunda etapa corresponde ao pré-processamento dos dados, etapa fundamental para o uso eficaz de técnicas de aprendizado de máquina. Durante esse processo, os dados são preparados para análises mais refinadas. Uma análise exploratória é conduzida para verificar a adequação da base à aplicação de algoritmos de aprendizado de máquina. Uma vez verificada a conformidade da base, procede-se ao seu tratamento. Caso contrário, é necessário iniciar um novo experimento de coleta de dados.

Durante a etapa de tratamento da base de dados, são identificadas e corrigidas inconsistências e dados defeituosos. Posteriormente, procede-se à normalização dos dados e à seleção das variáveis mais relevantes para a análise. Essa fase é fundamental para assegurar a

integridade, a qualidade e a representatividade dos dados no processo subsequente utilizados no treinamento do modelo.

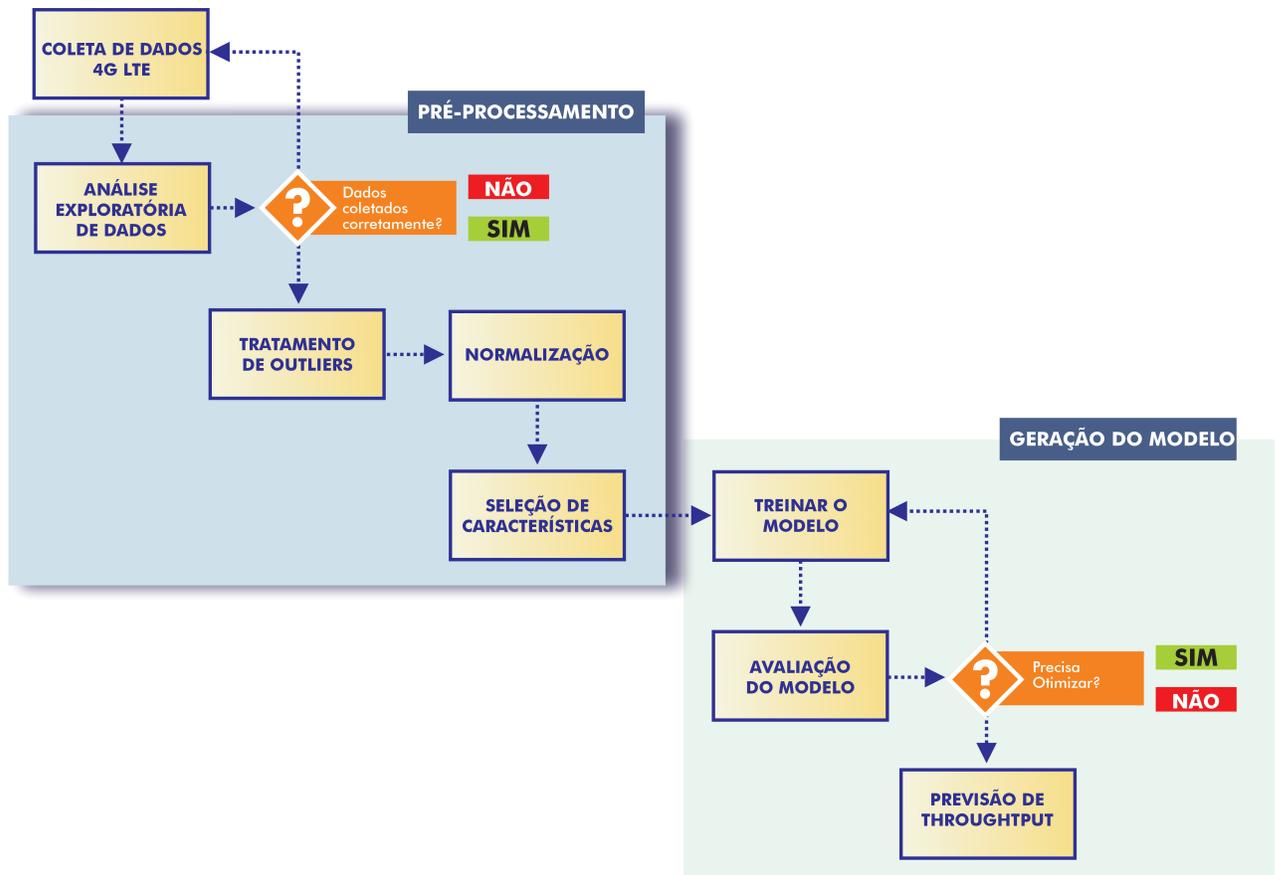


Figura 4.1: visão geral do método proposto
Fonte: elaboração própria (2025).

Finalmente, na terceira etapa, o modelo é treinado com os algoritmos disponíveis e, em seguida, avaliado para determinar se os resultados são satisfatórios. Caso os resultados não atinjam o padrão desejado, o modelo deve ser treinado novamente até atingir um bom desempenho. Ao concluir esses estágios, é possível fazer previsões da vazão de *uplink*.

4.1.1 Coleta de Dados

A coleta de dados representa uma etapa essencial na aplicação de algoritmos de aprendizado de máquina para previsão da vazão de *uplink* em redes móveis 4G LTE. Para garantir a representatividade e a qualidade dos dados, é fundamental adotar um método de coleta adequado e considerar diversos fatores que influenciam a conectividade, tais como a localização

geográfica, o tipo de dispositivo utilizado, o horário das medições, o volume da amostra, a velocidade da conexão e o tipo de dados registrados.

A qualidade da rede 4G LTE pode variar significativamente de uma região para outra, e até mesmo de um local para outro dentro da mesma região. Portanto, é importante colher dados de uma variedade de visões geográficas para obter uma abrangência da qualidade da rede.

O desempenho da rede 4G LTE pode ser afetado pelo tipo de dispositivo móvel usado para acessá-la. Por exemplo, dispositivos mais antigos podem ter dificuldade em se conectar a redes 4G LTE, enquanto dispositivos mais recentes podem suportar tecnologias mais avançadas de rede. É importante, portanto, coletar dados usando vários tipos de dispositivos móveis para ter uma compreensão mais precisa do desempenho da rede.

A qualidade da rede 4G LTE pode variar dependendo do horário do dia em que é acessada. Por exemplo, em horários de pico, quando há muitos usuários simultâneos na rede, a qualidade da conexão pode não ser garantida. Portanto, é importante coletar dados em diferentes horários do dia para avaliar a qualidade da rede em diferentes condições.

O tamanho da amostra desempenha um papel essencial na predição da vazão, uma vez que amostras reduzidas podem resultar em estimativas imprecisas ou enviesadas em relação à vazão real. Quando a amostragem é limitada, há risco de não refletir adequadamente a população de interesse, comprometendo a qualidade das predições. Em contrapartida, uma base de dados mais ampla tende a produzir resultados mais consistentes e confiáveis, reduzindo a probabilidade de variações não representativas. Dessa forma, garantir uma amostra suficientemente grande e bem distribuída é fundamental para assegurar a robustez e a precisão dos modelos preditivos.

A velocidade na coleta de dados pode afetar a qualidade dos dados coletados e, conseqüentemente, o desempenho dos modelos de aprendizado de máquina treinados com esses dados. Se a coleta de dados for muito rápida, pode haver perda de dados importantes ou ruído nos dados coletados, o que pode resultar em modelos de aprendizado de máquina imprecisos.

Os tipos dados coletados podem incluir informações sobre data, horário, e outras métricas relevantes para a qualidade da rede como RSRP, RSRQ, SNR, *uplink* e *dowlink*.

É importante coletar uma ampla variedade de dados para obter uma visão completa do desempenho da rede e garantir que os modelos de aprendizagem de máquina sejam treinados com informações precisas e monitoradas. Portanto, a coleta de dados de rede 4G LTE requer a consideração cuidadosa de vários fatores para garantir que os dados sejam precisos e representativos.

4.1.2 Pré-processamento

O pré-processamento da base de dados representa uma fase fundamental para a predição da vazão de *uplink* da rede 4G LTE. Seu principal objetivo é preparar os dados, garantindo que eles sejam precisos, relevantes e adequados para a geração do modelo de predição. Essa etapa contribui significativamente para a mitigação de problemas presentes na base de dados por meio de tratamento de *outliers*¹, substituição de valores ausentes, normalização dos dados, e principalmente, a seleção de variáveis.

A. Análise Exploratória de Dados

Conhecer a base de dados é fundamental no processo da predição da vazão de *uplink*. Isso porque a qualidade dos resultados obtidos depende diretamente da qualidade dos dados de entrada utilizados. É importante entender as características e os padrões dos dados em questão, a fim de identificar e corrigir eventuais erros e inconsistências tais como valores faltantes, dados duplicados, erros de entrada de dados, *outliers*, entre outros.

B. Tratamento de *Outliers*

Outliers são valores que estão muito distantes da maioria dos outros valores em um conjunto de dados. Esses valores podem ser causados por erros na coleta de dados, erros de entrada, ou podem indicar algo significativo sobre o fenômeno que está sendo medido. Em muitos casos, esses valores extremos podem ser problemáticos para a análise, pois podem distorcer os resultados e prejudicar a precisão e a confiabilidade do modelo. Antes de tratá-los, é importante identificar os *outliers* em um conjunto de dados. Isso pode ser feito visualmente, usando gráficos, ou estatisticamente, usando técnicas de análise de desvio e análise de distribuição de frequência. Se os *outliers* são causados por erros de entrada ou coleta de

¹ *Outliers* são valores atípicos ou excepcionais em um conjunto de dados que se diferenciam significativamente dos demais valores. Esses valores estão fora do padrão esperado e podem distorcer a análise estatística ou a interpretação dos resultados.

dados, a remoção desses valores pode ser uma opção. No entanto, é importante lembrar que a remoção de valores extremos pode afetar a distribuição e a precisão dos dados.

No tratamento de dados ausentes, pode-se utilizar a interpolação linear, que é uma técnica utilizada para estimar valores intermediários entre pontos conhecidos em uma função. Essa técnica é baseada na suposição de que a função varia de forma linear entre os pontos conhecidos. É comumente usada em análise de dados, processamento de imagens, previsão do tempo e modelagem matemática, entre outras áreas.

A interpolação linear é relativamente simples e envolve apenas a conexão dos pontos conhecidos por segmentos de reta. Por exemplo, se tivermos dois pontos conhecidos em um gráfico (x_1, y_1) e (x_2, y_2) , podemos encontrar o valor de y em um ponto intermediário x usando a equação da reta que conecta esses dois pontos:

$$y = y_1 + \frac{y_2 - y_1}{x_2 - x_1} \cdot (x - x_1) \quad (4.1)$$

Essa equação nos permite calcular o valor de y para qualquer valor de x dentro do intervalo entre os dois pontos conhecidos.

C. Normalização

A normalização da base de dados é uma técnica usada para transformar os dados em uma escala comum e predefinida, a fim de remover qualquer viés ou influência que possa ter no modelo de aprendizado de máquina que está sendo criado. É uma etapa importante no pré-processamento de dados e pode melhorar a precisão e a confiabilidade do modelo.

A normalização Min-Max é uma técnica amplamente utilizada para transformar valores de uma variável em um intervalo específico, normalmente entre 0 e 1. Para aplicar essa técnica, subtrai-se o valor mínimo da variável de todos os valores e, em seguida, divide-se o resultado pelo intervalo, que é a diferença entre o valor máximo e o valor mínimo da variável. A fórmula para a normalização Min-Max é a seguinte:

$$X_{\text{norm}} = \frac{X - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}}$$

Onde:

X representa o valor original da variável,

X_{\min} é o valor mínimo da variável,

X_{\max} é o valor máximo da variável, e

X_{norm} é o valor normalizado resultante, que estará no intervalo entre 0 e 1.

D. Seleção de Características

A seleção de variáveis, também conhecida como "*feature selection*" em inglês, é um processo importante na área de aprendizado de máquina. É uma técnica que consiste em escolher as características mais relevantes ou informativas em um conjunto de dados, que serão usadas para construir um modelo de aprendizado de máquina. A ideia por trás da seleção de variáveis é identificar aquelas que têm maior impacto na saída do modelo, reduzindo o número de variáveis no final e, assim, melhorando a precisão e a eficiência do modelo.

Esta etapa é importante porque ajuda a evitar a inclusão de variáveis irrelevantes ou redundantes no modelo, o que pode levar a um modelo superajustado ou com baixa precisão. Além disso, a seleção de variáveis pode ajudar a reduzir o tempo e o custo de construção do modelo. Existem diversas técnicas para seleção de características relevantes, dentre as quais se destacam o teste Qui-Quadrado (*Chi-Square Test*), a informação mútua (*Mutual Information*), e as análises de correlação de Spearman e Pearson. No escopo deste trabalho, destaca-se a utilização da correlação de Spearman, por se tratar de uma abordagem eficaz na identificação de relações monotônicas entre variáveis contínuas e a variável-alvo de vazão.

Embora a correlação de Pearson seja uma técnica amplamente reconhecida para medir relações lineares entre variáveis contínuas, neste trabalho opta-se por utilizar exclusivamente a correlação de Spearman. Esta abordagem é mais adequada para a identificação de relações monotônicas, crescentes ou decrescentes, entre as variáveis preditoras e a variável-alvo de vazão, independentemente de linearidade. Ao utilizar os postos (*ranks*) dos dados em vez dos seus valores brutos, a correlação de Spearman oferece maior robustez frente à presença de (*outliers*) e à ausência de distribuição normal, características comuns em conjuntos

de dados provenientes de redes móveis.

4.1.3 Geração do Modelo

Nesta etapa da geração do modelo, é realizado um processo trifásico: treinamento, avaliação e obtenção da predição da vazão de *uplink*. Na primeira fase, o treinamento alimenta o algoritmo de aprendizado de máquina com a base de dados coletada de rede de dados móveis, permitindo que o modelo aprenda a relação entre as variáveis de entrada e saída. Em seguida, tem-se a avaliação do modelo; nesta etapa, é testado seu desempenho e capacidade de generalização em um conjunto de teste independente. A predição é comparada com os valores reais, com o auxílio de métricas de avaliação. O último passo é a obtenção da predição final; aplica-se o modelo treinado e avaliado em novos dados.

A. Treinar o Modelo

Para treinar o modelo de aprendizagem de máquina, é necessário dividir a base de dados em um conjunto de treinamento e outro de teste. O de treinamento é usado para ajustar os parâmetros do modelo, enquanto o de teste é usado para avaliar a precisão.

É comum recomendar que cerca de 70-80% dos dados sejam usados para formar o conjunto de treinamento, enquanto o restante, ou seja, cerca de 30-20% dos dados, deve ser destinado ao conjunto de testes. No método proposto, será utilizado 70% dos dados para o treinamento e 30% para o teste. Para treinar o modelo, é preciso escolher um algoritmo adequado e ajustar os parâmetros para minimizar o erro na predição. O aprendizado de máquina supervisionado usando regressão é uma técnica que pode ser utilizada para predição de valores contínuos, como a vazão de *uplink* da rede 4G LTE de telefonia móvel. Dentre alguns algoritmos de regressão estão: RF, LR, SVM, KNN e NN.

B. Avaliação do Modelo

Após treinar o modelo de aprendizado de máquina supervisionado, usando regressão para predição da vazão de *uplink*, é importante avaliar o desempenho do modelo para verificar sua capacidade de fazer predições precisas em dados não vistos durante o treinamento.

A avaliação do modelo pode ser realizada usando um conjunto de dados de teste separados, que não foram usados durante o treinamento. O conjunto de teste deve incluir exemplos de entradas e saídas correspondentes, da mesma forma que o conjunto de treina-

mento.

Uma das métricas que podem ser usadas para a avaliação é o erro médio absoluto (MAE) ou o erro médio quadrático (MSE), que medem a diferença entre os valores previstos e reais. O MAE é a média da diferença absoluta entre as predições do modelo e os valores reais, enquanto o MSE é a média dos erros quadráticos. Ambos os valores são medidas de erro, e quanto menor, melhor será o desempenho do modelo.

O coeficiente de determinação, também conhecido como coeficiente de determinação R^2 , é outra métrica estatística comumente usada para avaliar a qualidade do modelo de regressão supervisionado. Ele quantifica a proporção da variância na variável dependente que é previsível a partir da(s) variável(is) independente(s). Seu valor varia entre 0 e 1, onde um R^2 de 1 indica que o modelo é capaz de prever perfeitamente a variável dependente usando a variável independente, enquanto um R^2 de 0 sugere que o modelo não consegue explicar qualquer variação na variável dependente.

Após a avaliação inicial do modelo, pode-se realizar ajustes nos parâmetros e proceder com um novo treinamento, se julgado necessário. Quando o modelo atinge um nível adequado de desempenho, está pronto para ser implementado e poderá realizar predições sobre conjuntos de dados inéditos.

C. Predição da Vazão

Após a geração do modelo de predição da vazão de *uplink*, o mesmo pode ser utilizado para realizar previsões com base em novos dados. Para isso, é necessário passar as entradas relevantes do novo conjunto de dados para o modelo, o qual retornará a vazão do *uplink* correspondente. A aplicação prática do modelo pode incluir, por exemplo, a otimização do desempenho da rede de celular móvel 4G LTE. Com base nas predições do modelo, é possível identificar as áreas com menor vazão de *uplink* e tomar medidas para melhorar o desempenho dessas áreas, como a instalação de novas torres de celular, ajuste de antenas ou realocação de recursos de rede.

Além disso, o modelo também pode ser utilizado para fazer predição em tempo real para ajustar a alocação de recursos de rede, garantindo que a capacidade de rede seja melhorada em relação à velocidade, segurança e disponibilidade para atender às necessidades dos usuários.

Portanto, o modelo de predição da vazão de *uplink* pode ser uma ferramenta valiosa para ajudar as operadoras de telefonia móvel a melhorar o desempenho da rede e aprimorar a experiência do usuário, o que pode trazer benefícios significativos tanto para as empresas quanto para os consumidores.

4.2 Resumo

Neste capítulo foram apresentadas as etapas necessárias para predizer a vazão de *uplink* da rede de telefonia móvel 4G LTE. Na seção 4.1 foi apresentada a visão geral do método adotado, com a descrição dos processos:

- Coleta de dados: descreve qual foi o método utilizado para fazer a coleta de dados, localização geográfica, o tipo de dispositivo móvel, horário do dia, tamanho da amostra, velocidade, tipo de dados coletados. Todos esses dados foram descritos para melhor compreensão dos dados e para ajudar em trabalhos futuros;
- Pré-processamento: essa fase tem como objetivo preparar os dados de modo a garantir sua precisão, relevância e adequação ao modelo preditivo. É uma etapa essencial, dividida em quatro fases: conhecimento da base, tratamento de *Outliers*, normalização e seleção das variáveis;
- Geração do modelo: tem como foco treinar um modelo de aprendizagem de máquina, seguido de sua avaliação por meio de métricas como MAE, MSE, R^2 entre outros. Após validá-lo, o modelo é aplicado na prática para realizar predições em novos conjuntos de dados.

Capítulo 5

Resultados e Discussões

Este capítulo descreve os experimentos realizados para a aplicação do modelo de predição da vazão de uplink em redes de telefonia móvel. O objetivo dos experimentos foi verificar a eficácia do modelo e sua capacidade de predizer a vazão em ambientes reais de rede móvel. Para isso, o capítulo apresenta o protocolo experimental adotado, incluindo a descrição da base de dados, bem como as tecnologias e metodologias utilizadas para a coleta e a análise dos dados.

5.1 Protocolo de Experimento

Nesta seção, apresenta-se o protocolo do experimento, no qual são descritas detalhadamente as diferentes etapas do processo experimental, desde as tecnologias empregadas até a metodologia de coleta de dados utilizada para a análise dos resultados obtidos a partir da base de dados. Com essa abordagem, busca-se oferecer uma visão abrangente do experimento, permitindo a plena compreensão da aplicação do método proposto.

Os experimentos realizados e descritos ao longo deste capítulo buscam preencher uma lacuna identificada nos estudos avaliados, fornecendo uma base sólida para futuras investigações na área. A análise dos resultados, em conjunto com a execução estruturada do experimento, permitiu avaliar a eficácia do modelo de predição proposto, bem como sua capacidade de estimar com precisão a vazão em ambientes de rede móvel. Dessa forma, este trabalho contribui efetivamente para o avanço do conhecimento técnico-científico no campo de otimização de redes móveis.

5.1.1 Tecnologias Utilizadas

Na execução do experimento, a [Tabela 5.1](#) apresenta as ferramentas utilizadas, destacando o smartphone, o aplicativo de coleta, o meio de transporte e o chip da operadora de rede para a obtenção dos dados. Para a análise dos dados, foram utilizados o ambiente Google Colab, a linguagem Python e as bibliotecas Pandas, NumPy, Matplotlib, Seaborn e Scikit-learn, além de um computador para a execução dos procedimentos.

Tabela 5.1: ferramentas e recursos utilizados na aplicação do método proposto

Ferramentas	Objetivo
Celular Motorola	Coletar os dados de rede
Aplicativo Armadeira	Realizar as medições dos parâmetros de rede em sistema Android
Motocicleta	Realizar a locomoção para a coleta
Chip da operadora Vivo	Garantir a conexão com as torres e o acesso aos dados
Python	Linguagem de programação utilizada para a execução do método
Google Colab	Ambiente de execução para os códigos em Python
Notebook Dell	Utilizado para realizar as atividades de programação
Pandas	Biblioteca para manipulação e análise de dados
NumPy	Biblioteca utilizada para realizar uma ampla variedade de operações matemáticas
Matplotlib e Seaborn	Bibliotecas destinadas à visualização de dados
Scikit-learn	Biblioteca para implementação de algoritmos de aprendizado de máquina

Fonte: elaboração própria (2025).

Foi utilizado um smartphone Motorola modelo Moto G10, equipado com um chip da operadora Vivo para a conexão à internet, juntamente com o aplicativo Armadeira, desenvolvido para a coleta de parâmetros de rede. A escolha do dispositivo foi fundamentada na necessidade de mobilidade, flexibilidade na configuração de coleta e capacidade de obtenção de informações em tempo real. A operadora Vivo, por sua vez, foi selecionada em função da boa cobertura de rede na área estudada, fator essencial para garantir a qualidade e a confiabilidade dos dados. Para a locomoção durante a coleta, foi empregada uma motocicleta, possibilitando deslocamentos ágeis entre os diferentes pontos de amostragem.

Para a aplicação do modelo de aprendizado de máquina no experimento, optou-se por

utilizar a linguagem de programação Python, que oferece diversas bibliotecas e ferramentas voltadas para a análise de dados e a implementação de algoritmos de aprendizado de máquina. O Google Colab foi utilizado para facilitar o acesso e a execução do código, permitindo a execução em uma infraestrutura poderosa e escalável. A execução do código foi realizada em um notebook da marca Dell, o qual forneceu os recursos computacionais necessários para a condução dos experimentos de forma eficiente e adequada.

No experimento, utilizaram-se diversas tecnologias para análise e visualização de dados. Pandas foi empregada para manipulação de dados, NumPy para operações matemáticas, Matplotlib e Seaborn para geração de gráficos, e Scikit-learn para preparação e aplicação de algoritmos de aprendizado de máquina.

A combinação dessas ferramentas permite a aplicação do modelo de aprendizado de máquina no experimento com eficiência e precisão, oferecendo resultados confiáveis e relevantes para a pesquisa em questão.

5.1.2 Aplicativo para Coleta de Dados

O Armadeira é um aplicativo para smartphone desenvolvido pelo projeto "Técnicas de Inteligência Artificial para Análise e Otimização de Desempenho de Software" (SW-PerfI), especificamente pela equipe de inteligência de hardware, em parceria entre a Universidade Federal do Amazonas, Instituto de Pesquisa Eldorado e Motorola. O projeto de pesquisa, desenvolvimento e inovação visa estudar e investigar possíveis melhorias no desempenho em dispositivos móveis. Como parte deste esforço, o aplicativo Armadeira é projetado para coletar dados de parâmetros de rede em smartphones Android.

As [Figura 5.1](#) e [Figura 5.2](#) mostram as duas telas principais do aplicativo de coleta de dados de rede. O Armadeira é projetado para coletar informações sobre a rede Wi-Fi e dados móveis de um smartphone Android em uma granularidade de um segundo. O aplicativo consegue coletar até 38 parâmetros, incluindo informações sobre a qualidade do sinal, tipo de rede (4G, 5G, Wi-Fi), operadora e muito mais.

As [Figura 5.3](#) e [Figura 5.4](#) mostram o armazenamento e uma parte do arquivo CSV. Os dados coletados pelo aplicativo são salvos em um arquivo CSV. A escolha desse formato foi pensada para permitir a realização de análises futuras de desempenho da rede com facilidade

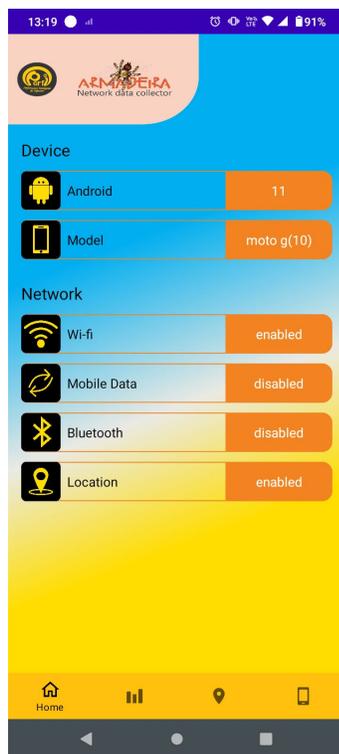


Figura 5.1: tela home do aplicativo
Fonte: elaboração própria (2025).



Figura 5.2: tela de dados do aplicativo
Fonte: elaboração própria (2025).

e eficiência, uma vez que o CSV é um formato simples e flexível que pode ser importado em diversas ferramentas de análise de dados.

Para proteger a privacidade dos usuários, o Armadeira na versão (armadeira-1.01.009) armazena os dados coletados no próprio dispositivo do usuário. Isso garante que os dados não sejam compartilhados com terceiros e que os usuários tenham total controle sobre suas informações. Com esses dados, é possível fornecer análises valiosas sobre o desempenho da rede, identificar possíveis gargalos e ajudar os usuários a tomarem decisões relacionadas à sua conexão à Internet. O aplicativo permanece em desenvolvimento e pode ser aprimorado com novas funcionalidades.

5.1.3 Fluxo de Coleta de Dados

O processo de coleta de dados de rede de telefonia móvel adotado é composto por um fluxo que envolve o envio de um arquivo de 1,6 Gigabyte para a nuvem. Esse fluxo é apresentado na [Figura 5.5](#). A primeira etapa desse fluxo consiste em ligar a conexão de dados móveis

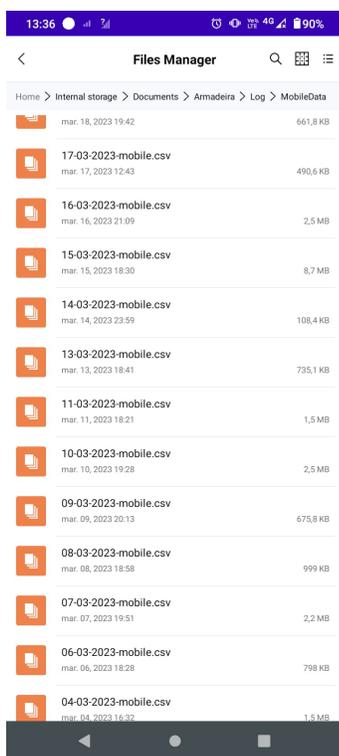


Figura 5.3: armazenamento dos dados
Fonte: elaboração própria (2025).

No.	date	time	locationStatus
1	15/03/2023	00:00:01	1
2	15/03/2023	00:00:02	1
3	15/03/2023	00:00:03	1
4	15/03/2023	00:00:04	1
5	15/03/2023	00:00:05	1
6	15/03/2023	00:00:06	1
7	15/03/2023	00:00:07	1
8	15/03/2023	00:00:08	1
9	15/03/2023	00:00:09	1
10	15/03/2023	00:00:10	1
11	15/03/2023	00:00:11	1
12	15/03/2023	00:00:12	1
13	15/03/2023	00:00:13	1
14	15/03/2023	00:00:14	1

Figura 5.4: arquivo.csv dos dados
Fonte: elaboração própria (2025).

e enviar o arquivo. Em seguida, é iniciado o aplicativo de coleta, responsável por capturar e armazenar informações relevantes sobre a rede de telefonia móvel em que o usuário está conectado. Uma vez que o aplicativo esteja em execução, o usuário pode iniciar a coleta de dados.

Durante a etapa de coleta de dados, o sistema realizou o monitoramento periódico do envio de arquivos à nuvem. Ao identificar a conclusão do envio, uma solicitação era apresentada ao usuário para confirmação do término da operação. Caso a resposta fosse negativa, o monitoramento era mantido; em caso afirmativo, o sistema questionava se o usuário desejava prosseguir com o experimento. Se a resposta fosse positiva, um novo envio era iniciado, reiniciando o ciclo. Caso contrário, o aplicativo de coleta era suspenso e o processo finalizado. Esse fluxo foi cuidadosamente projetado visando garantir uma coleta de dados eficiente, precisa e confiável, promovendo a consistência dos dados e assegurando a integridade dos resultados.

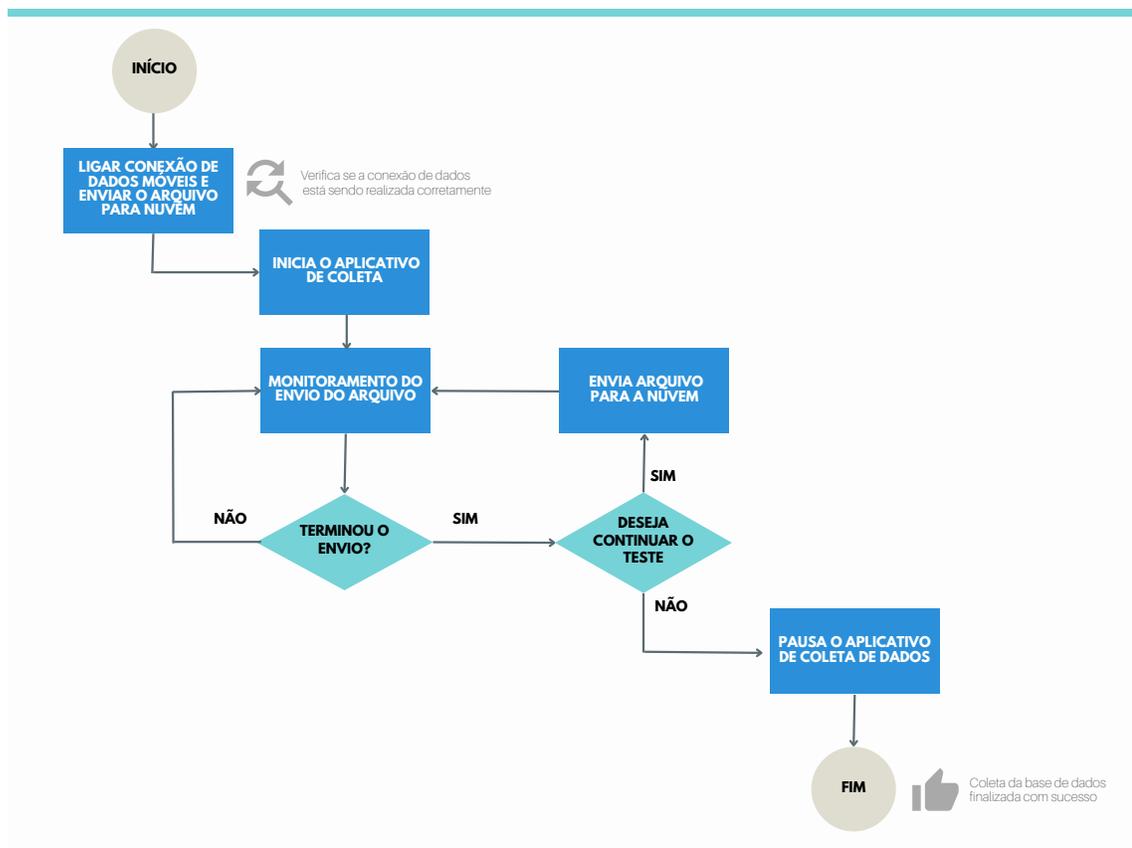


Figura 5.5: fluxo de coleta de dados
Fonte: elaboração própria (2025).

5.2 Execução do Experimento

As coletas da base de dados foram realizadas do dia 03 a 24 de março de 2023, na cidade de Itacoatiara, no estado do Amazonas, no país Brasil. A [Figura 5.6](#) mostra a rota utilizada para a coleta da base de dados, a foto foi retirada do aplicativo Google Maps. É possível notar que essa rota possui 6,5 quilômetros. O veículo automotor foi usado a uma velocidade média de 40 quilômetros/hora.

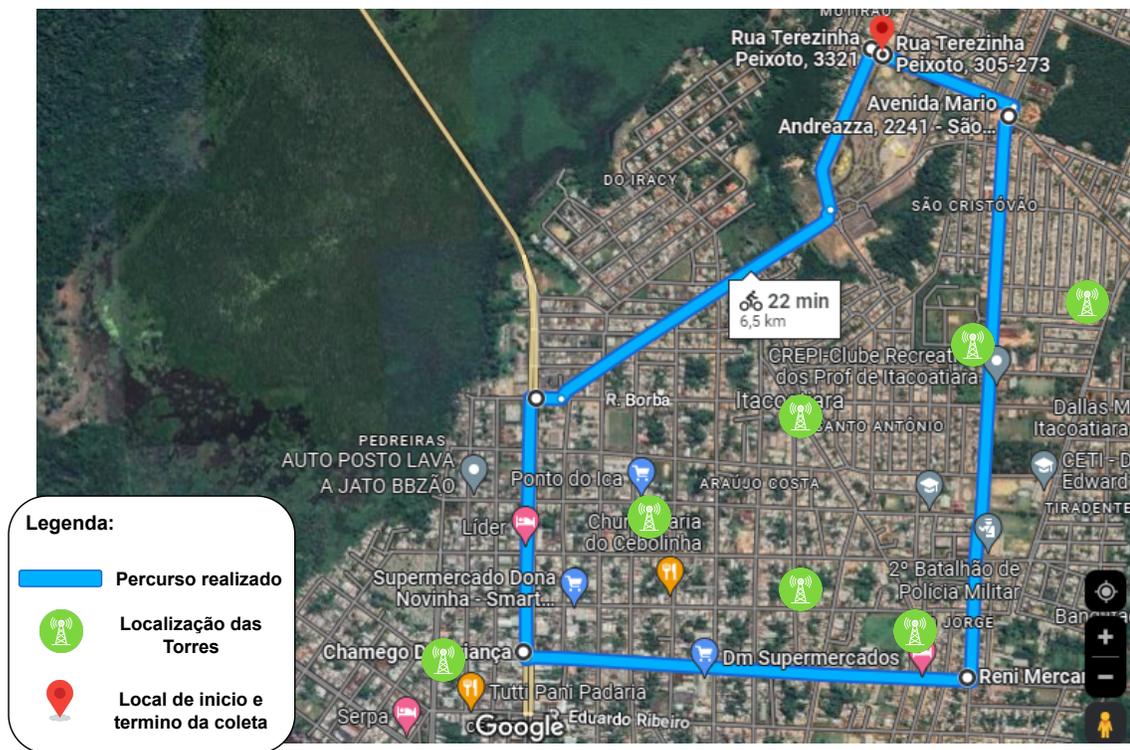


Figura 5.6: rota da coleta de dados da cidade
 Fonte: elaboração própria (2025).

5.2.1 Base de Dados

A [Tabela 5.2](#) apresenta os dados obtidos por meio do aplicativo desenvolvido, os quais englobam uma variedade de informações relevantes para a análise. Entre os dados coletados, destacam-se o identificador do dispositivo, a data e o horário em que cada coleta foi realizada, permitindo o rastreamento temporal das amostras. Além desses, foram obtidas informações geográficas, especificamente as coordenadas de latitude e longitude, possibilitando a análise espacial dos dados. Adicionalmente, foram coletados parâmetros relacionados à conectividade móvel, incluindo o estado do dispositivo (ativo ou desconectado no momento da coleta) e o tipo de rede móvel à qual estava conectado, como 3G, 4G, entre outras tecnologias.

Além disso, foram coletados parâmetros essenciais para caracterizar a qualidade da rede móvel, como RSRP, RSRQ, RSSI, RSSNR, além das taxas de *downlink* e *uplink*. Esses dados são fundamentais para avaliar o desempenho da rede em diferentes regiões e identificar possíveis falhas de cobertura ou conexão. A lista completa das informações coletadas pode ser consultada no Apêndice [Apêndice A](#).

Tabela 5.2: parâmetros coletados pelo aplicativo durante o experimento

Nome da Feature	Tipo	Unidade	Descrição
mobileStatus	boolean	–	(1) indica que o dispositivo móvel está ligado; (0) indica que está desligado.
netType	string	–	Parâmetro que indica o tipo de rede móvel atualmente em uso pelo dispositivo.
downstream Bandwidth	int	Kbps	Velocidade de download de dados na rede móvel, medida em kilobits por segundo (Kbps).
upstream Bandwidth	int	Kbps	Velocidade de upload de dados na rede móvel, medida em Kbps.
uplink	long	Bytes	Taxa de dados recebidos, em Bytes.
downlink	long	Bytes	Taxa de dados transmitidos, em Bytes.
networkOperator	string	–	Identifica a operadora de rede móvel à qual o dispositivo está conectado.
mcc	string	–	Código do país da rede móvel (MCC) — identificador único atribuído a um país para fins de rede móvel.
mnc	string	–	Código da rede móvel (MNC) — identificador único atribuído a uma rede móvel dentro de um país.
dbm	int	dBm	Potência do sinal recebido, em decibéis em relação a um miliwatt (dBm).
level	int	dBm	Indicador de intensidade do sinal recebido (RSSI), em dBm.
rsrp [dBm]	int	dBm	Potência do sinal de referência recebido (RSRP), em dBm.
rsrq	int	dB	Qualidade do sinal de referência recebido (RSRQ), em dB.
rsqi [dBm]	int	dBm	Indicador de intensidade do sinal recebido (RSSI), em dBm.
rssnr	int	dB	Relação sinal-ruído do sinal de referência (RSSNR), em dB.
band	int	–	Banda de frequência na qual o dispositivo móvel está operando.
frequency [MHz]	int	MHz	Frequência, em MHz, na qual o dispositivo móvel está operando na rede celular.

Fonte: elaboração própria (2025).

5.2.2 Análise Exploratória dos Dados

A próxima etapa tratou da análise exploratória dos dados. Inicialmente, utilizou-se o ambiente do Google Colab para importar o arquivo .csv e ter acesso aos detalhes da base de dados. A [Figura 5.7](#) apresenta as dez primeiras linhas desses dados.

	deviceId	data	time	locationStatus	latitude	longitude	mobileStatus	netType	RSRP	RSRQ	...	band
0	3d4c6129617d84c4	02/03/2023	18.02.38	1	-3.122741	-58.433811	1	4G	-93	-15	...	1
1	3d4c6129617d84c4	02/03/2023	18.02.38	1	-3.122715	-58.433814	1	4G	-93	-15	...	1
2	3d4c6129617d84c4	02/03/2023	18.02.55	1	-3.122723	-58.433919	1	4G	-100	-14	...	1
3	3d4c6129617d84c4	02/03/2023	18.02.57	1	-3.122669	-58.434053	1	4G	-95	-14	...	1
4	3d4c6129617d84c4	02/03/2023	18.02.59	1	-3.122605	-58.434201	1	4G	-91	-14	...	1
5	3d4c6129617d84c4	02/03/2023	18.03.01	1	-3.122563	-58.434316	1	4G	-88	-15	...	1
6	3d4c6129617d84c4	02/03/2023	18.03.03	1	-3.122516	-58.434424	1	4G	-95	-15	...	1
7	3d4c6129617d84c4	02/03/2023	18.03.06	1	-3.122494	-58.434518	1	4G	-89	-15	...	1
8	3d4c6129617d84c4	02/03/2023	18.03.16	1	-3.122508	-58.434637	1	4G	-93	-16	...	1
9	3d4c6129617d84c4	02/03/2023	18.03.19	1	-312.262	-58.434698	1	4G	-85	-16	...	1

10 rows x 29 columns

Figura 5.7: base de dados para análise exploratória dos dados
Fonte: elaboração própria (2025).

Essa base de dados contém 11.260 linhas e 29 colunas. Foi verificada também a quantidade de dados nulos na base. Foram identificados os parâmetros de latitude, longitude, RSRP, RSRQ, SNR, RSSI, *downlink* e *uplink*. Esses dados apresentavam uma quantidade pequena de dados nulos. A [Figura 5.8](#) apresenta a frequência dos dados coletados por data. É possível notar que os dados iniciam no dia 02.03.2023 e terminam no dia 24.03.2023. O dia com maior frequência foi 24.03.2023. Essas coletas foram feitas nos períodos da manhã, da tarde e da noite.

A [Figura 5.9](#) apresenta a quantidade dos dados divididos nos três períodos, evidenciando que o período da noite foi mais representativo. O período da noite tem início a partir das 18 horas até as 6 horas.

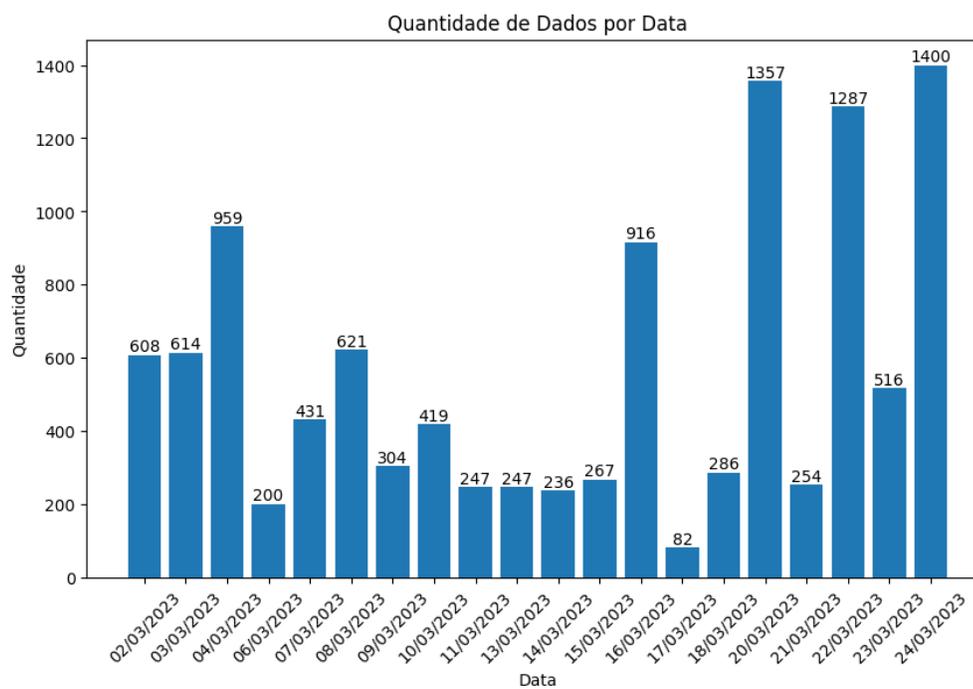


Figura 5.8: quantidade da coleta por data
Fonte: elaboração própria (2025).

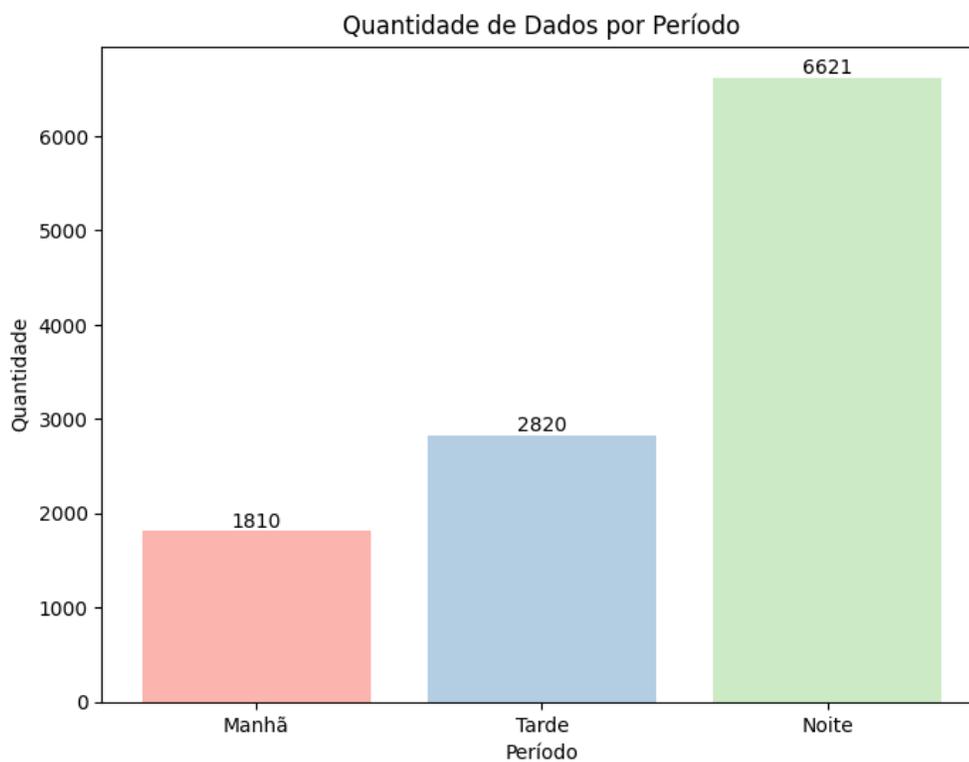
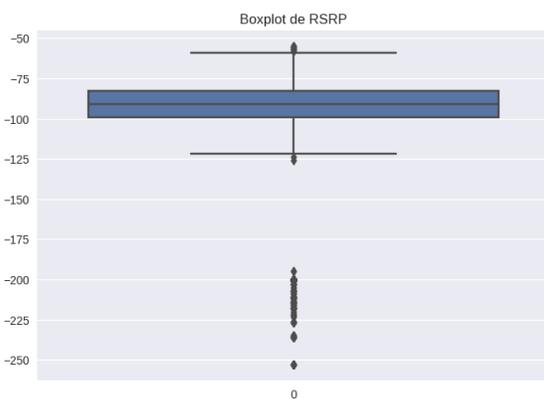


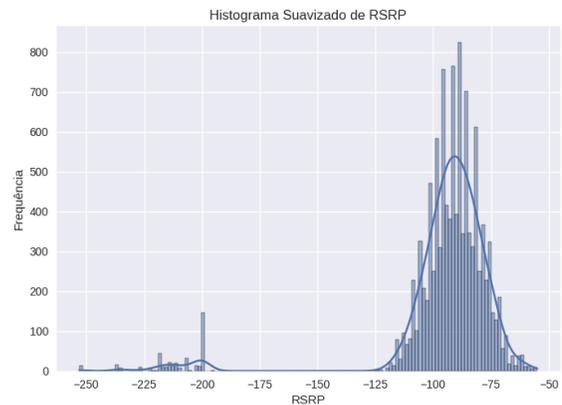
Figura 5.9: quantidade da coleta por período
Fonte: elaboração própria (2025).

A análise dos parâmetros de rede por meio de representações gráficas, como boxplots e histogramas, constitui uma ferramenta fundamental para a compreensão do comportamento dos dados e para a identificação de eventuais inconsistências, tais como valores atípicos (*outliers*) e variações acentuadas na distribuição. Essas características podem indicar problemas relevantes na infraestrutura de rede, como falhas de cobertura, perdas de sinal e degradação na qualidade da transmissão. A identificação e o tratamento dessas anomalias são imprescindíveis para assegurar a qualidade dos serviços e a confiabilidade dos modelos de predição.

Na [Figura 5.10](#), apresentam-se as representações gráficas do parâmetro RSRP, considerado uma das métricas mais importantes para a avaliação da qualidade do sinal em redes 4G LTE. O boxplot (a) permite visualizar a distribuição e possíveis assimetrias ou valores extremos, enquanto o histograma (b) evidencia a frequência de ocorrência dos diferentes níveis de intensidade do sinal. A correta interpretação dessas representações é essencial, uma vez que deslocamentos da mediana para valores mais baixos ou a presença de *outliers* podem sinalizar áreas com cobertura insuficiente, resultando em experiências negativas para o usuário, como quedas de chamadas ou lentidão na transmissão de dados.



(a) Boxplot do parâmetro de qualidade de rede RSRP



(b) Histograma do parâmetro de qualidade de rede RSRP

Figura 5.10: representações gráficas do parâmetro de qualidade de rede RSRP

Fonte: elaboração própria (2025).

Complementando a análise gráfica, a [Tabela 5.3](#) apresenta as estatísticas descritivas do RSRP, oferecendo uma visão quantitativa do comportamento deste parâmetro. A média registrada foi de -94.48 dBm, valor que caracteriza uma intensidade de sinal moderada, visto que quanto mais próximo de zero, melhor é a qualidade do sinal recebido.

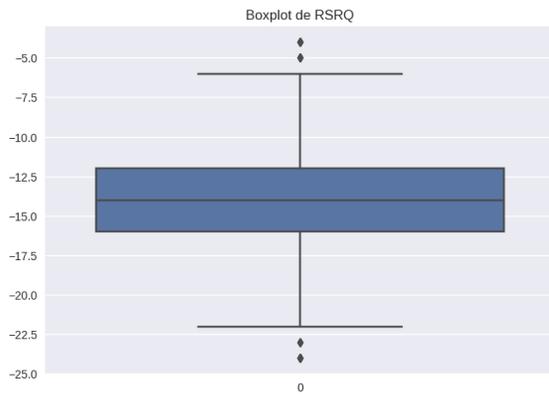
Tabela 5.3: estatísticas do RSRP

RSRP	Valor
Média	-94.48
Desvio Padrão	24.56
Valor Mínimo	-253.00
25%	-99.00
50%	-91.00
75%	-83.00
Valor Máximo	-55.00

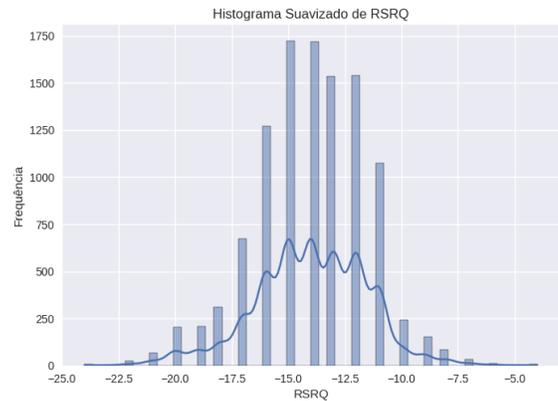
Fonte: elaboração própria (2025).

O desvio padrão, de 24.56 dBm, evidencia uma expressiva variabilidade nos valores observados, indicando heterogeneidade na qualidade do sinal ao longo das medições. O valor mínimo registrado, de -253.00 dBm, representa uma situação de extrema debilidade do sinal, enquanto o valor máximo, de -55.00 dBm, reflete uma condição de sinal relativamente forte. A análise dos percentis revela que 25% das medições encontram-se abaixo de -99.00 dBm, a mediana situa-se em -91.00 dBm, e 75% dos valores estão abaixo de -83.00 dBm. Esses dados indicam que a maioria das observações corresponde a níveis de sinal entre moderados e fracos, reforçando a necessidade de ações para a melhoria da cobertura em determinadas regiões.

A [Figura 5.11](#) apresenta as representações gráficas do parâmetro RSRQ, cuja função é avaliar a qualidade do sinal com base na relação entre a potência do sinal de referência e os níveis de interferência presentes na rede. Comparativamente ao RSRP, o RSRQ apresenta uma distribuição mais concentrada e com menor presença de *outliers*, indicando maior estabilidade e uniformidade na qualidade do sinal.



(a) Boxplot do parâmetro de qualidade de rede RSRQ



(b) Histograma do parâmetro de qualidade de rede RSRQ

Figura 5.11: representações gráficas do parâmetro de qualidade de rede RSRQ
 Fonte: elaboração própria (2025).

A [Tabela 5.4](#) exibe as estatísticas descritivas do RSRQ, que complementam a interpretação gráfica. A média observada foi de -14.00 dB, valor indicativo de uma qualidade moderada, uma vez que valores mais próximos de zero refletem melhor relação entre sinal e interferência.

Tabela 5.4: estatísticas do RSRQ

RSRQ	Valor
Desvio Padrão	2.49
Valor Mínimo	-24.00
25%	-16.00
50%	-14.00
75%	-12.00
Valor Máximo	-4.00

Fonte: elaboração própria (2025).

O desvio padrão de 2.49 dB indica uma dispersão relativamente baixa, demonstrando consistência nas medições. Os valores extremos variaram de -24.00 dB a -4.00 dB, sendo que a maior parte dos dados se concentrou entre -16.00 dB (percentil 25%) e -12.00 dB (percentil 75%), evidenciando uma distribuição estável e homogênea.

O parâmetro SNR, representado na [Figura 5.12](#), quantifica a relação entre a potência do sinal e o nível de ruído, sendo um dos principais indicadores da qualidade da comunicação em redes móveis. O boxplot (a) e o histograma (b) permitem analisar, respectivamente, a

distribuição e a frequência de ocorrência dos valores deste parâmetro.

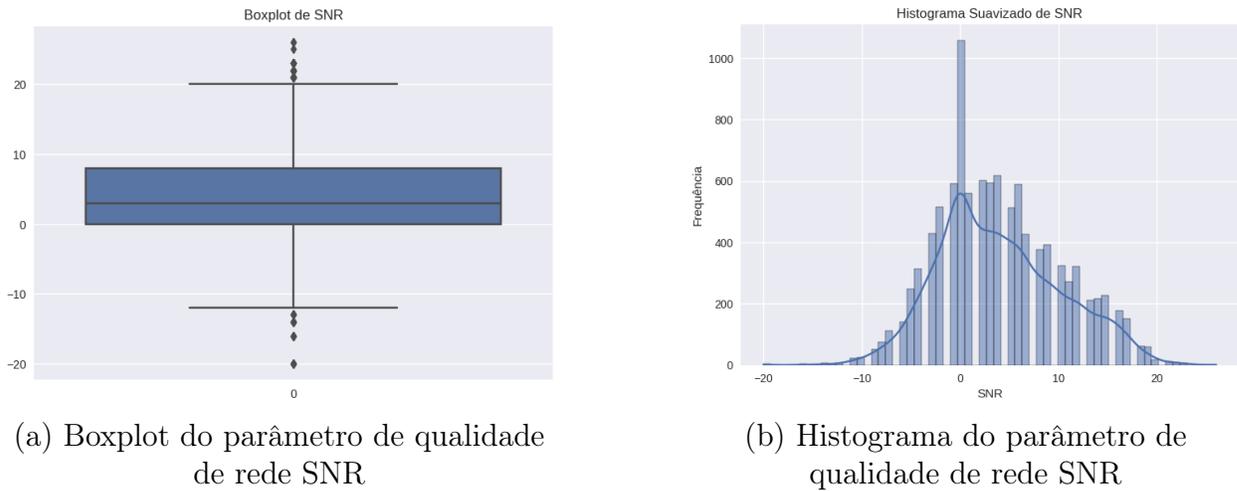


Figura 5.12: representações gráficas do parâmetro de qualidade de rede SNR
Fonte: elaboração própria (2025).

De acordo com a [Tabela 5.5](#), a média do SNR foi de 3.93 dB, enquanto o desvio padrão atingiu 6.31 dB, indicando uma dispersão moderada em torno da média.

Tabela 5.5: estatísticas do SNR

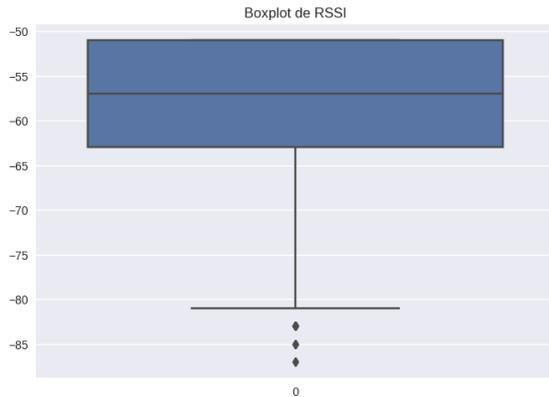
SNR	Valor
Média	3.93
Desvio Padrão	6.31
Valor Mínimo	-20.00
25%	0.00
50%	3.00
75%	8.00
Valor Máximo	26.00

Fonte: elaboração própria (2025).

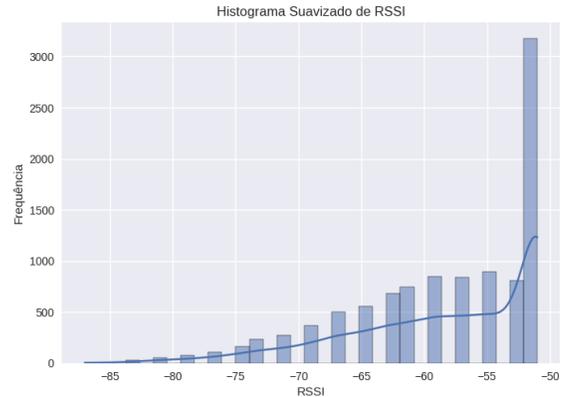
Os valores do SNR oscilaram entre -20.00 dB e 26.00 dB, sendo que níveis mais elevados refletem uma relação sinal-ruído mais favorável e, conseqüentemente, uma comunicação de maior qualidade. A distribuição dos percentis indica que 25% dos dados estão abaixo de 0.00 dB, a mediana corresponde a 3.00 dB e 75% das observações se situam abaixo de 8.00 dB, caracterizando uma predominância de condições de sinal de qualidade moderada.

Por fim, a [Figura 5.13](#) apresenta as representações gráficas do parâmetro RSSI, que mensura a intensidade do sinal recebido, sendo um elemento crucial na avaliação da qualidade

e da capacidade de recepção dos dispositivos móveis.



(a) Boxplot do parâmetro de qualidade de rede RSSI



(b) Histograma do parâmetro de qualidade de rede RSSI

Figura 5.13: representações gráficas do parâmetro de qualidade de rede RSSI

Fonte: elaboração própria (2025).

A [Tabela 5.6](#) sintetiza as estatísticas descritivas do RSSI, cuja média foi de -58.48 dBm, indicando uma intensidade moderada do sinal, considerando-se que valores mais próximos de zero refletem maior força de recepção.

Tabela 5.6: estatísticas do RSSI

RSSI	Valor
Média	-58.48
Desvio Padrão	7.58
Valor Mínimo	-87.00
25%	-63.00
50%	-57.00
75%	-51.00
Valor Máximo	-51.00

Fonte: elaboração própria (2025).

O desvio padrão, de 7.58 dBm, revela uma dispersão moderada em torno da média. As medições variaram de -87.00 dBm (intensidade mais fraca) a -51.00 dBm (mais forte). A análise dos percentis mostra que 25% dos valores estão abaixo de -63.00 dBm, a mediana corresponde a -57.00 dBm e 75% das observações encontram-se abaixo de -51.00 dBm, o que denota uma predominância de sinais de intensidade moderada.

Durante a verificação da rota é apresentada a plotagem dos pontos de latitude e

longitude, que podem ser verificados na [Figura 5.14](#). Essa figura representa visualmente a localização geográfica dos dados coletados no dia 24.03.2023. Através dessa plotagem é possível ter uma noção espacial da distribuição dos pontos ao longo da rota percorrida. A análise geoespacial é relevante para compreender a cobertura da rede e identificar possíveis áreas de interesse ou regiões com menor qualidade de sinal. Através da representação gráfica, será possível extrair *insights* adicionais que complementam a análise dos parâmetros de qualidade de rede. A plotagem das coordenadas de latitude e longitude proporciona uma visualização eficaz dos dados geográficos, facilitando a interpretação e o entendimento do contexto espacial dos registros coletados.

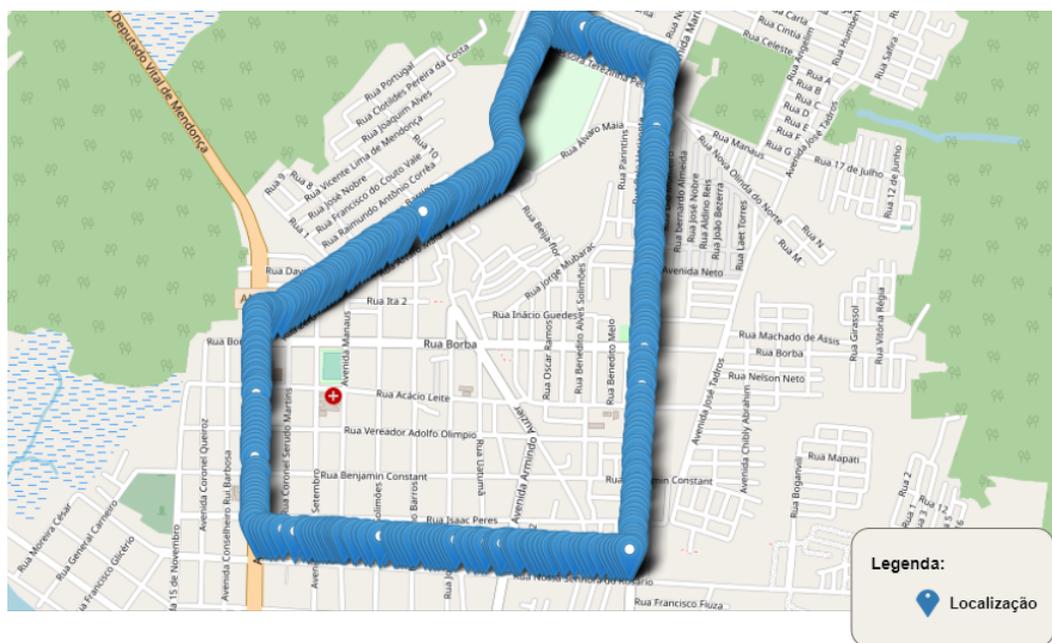


Figura 5.14: plotagem das coordenadas de latitude e longitude da base de dados coletada no dia 24.03.2023

Fonte: elaboração própria (2025).

A [Figura 5.15](#) apresenta a plotagem das coordenadas de localização no dia 24.03.2023 em relação ao parâmetro de RSRP. Essas coordenadas foram divididas em cinco categorias de sinal: Ruim, Médio, Bom, Muito Bom e Excelente. Esta imagem é de extrema importância para avaliar a distribuição e intensidade do sinal de RSRP em diferentes regiões.

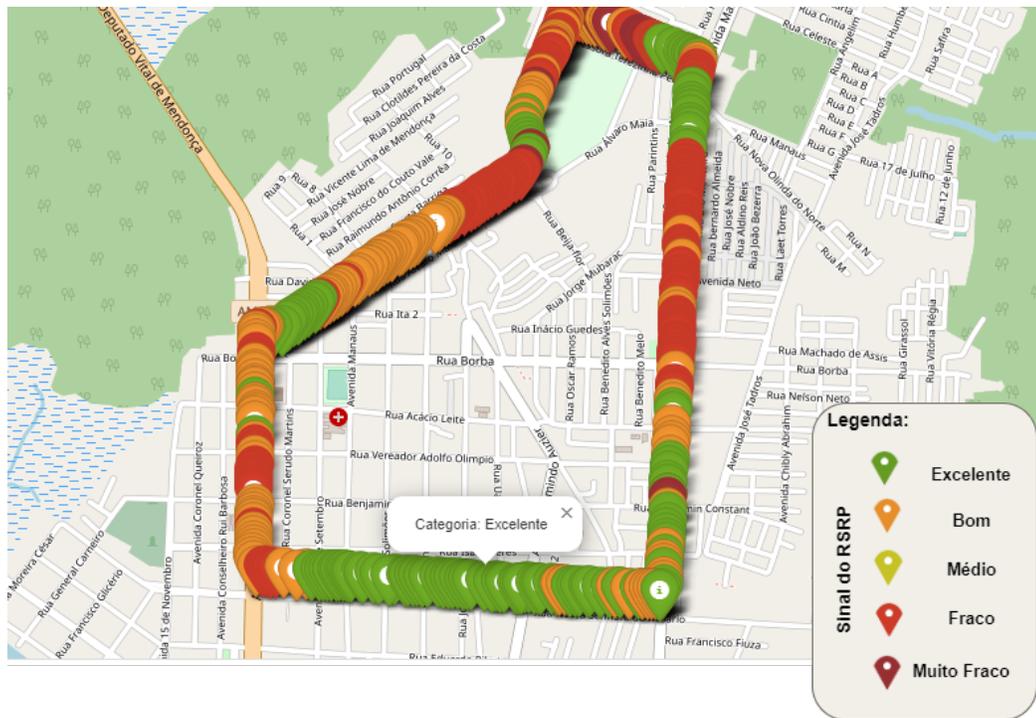


Figura 5.15: plotagem das coordenadas de localização com relação ao RSRP
 Fonte: elaboração própria (2025).

A análise exploratória dos dados revelou a presença de inconsistências associadas à localização geográfica e a parâmetros relacionados à qualidade da rede, aspectos frequentemente observados em bases de dados coletadas a partir de dispositivos móveis operando em ambientes reais. Adicionalmente, foram identificados registros nulos e valores ausentes, os quais, caso não tratados, poderiam comprometer a integridade das análises subsequentes.

Diante desse contexto, o pré-processamento dos dados desempenhou um papel fundamental na mitigação dessas falhas. As inconsistências e lacunas foram devidamente tratadas para assegurar a qualidade e a confiabilidade do conjunto de dados. Um dos principais focos dessa etapa consistiu na detecção e no tratamento de *outliers*, ou valores atípicos, que podem distorcer análises estatísticas e prejudicar o desempenho dos modelos preditivos.

A correção adequada dessas anomalias resultou na construção de uma base de dados robusta, o que favoreceu a obtenção de resultados com maior consistência estatística e tecnicamente mais válidos. Ressalta-se que a qualidade dos dados é um requisito imprescindível para garantir previsões consistentes e eficazes em aplicações baseadas em algoritmos de aprendizado de máquina, especialmente no contexto de redes móveis, onde a variabilidade e

a dinamicidade são características inerentes ao ambiente de coleta.

5.2.3 Limitação do Experimento

Abaixo seguem algumas limitações encontradas durante a execução do experimento de coleta de dados:

- Plano de dados: para medir a vazão de *uplink*, como apresentada na [Figura 5.5](#), foi necessário enviar continuamente dados para a nuvem durante a coleta (esse envio constante de dados faz parte do experimento). Como resultado, houve um consumo significativo de dados, o que pode ser considerado uma limitação, uma vez que exige um plano de dados adequado para a realização dos testes;
- Variação do horário: a variação do horário durante as coletas de dados pode afetar a sua continuidade e completude, principalmente em locais com maior risco de violência. Algumas coletas noturnas tiveram que ser interrompidas precocemente devido à preocupação com roubos, o que pode prejudicar a qualidade e representatividade dos dados coletados;
- Clima: o clima pode ser uma limitação durante a coleta de dados. Durante o período chuvoso, em dias com incidência de raios, foi necessário interromper a coleta para garantir a segurança dos participantes e dos equipamentos. Isso pode levar a interrupções imprevistas e afetar a continuidade e qualidade da coleta de dados;
- Orçamento: o uso de apenas um telefone na coleta de dados pode limitar a abrangência e precisão das medições devido a restrições orçamentárias e de plano de dados. O uso de mais de um telefone poderia permitir a realização de medições mais abrangentes e uma análise mais completa das condições de rede, aumentando a confiabilidade dos resultados obtidos.

5.3 Aplicação do Modelo de Predição de Uplink

Nesta etapa da dissertação, foi conduzido o processo de aplicação prática do modelo de predição da vazão de *uplink*, estruturado em três fases principais: análise de correlação, treinamento com algoritmos e geração das predições. Inicialmente, realizou-se uma análise de correlação para identificar os parâmetros da camada física que apresentaram maior relação com a variável de interesse. Essa análise foi fundamental para orientar a seleção das variáveis mais relevantes, garantindo que o modelo fosse alimentado com atributos representativos. E por fim, apresenta a discussão dos resultados da predição da vazão de *uplink*.

5.3.1 Análise de Correlação

A predição da vazão de *uplink* em redes móveis exige uma compreensão detalhada da relação entre os parâmetros da camada física e o desempenho da rede. Para isso, foi realizada uma análise de correlação com o objetivo de quantificar o grau de associação entre as variáveis. A correlação pode ser positiva, quando ambas as variáveis variam na mesma direção, ou negativa, quando variam em direções opostas.

Para avaliar a relação entre os principais parâmetros de qualidade da rede, o coeficiente de Spearman é calculado conforme a [Equação 5.1](#):

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \quad (5.1)$$

onde:

ρ é o coeficiente de correlação de Spearman,

d_i representa a diferença entre as posições no ranking da observação i ,

n é o número total de pares de observações.

Neste estudo, foi utilizada a matriz de correlação de Spearman (veja a [Figura 5.16](#)), na imagem foi possível destacar os principais parâmetros de qualidade da rede (RSRP, RSRQ, SNR e RSSI) e as taxas de vazão de *uplink* e *downlink*. Observa-se que o RSRP possui forte correlação positiva com o RSSI (0,86), o que é esperado, dado que ambos refletem a

intensidade do sinal recebido. Além disso, destaca-se a correlação significativa entre o RSSI e as vazões de *uplink* (0,52) e *downlink* (0,47), indicando sua influência direta no desempenho da rede. A elevada correlação (0,93) entre as vazões de *uplink* e *downlink* também sugere que ambos os sentidos de transmissão podem ser impactados de forma semelhante pelas condições da rede. Esses resultados servem como base para a seleção dos atributos mais relevantes a serem utilizados na modelagem preditiva.

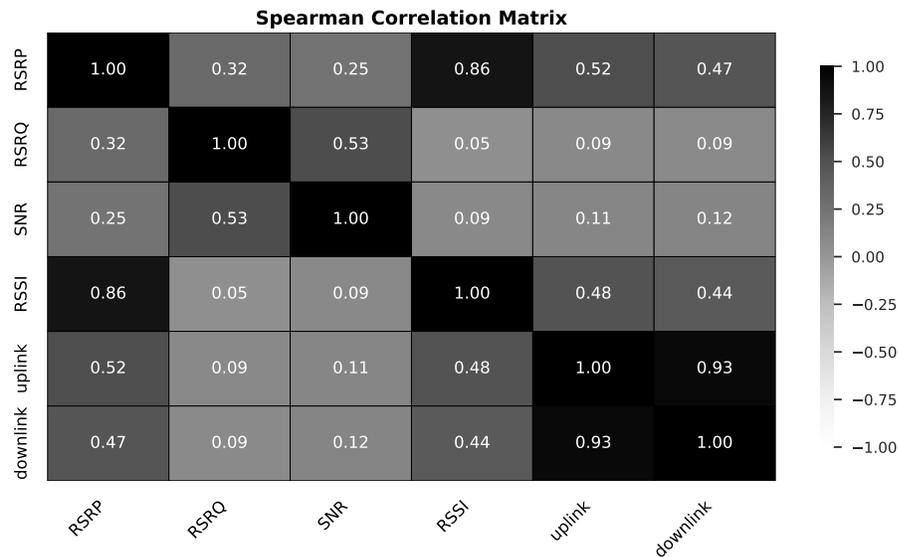


Figura 5.16: matriz de correlação de Spearman destacando a relação entre os parâmetros de qualidade da rede e a vazão de *uplink* em redes 4G LTE

Fonte: elaboração própria (2025).

A análise de dispersão reforça essa tendência. Conforme ilustrado na [Figura 5.17a](#), quando o RSRP está abaixo de -100 dBm, a vazão de *uplink* é significativamente reduzida. Em contraste, valores entre -90 dBm e -70 dBm estão associados a taxas de transmissão mais altas, reforçando a importância do RSRP como principal indicador de desempenho de *uplink*.

O RSRQ, por sua vez, mede a qualidade do sinal de referência recebido e é influenciado pela carga da rede e pela interferência entre células. Seus valores variam de -20 dB (alta interferência) a -8 dB (baixa interferência), conforme mostrado na [Figura 5.17b](#). No entanto, sua correlação mais fraca indica que esse parâmetro é menos determinante para a predição de vazão.

A relação sinal-ruído (SNR) representa a relação entre a potência do sinal útil e o

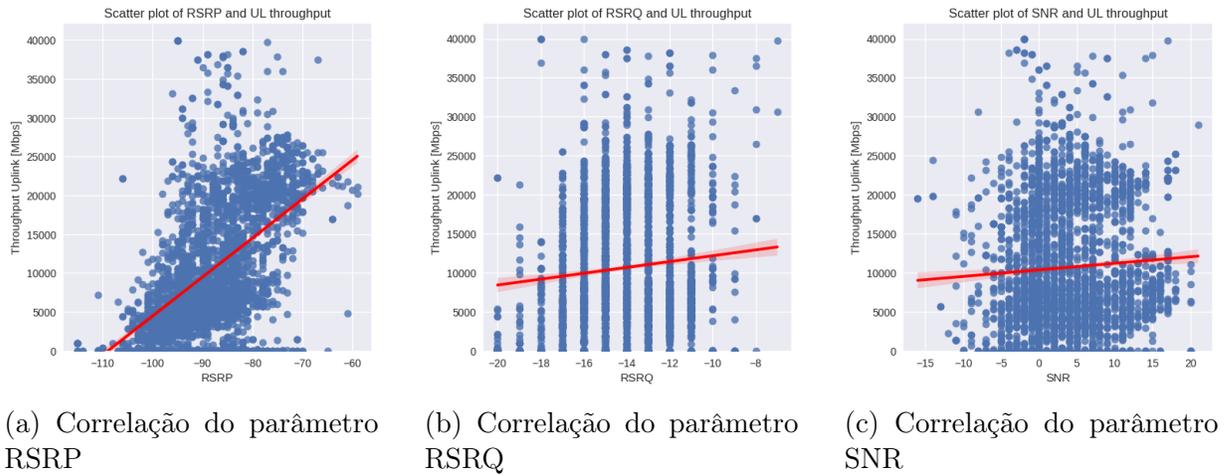


Figura 5.17: correlação entre parâmetros de qualidade da rede e vazão de *uplink*
 Fonte: elaboração própria (2025).

ruído de fundo da rede. Embora sua correlação com a vazão não seja expressiva, valores elevados de SNR indicam melhores condições de transmissão. A regressão linear apresentada na Figura 5.17c sugere uma relação positiva entre SNR e vazão de *uplink*, embora mais fraca que a observada com o RSRP.

Com base na análise de correlação, o conjunto de dados foi particionado em 70% para treinamento e 30% para teste. Foram avaliados cinco modelos de aprendizado de máquina para a predição da vazão de *uplink*: regressão linear (LR), floresta aleatória (RF), k-vizinhos mais próximos (KNN), máquina de vetores de suporte (SVM) e rede neural (NN).

A avaliação do desempenho preditivo dos modelos foi conduzida por meio de quatro métricas de erro: erro quadrático médio (MSE), erro absoluto médio (MAE), raiz do erro quadrático médio (RMSE) e coeficiente de determinação (R^2).

Essas métricas fornecem uma avaliação abrangente do desempenho preditivo dos modelos. Enquanto o MSE e o RMSE penalizam mais intensamente grandes erros de predição, o MAE oferece uma medida mais intuitiva do erro médio absoluto. Já o coeficiente de determinação (R^2) expressa a capacidade explicativa do modelo em relação à variabilidade dos dados, sendo que valores próximos de 1 indicam elevada capacidade preditiva.

A Figura 5.16 aponta que o RSRP apresenta a maior correlação com a vazão de *uplink* (0,53), indicando que valores mais elevados de RSRP estão associados a maiores taxas de transmissão. O RSSI possui a segunda maior correlação (0,49), enquanto RSRQ e SNR

apresentam correlações mais baixas (0,13 e 0,15, respectivamente), sugerindo um impacto menos significativo na vazão.

5.3.2 Discussão dos Resultados da Predição da Vazão de Uplink

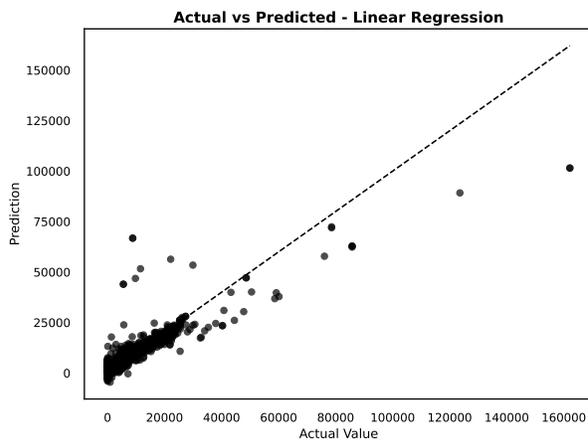
A [Figura 5.18](#) apresenta a comparação entre os valores observados e os valores previstos pelos diferentes algoritmos de aprendizado de máquina avaliados. O eixo x representa os valores reais da vazão de *uplink*, enquanto o eixo y corresponde aos valores previstos. Cada ponto no gráfico refere-se a uma amostra do conjunto de teste, o qual representa 30% dos dados coletados e foi utilizado exclusivamente para fins de avaliação preditiva.

Cabe ressaltar que 70% dos dados foram utilizados para o treinamento dos modelos, sendo os gráficos uma forma de ilustrar a capacidade dos algoritmos de generalizar para novos dados.

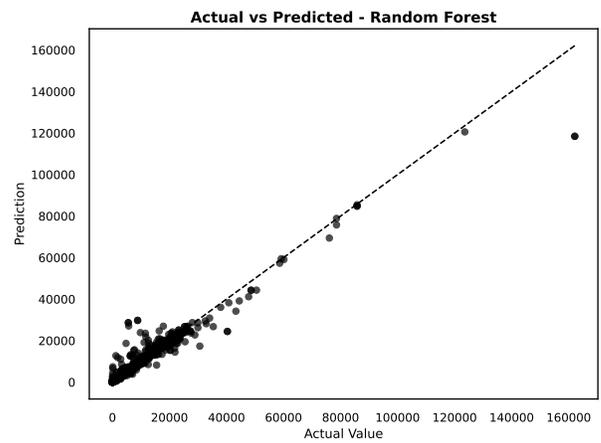
Em um cenário ideal de predição, todos os pontos estariam alinhados sobre a diagonal $y = x$, indicando correspondência perfeita entre os valores reais e os valores previstos. Assim, a dispersão dos pontos em torno dessa linha diagonal permite uma análise visual da acurácia do modelo: quanto mais próximos da diagonal, melhor o desempenho preditivo.

O modelo de regressão linear, conforme ilustrado na [Figura 5.18a](#), apresenta um ajuste limitado, demonstrando dificuldade em capturar as não-linearidades inerentes ao problema. Em contraste, o modelo floresta aleatória, (veja a [Figura 5.18b](#)), exibe maior proximidade entre os valores previstos e observados, evidenciando um desempenho mais robusto.

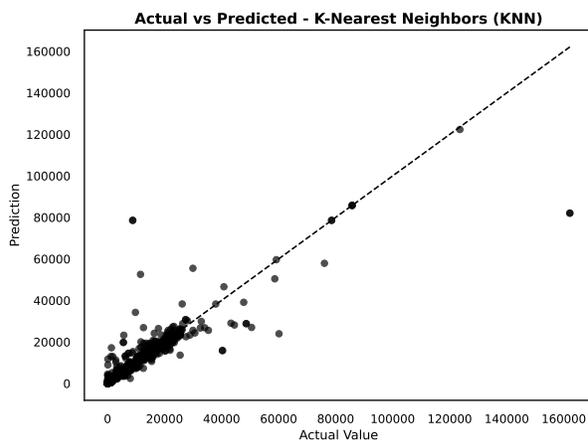
Os modelos k vizinhos mais próximos e máquina de vetores de suporte, representados na [Figura 5.18c](#) e [Figura 5.18d](#), respectivamente, exibem padrões semelhantes, com desvios mais acentuados em valores extremos, indicando limitações em suas capacidades de generalização. Por fim, a rede neural, apresentada na [Figura 5.18e](#), exibe uma tendência de ajuste, embora com maior dispersão em torno da linha ideal, o que pode sugerir desafios na convergência do treinamento ou a necessidade de otimização dos hiperparâmetros.



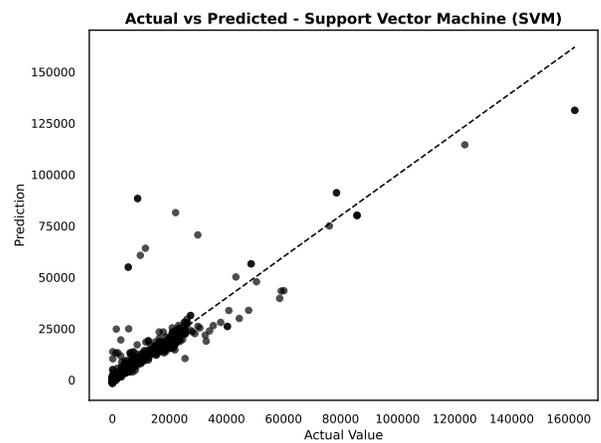
(a) regressão linear (LR)



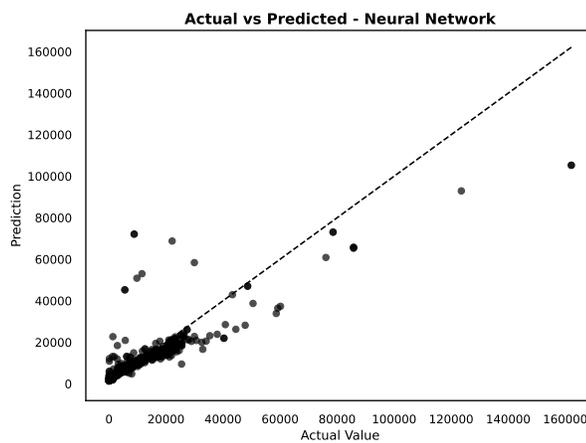
(b) floresta aleatória (RF)



(c) k vizinhos mais próximos (KNN)



(d) máquina de vetores de suporte (SVM)



(e) Rede Neural (NN)

Figura 5.18: comparação entre valores observados e previstos para diferentes algoritmos de aprendizado de máquina

Fonte: elaboração própria (2025).

A [Tabela 5.7](#) apresenta a comparação de desempenho entre os modelos utilizados para a predição da vazão de *uplink*. As métricas avaliadas incluem MSE, R^2 , MAE e RMSE.

Tabela 5.7: resultados finais da predição da vazão de *uplink* utilizando diferentes algoritmos de aprendizado de máquina.

Modelo	MSE	R^2	MAE	RMSE
LR	56.540	0.760	3591.700	7519.250
RF	16.480	0.930	1654.500	4060.030
KNN	56.250	0.760	2427.630	7500.020
SVM	60.360	0.740	2713.780	7769.130
NN	60.440	0.740	3644.800	7774.060

Fonte: elaboração própria (2025).

Cada modelo foi treinado com o conjunto de treinamento e, posteriormente, o conjunto de teste foi utilizado para gerar as predições. Os resultados apresentados na tabela correspondem ao desempenho de cada modelo sobre os dados de teste.

A predição da vazão de *uplink* foi realizada por meio de aprendizado supervisionado, no qual os modelos são treinados com exemplos rotulados e aprendem a mapear as variáveis de entrada aos valores de saída. Durante o treinamento, aplicou-se a técnica de validação cruzada *k-fold*, com o objetivo de proporcionar uma avaliação mais robusta do desempenho dos modelos e mitigar o risco de *overfitting*.

O valor de k foi definido como 5, dividindo-se o conjunto de treinamento em cinco subconjuntos; em cada iteração, quatro foram utilizados para treinamento e um para validação, até que todos fossem empregados como subconjunto de validação. A escolha por $k = 5$ está alinhada com recomendações amplamente aceitas na literatura, como as de [Kohavi \(1995\)](#) e [James et al. \(2013\)](#), que indicam que valores entre 5 e 10 representam um equilíbrio adequado entre viés, variância e custo computacional. Dessa forma, $k = 5$ foi considerado suficiente para fornecer estimativas confiáveis do erro de generalização, mantendo a eficiência computacional.

Os resultados obtidos indicam que o modelo RF apresentou o melhor desempenho geral, evidenciado pelo menor (MSE = 16,480), maior ($R^2 = 0,930$), menor (MAE = 1654,500) e menor (RMSE = 4060,030). Esses valores sugerem que o modelo foi o mais eficaz em capturar as relações complexas entre os parâmetros de entrada e a vazão de *uplink*.

Os modelos de LR e KNN apresentaram desempenhos semelhantes, com coeficiente de determinação em torno de $R^2 = 0,76$, indicando capacidade preditiva moderada. Contudo, ambos registraram valores mais elevados para MSE, MAE e RMSE em comparação ao RF, refletindo maior erro nas previsões.

Por sua vez, os modelos SVM e NN exibiram o pior desempenho, com R^2 aproximado de 0,74 e métricas de erro superiores. Esses resultados indicam que tais abordagens podem não ser as mais adequadas para este conjunto de dados, possivelmente em função da complexidade das relações entre as variáveis ou da necessidade de ajustes mais refinados nos hiperparâmetros.

5.4 Resumo do Capítulo

Este capítulo apresentou a aplicação do modelo de predição da vazão de *uplink* em redes móveis, desde a descrição do protocolo experimental até a análise dos resultados obtidos. A coleta de dados foi realizada utilizando tecnologias acessíveis, como smartphone, chip de operadora e aplicativo Armadeira, com deslocamento em motocicleta. A estrutura experimental foi cuidadosamente planejada para assegurar a confiabilidade dos dados, utilizando ambientes como Google Colab e bibliotecas de ciência de dados em Python.

Foram descritos o processo de coleta, o fluxo de dados e as principais características da base construída, com análise exploratória minuciosa dos parâmetros de qualidade de rede, como RSRP, RSRQ, RSSI e SNR. A análise gráfica, estatística e espacial revelou a distribuição, a consistência e as possíveis inconsistências dos dados, fundamentais para as etapas subsequentes de modelagem.

A partir da análise de correlação, foram selecionadas as variáveis mais relevantes para a predição da vazão de *uplink*, sendo avaliados cinco modelos de aprendizado de máquina: regressão linear, floresta aleatória, k vizinhos mais próximos, máquina de vetores de suporte e rede neural. A floresta aleatória destacou-se como o modelo de melhor desempenho, alcançando o menor erro e o maior coeficiente de determinação (R^2). O estudo comprovou a eficácia do modelo proposto e contribuiu para a predição da vazão de *uplink* de forma estruturada, replicável e aplicável a diferentes contextos de redes móveis.

Capítulo 6

Conclusão

Este capítulo apresenta as conclusões gerais deste trabalho de dissertação, sintetizando as atividades desenvolvidas, as contribuições alcançadas e as perspectivas futuras de continuidade e aprofundamento da pesquisa.

6.1 Considerações Finais

O estudo da predição da vazão em dispositivos móveis, com foco na vazão de *uplink*, mostrou-se fundamental para o aprimoramento da qualidade de serviços em redes móveis e para a melhoria da experiência do usuário. Este trabalho abordou conceitos fundamentais de redes de comunicação móvel e técnicas de aprendizado de máquina, embasando-se em fontes científicas atualizadas, como artigos, livros e literatura especializada.

Com o objetivo de aprofundar o conhecimento no tema, foi realizada uma revisão sistemática da literatura que permitiu identificar os principais métodos já propostos, suas limitações e as lacunas existentes, especialmente no que tange à predição de *uplink*, que permanece subexplorada em comparação ao *downlink*.

A análise crítica dos estudos correlatos serviu de base para a formulação da metodologia proposta, que envolveu o desenvolvimento de um conjunto experimental próprio, focado na coleta estruturada de parâmetros de rede móvel. Essa abordagem permitiu a criação de uma base de dados específica e adequada para o treinamento e validação de modelos de aprendizado de máquina.

A partir da aplicação do método desenvolvido, foi possível avaliar diferentes algoritmos preditivos, destacando-se RF como a melhor solução para a predição da vazão de *uplink*. Os resultados obtidos demonstram a eficácia da metodologia proposta, além de abrir caminho para novas investigações sobre otimização e predição em redes móveis. No entanto, algumas limitações devem ser consideradas para a interpretação e generalização dos resultados.

O método adotado baseou-se em coleta de dados em campo utilizando o smartphone, o que, embora seja uma abordagem prática e acessível, exige a contratação de um plano de dados ativo para a medição em tempo real da vazão de *uplink*. Essa demanda pode limitar a aplicabilidade do método em larga escala, especialmente em regiões com restrições de conectividade móvel.

Além disso, as variações nos horários de coleta de dados afetaram a continuidade do experimento. Em determinados momentos, especialmente no período noturno, a coleta teve que ser interrompida por questões de segurança, devido à exposição a áreas de alto risco. Essa limitação pode impactar a representatividade dos dados, uma vez que determinados períodos do dia (como o noturno) apresentaram maior volume de registros, enquanto outros foram menos explorados.

Outro fator relevante refere-se às condições climáticas, que influenciaram diretamente a coleta. Chuvas intensas e tempestades exigiram a interrupção das atividades para garantir a segurança dos pesquisadores e dos equipamentos, o que pode ter gerado lacunas pontuais no conjunto de dados.

O principal objetivo desta dissertação foi desenvolver um modelo preditivo capaz de estimar a vazão de *uplink* em redes de telefonia móvel, com base em dados coletados por dispositivos portáteis. Para atingir esse propósito, foram realizadas quatro etapas principais: a coleta estruturada de parâmetros da camada física da rede móvel; a identificação dos fatores com maior correlação com a vazão de *uplink*; a construção do modelo de predição; e, por fim, a avaliação experimental da proposta, incluindo a comparação com abordagens presentes na literatura. O trabalho buscou, assim, oferecer uma solução eficiente, prática e cientificamente fundamentada para auxiliar na compreensão e otimização do desempenho de redes móveis.

Por fim, aspectos ambientais como densidade urbana, velocidade de deslocamento e interferência de sinal também podem influenciar a variação da vazão de *uplink*, exigindo

que esses fatores sejam considerados em futuras investigações que visem validar ou estender o modelo proposto.

6.2 Trabalhos Futuros

Embora esta dissertação tenha alcançado importantes resultados, algumas atividades complementares são sugeridas para a continuidade dos trabalhos:

1. Realizar coletas adicionais de dados utilizando diferentes dispositivos móveis e operadoras, ampliando a robustez do conjunto de dados;
2. Avaliar a influência de variáveis ambientais (como clima e horário) na qualidade da rede e no desempenho dos modelos preditivos;
3. Aplicar novos algoritmos de aprendizado de máquina, incluindo modelos baseados em *deep learning*, para comparação de desempenho;
4. Realizar estudos de validação externa em outras cidades e regiões, visando testar a capacidade de generalização do modelo proposto.

6.3 Principais Contribuições

Abaixo estão apresentadas as principais contribuições desta dissertação de mestrado, destacando os pontos relevantes obtidos a partir da pesquisa:

- A primeira contribuição foi a realização de uma revisão sistemática sobre a predição da vazão que permitiu uma análise aprofundada das pesquisas já existentes nesta área;
- Além disso, foi desenvolvido um aplicativo de coleta de dados de rede, denominado Armadeira, que permite a coleta de informações relevantes sobre dados móveis de rede em smartphones Android;
- Descrição da coleta de dados de rede, com ênfase nas ferramentas e etapas necessárias para sua execução. Fornece uma visão geral de como esse processo pode ser realizado de maneira eficiente e estruturada;

- Elaborou-se um conjunto de dados 4G LTE de parâmetros de rede sem fio. A base de dados coletada será divulgada, permitindo que outros pesquisadores possam utilizá-la em seus estudos e pesquisas;
- Outra importante contribuição deste estudo foi a apresentação de um método para predição da vazão em redes de telefonia móvel. Foram avaliados diferentes algoritmos a fim de identificar aqueles com melhor desempenho em termos de precisão e eficiência.

As contribuições do estudo fortalecem o conhecimento científico sobre a predição de vazão em redes de telefonia móvel, oferecendo subsídios relevantes para o desenvolvimento de soluções que aprimoram o desempenho das redes, proporcionando melhorias de qualidade na entrega do serviço e uma experiência mais satisfatória para os usuários finais.

6.4 Produção Científica

Como resultado desta dissertação, dois artigos científicos foram desenvolvidos e submetidos para publicação, consolidando as contribuições alcançadas durante a pesquisa:

- **Network Data Throughput on Mobile Devices: A Systematic Literature Review** — submetido à revista *Journal of Network and Computer Applications* (2025), abordando uma revisão sistemática da literatura sobre predição de throughput em dispositivos móveis;
- **Predictive Modeling of Uplink Throughput in Mobile Networks: Insights from Machine Learning** — submetido à revista *Digital Communications and Networks* (2025), Modelagem Preditiva da Vazão de Uplink em Redes Móveis: Percepções a partir do Aprendizado de Máquina;
- **A multi-device and multi-operator dataset from mobile network coverage on Android devices** — publicado *Data in Brief* (2024), O conjunto de dados compreende amostras coletadas.

Esses artigos representam a consolidação dos principais resultados obtidos e contribuem para o avanço do conhecimento na área de redes móveis e aprendizado de máquina.

Referências Bibliográficas

- Afroz, F., Subramanian, R., Heidary, R., Sandrasegaran, K. e Ahmed, S. (2015), Sinr, rsrp, rssi and rsrq measurements in long term evolution networks, *International Journal of Wireless Mobile Networks* 7, 113–123. URL : <https://doi.org/10.5121/ijwmn.2015.7409>
- Al-Thaedan, A., Shakir, Z., Mjhoor, A. Y., Alsabah, R., Al-Sabbagh, A., Salah, M. e Zec, J. (2023), Downlink throughput prediction using machine learning models on 4g-lte networks, *International Journal of Information Technology* 15, 1347–1355.
- Al-Thaedan, M. *et al.* (2024), A machine learning framework for predicting downlink throughput in 4g-lte/5g cellular networks, *Journal of Computer Networks and Communications* 2024, 1–13. URL : <https://doi.org/10.1155/2024/8856391>
- Altman, N. S. (1992), An introduction to k-nearest neighbor, *The American Statistician* 46(3), 175–177.
- Amin, F., Majeed, A., Mateen, A., Abbasi, R. e Hwang, S. O. (2022), A systematic survey on the recent advancements in the social internet of things, *IEEE Access* 10, 63867–63884. URL : <https://doi.org/10.1109/ACCESS.2022.3183261>
- Aprillia, O. M., Saragih, Y., Lammada, I., Agatha, E. S., Wijaya, A. e Van, N. T. N. (2020), Rf 4g network performance result based on android drive test collection tools automatically extracted by yaiao application at the universitas singaperbangsa karawang case study, IEEE, pp. 291–295. URL : <https://doi.org/10.1109/MECnIT48290.2020.9166673>
- Batool, A., Fouda, M. M. e Fadlullah, Z. M. (2024), Deep learning-based throughput prediction in 5g cellular networks, *Proceedings of the 2024 International Conference on Advanced Communication Technologies* . Disponível em: *IEEE Xplore*, acesso em maio de 2025.

- Baxhaku, I., Rimac-Drlje, S. e Shahu, D. (2018), Video quality testing for skype and viber over 4g mobile network, *2018 International Symposium ELMAR* pp. 165–168. URL : <https://doi.org/10.23919/ELMAR.2018.8534694>
- Bayram, K. S., Kızrak, M. A. e Bolat, B. (2013), Classification of eeg signals by using support vector machines, *in 2013 IEEE INISTA*, pp. 1–3. URL : <https://doi.org/10.1109/INISTA.2013.6577636>
- Ben Hcine, M. e Bouallegue, R. (2015), Analysis of uplink effective sinr in lte networks, *in 2015 International Wireless Communications and Mobile Computing Conference (IWCMC)*, pp. 425–430. URL : <https://doi.org/10.1109/IWCMC.2015.7289121>
- Bhaskaraputra, A., Sutojo, F., Ramadhan, A. N., Gunawan, A. A. S. e Anderies (2022), Systematic literature review on solving personalization problem in digital marketing using machine learning and its impact, pp. 178–182. URL : <https://doi.org/10.1109/iSemantic55962.2022.9920387>
- Bishop, C. M. (2011), *Pattern Recognition and Machine Learning*, 2nd corr. ed. edn, Springer, New York.
- Biswas, M., Chakraborty, A. e Palit, B. (2024), A kalman filter based low complexity throughput prediction algorithm for 5g cellular networks, *IEEE Transactions on Vehicular Technology* 73(5), 7089–7101. URL : <https://doi.org/10.1109/TVT.2023.3339649>
- Cao, J., Ma, M., Li, H., Zhang, Y. e Luo, Z. (2014), A survey on security aspects for lte and lte-a networks, *IEEE Communications Surveys Tutorials* 16, 283–302. URL : <https://doi.org/10.1109/SURV.2013.041513.00174>
- Carvalho, M., Silva, E. B., Silva, V. F. e Macedo, D. F. (2019), Qd4g: Qoe para vídeo em redes d2d/4g com aprendizado de máquina, *Anais do XXXVII Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos (SBRC 2019)* pp. 183–196. URL : <https://doi.org/10.5753/sbrc.2019.7359>
- Chen, T. e Guestrin, C. (2016), Xgboost: A scalable tree boosting system, *in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data*

- Mining, KDD '16, Association for Computing Machinery, New York, NY, USA, p. 785–794. URL : <https://doi.org/10.1145/2939672.2939785>
- Chmieliauskas, D. e Paulikas, (2025), Evaluation of uplink video streaming qoe in 4g and 5g cellular networks using real-world measurements, *IEEE Access* 13, 53996–54018. URL : <https://doi.org/10.1109/ACCESS.2025.3554340>
- Dahlman, E., Parkvall, S. e Sköld, J. (2011), *4G: LTE/LTE-Advanced for Mobile Broadband*, academic press edn.
- Elsherbiny, H., Nagib, A. M., Abou-zeid, H., Abbas, H. M., Hassanein, H. S., Noureldin, A., Bin Sediq, A. e Boudreau, G. (2020), 4g lte network data collection and analysis along public transportation routes, *in* GLOBECOM 2020 - 2020 IEEE Global Communications Conference, pp. 1–6. URL : <https://doi.org/10.1109/GLOBECOM42002.2020.9348031>
- Eyceyurt, E., Egi, Y. e Zec, J. (2022), Machine-learning-based uplink throughput prediction from physical layer measurements, *Electronics* 11, 1227. URL : <https://doi.org/10.3390/electronics11081227>
- Ferreira, B., Gaspar, B., Paiva, S., Santos, A. e Cabral, J. (2021), Coverage and deployment analysis of nb-iot technology under various environment scenarios, *in* 2020 2nd International Conference on Societal Automation (SA), pp. 1–7. URL : <https://doi.org/10.1109/SA51175.2021.9507153>
- Gaber, A., Zaki, M. M. e Mohamed, A. M. (2020), 4g uplink power control tuning approach using unsupervised machine learning, pp. 437–442. URL : <https://doi.org/10.1109/ICOIN48656.2020.9016527>
- Gautam, A. e Dharmaraja, S. (2023), Reliability and survivability assessment of lte-a architecture and networks, *OPSEARCH* 60, 370–392. URL : <https://doi.org/10.1007/s12597-022-00607-y>
- George, S. V., Mathews, L. e Pillai, S. S. (2015), Performance analysis of transmission modes in 4g lte/lte-a downlink systems, *in* 2015 International Conference on Control,

- Instrumentation, Communication and Computational Technologies (ICCICCT), pp. 219–224. URL : <https://doi.org/10.1109/ICCICCT.2015.7475279>
- Geron, A. (2019), *Mãos à Obra: Aprendizado de Máquina com Scikit-Learn, Keras Tensor-Flow*, Alta Books.
- Goodfellow, I., Bengio, Y. e Courville, A. (2016), *Deep Learning*, MIT Press. URL : <http://www.deeplearningbook.org>
- Gódor, G., Jakó, Z., Ádám Knapp e Imre, S. (2015), A survey of handover management in lte-based multi-tier femtocell networks: Requirements, challenges and solutions, *Computer Networks* 76, 17–41. URL : <https://doi.org/10.1016/j.comnet.2014.10.016>
- Hadian, S., Sikora, A., Welte, D., Schappacher, M. e Sowieja, F. (2024), A performance scoring approach for qos optimization in private 5g networks, pp. 1–7. URL : <https://doi.org/10.1109/ICETI63946.2024.10777279>
- Hajlaoui, E., Khelifi, A., Zaier, A., Ghodhbane, J., Ben Hamed, M. e Sbita, L. (2019), Performance evaluation of lte physical layer, *in* 2019 International Conference on Internet of Things, Embedded Systems and Communications (IINTEC), pp. 106–111. URL : <https://doi.org/10.1109/IINTEC48298.2019.9112117>
- Holma, H. e Toskala, A. (2011), *LTE para UMTS: Evolução para LTE-Advanced*, john wiley filhos edn.
- Ismail, M. A., Roslee, M., Nordin, M. S., Hashim, S. H. S. e Mohd Razilan, A. R. (2024), Analyzing the influence of timing advance (ta) and reference signal received power (rsrp) on lte geolocation, pp. 377–380. URL : <https://doi.org/10.1109/APACE62360.2024.10877347>
- James, G., Witten, D., Hastie, T. e Tibshirani, R. (2013), *An Introduction to Statistical Learning: with Applications in R*, Springer, New York.
- Jha, H. e Vijayarajan, V. (2020), Mobile internet throughput prediction using machine learning techniques, *2020 International Conference on Smart Electronics and Communication (ICOSEC)* pp. 253–257. URL : <https://doi.org/10.1109/ICOSEC49089.2020.9215436>

- Jordan, M. I. e Mitchell, T. M. (2015), Machine learning: Trends, perspectives, and prospects, *Science* 349, 255–260. URL : <https://doi.org/10.1126/science.aaa8415>
- Kassim, M., Rahman, R. A., Aziz, M. A. A., Idris, A. e Yusof, M. I. (2017), Performance analysis of voip over 3g and 4g lte network, *2017 International Conference on Electrical, Electronics and System Engineering (ICEESE)* pp. 37–41. URL : <https://doi.org/10.1109/ICEESE.2017.8298391>
- Kelleher, J. D., Tierney, B. e Tierney, A. D. (2015), *Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies*, The MIT Press.
- Kitchenham, B., Brereton, O. P., Budgen, D., Turner, M., Bailey, J. e Linkman, S. (2009), Systematic literature reviews in software engineering – a systematic literature review, *Information and Software Technology* 51, 7–15. URL : <https://doi.org/10.1016/j.infsof.2008.09.009>
- Kitchenham, B. e Brereton, P. (2013), A systematic review of systematic review process research in software engineering, *Inf. Softw. Technol.* 55, 2049–2075. URL : <https://doi.org/10.1016/j.infsof.2013.07.010>
- Kohavi, R. (1995), A study of cross-validation and bootstrap for accuracy estimation and model selection, *in* Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 2, IJCAI'95, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, p. 1137–1143.
- Kousias, K., Alay, O., Argyriou, A., Lutu, A. e Riegler, M. (2019), Estimating downlink throughput from end-user measurements in mobile broadband networks, *2019 IEEE 20th International Symposium on "A World of Wireless, Mobile and Multimedia Networks"(WoWMoM)* pp. 1–10. URL : <https://doi.org/10.1109/WoWMoM.2019.8792968>
- Krasniqi, F., Maraj, A. e Blaka, E. (2018), Performance analysis of mobile 4g/lte networks, *2018 South-Eastern European Design Automation, Computer Engineering, Com-*

- puter Networks and Society Media Conference (SEEDA_CECNSM)* pp. 1 – 5. URL : <https://doi.org/10.23919/SEEDA – CECNSM.2018.8544937>
- Lawal, I. A. (2020), Gmdh modelling for mobile user throughput forecasting, *Proceedings of the 35th Annual ACM Symposium on Applied Computing* pp. 1021–1025. URL : <https://doi.org/10.1145/3341105.3373843>
- Mafra, S. N. e Travassos, G. H. (2006), Estudos primários e secundários apoiando a busca por evidência em engenharia de software, Relatório Técnico RT–ES 687/06, Programa de Engenharia de Sistemas e Computação, COPPE/UFRJ, Rio de Janeiro. URL : <http://www.cos.ufrj.br/uploadfile/1149103120.pdf>
- Mamalakis, M., Dwivedi, K., Sharkey, M., Alabed, S., Kiely, D. e Swift, A. J. (2023), A transparent artificial intelligence framework to assess lung disease in pulmonary hypertension, *Scientific Reports* 13, 3812. URL : <https://doi.org/10.1038/s41598-023-30503-4>
- Markappa, B., Thippeswamy, M. e Raghavendra, V. (2024), Comparative analysis of lstm and gru for uplink data rate prediction in 5g networks, *in Proceedings of the International Conference on Advances in Computing*.
- Mauro, M. D. e Liotta, A. (2020), An experimental evaluation and characterization of voip over an lte-a network, *IEEE Transactions on Network and Service Management* 17, 1626–1639. URL : <https://doi.org/10.1109/TNSM.2020.2995505>
- Meirelles, F. S. (2024), Pesquisa do uso da ti - tecnologia de informação nas empresas, Relatório técnico, Fundação Getulio Vargas (FGV). Acesso em: jan. 2025. URL : https://eaesp.fgv.br/sites/eaesp.fgv.br/files/u68/pesti-fgvicia-2024_0.pdf
- Minovski, D., Ahlund, C., Mitra, K. e Johansson, P. (2019), Analysis and estimation of video qoe in wireless cellular networks using machine learning, *2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX)* pp. 1–6. URL : <https://doi.org/10.1109/QoMEX.2019.8743281>

- Minovski, D., Ogren, N., Ahlund, C. e Mitra, K. (2021), Throughput prediction using machine learning in lte and 5g networks, *IEEE Transactions on Mobile Computing* pp. 1–1. URL : <https://doi.org/10.1109/TMC.2021.3099397>
- Miyim, A. M. e Wakili, A. (2019), Performance evaluation of lte networks, *2019 15th International Conference on Electronics, Computer and Computation (ICECCO)* pp. 1–6. URL : <https://doi.org/10.1109/ICECCO48375.2019.9043271>
- Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S. e Hassabis, D. (2015), Human-level control through deep reinforcement learning, *Nature* 518, 529–533. URL : <https://doi.org/10.1038/nature14236>
- Mohri, M., Rostamizadeh, A. e Talwalkar, A. (2018), *Foundations of Machine Learning*, MIT Press.
- Mostafa, A., Elattar, M. A. e Ismail, T. (2022), Downlink throughput prediction in lte cellular networks using time series forecasting, *in 2022 International Conference on Broadband Communications for Next Generation Networks and Multimedia Applications (CoBCom)*, pp. 1–4. URL : <https://doi.org/10.1109/CoBCom55489.2022.9880654>
- Narayanan, A., Ramadan, E., Mehta, R., Hu, X., Liu, Q., Fezeu, R. A. K., Dayalan, U. K., Verma, S., Ji, P., Li, T., Qian, F. e Zhang, Z.-L. (2020), Lumos5g: Mapping and predicting commercial mmwave 5g throughput, *ACM*, pp. 176–193. URL : <https://doi.org/10.1145/3419394.3423629>
- Nikula, M., Correia, L. M., Grilo, A., Mähönen, P., Santo, L. e Dinis, R. (2025), Machine learning-based anomaly detection and root cause analysis in mobile networks, *in 2025 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 01–06. URL : <https://doi.org/10.1109/WCNC61545.2025.10978282>
- Oussakel, I., Owezarski, P. e Berthou, P. (2019a), Cellular uplink bandwidth prediction

- based on radio measurements, Association for Computing Machinery, pp. 111–118. URL : <https://doi.org/10.1145/3345770.3356733>
- Oussakel, I., Owezarski, P. e Berthou, P. (2019b), Experimental estimation of lte-a performance, *2019 15th International Conference on Network and Service Management (CNSM)* pp. 1–9. URL : <https://doi.org/10.23919/CNSM46954.2019.9012663>
- Palaios, A., Vielhaus, C. L., Külzer, D. F., Watermann, C., Hernangómez, R., Partani, S., Geuer, P., Krause, A., Sattiraju, R., Kasparick, M., Fettweis, G. P., Fitzek, F. H. P., Schotten, H. D. e Stańczak, S. (2023), Machine learning for qos prediction in vehicular communication: Challenges and solution approaches, *IEEE Access* 11, 92459–92477. URL : <https://doi.org/10.1109/ACCESS.2023.3303528>
- Perla, S., K, N. N. e Potta, S. (2022), Implementation of autonomous cars using machine learning, *2022 International Conference on Edge Computing and Applications (ICECAA)* pp. 1444–1451. URL : <https://doi.org/10.1109/ICECAA55415.2022.9936102>
- Polak, L., Baranek, M., Kufa, J., Sotner, R. e Dluha, J. (2025), Cost-effective measurement setup for analyzing signal coverage in 4g/5g mobile networks, pp. 1–6. URL : <https://doi.org/10.1109/RADIOELEKTRONIKA65656.2025.11008431>
- Prajwala, J., Mathew, R. e Taj, N. (2018), Analysis of voip traffic over lte for different codecs, *2018 3rd IEEE International Conference on Recent Trends in Electronics, Information Communication Technology (RTEICT)* pp. 1858–1862. URL : <https://doi.org/10.1109/RTEICT42901.2018.9012408>
- Prasanth, S. V., Haritha, P., Likhitha, B., Mahalakshmi, K. e Akhther, A. P. (2025), Artificial intelligence based supervised learning model for attack detection of botnet and defence in iot cybersecurity infrastructures using convolutional neural network, *in 2025 International Conference on Inventive Computation Technologies (ICICT)*, pp. 1535–1540. URL : <https://doi.org/10.1109/10.1109/ICICT64420.2025.11004980>
- Putra, G. M., Budiman, E., Malewa, Y., Cahyadi, D., Taruk, M. e Hairah, U. (2021), 4g lte

- experience: Reference signal received power, noise ratio and quality, pp. 139–144. URL : <https://doi.org/10.1109/EIConCIT50028.2021.9431853>
- Raca, D., Zahran, A. H., Sreenan, C. J., Sinha, R. K., Halepovic, E., Jana, R. e Gopalakrishnan, V. (2020), On leveraging machine and deep learning for throughput prediction in cellular networks: Design, performance, and challenges, *IEEE Communications Magazine* 58, 11–17. URL : <https://doi.org/10.1109/MCOM.001.1900394>
- Rida, J. F. A. (2021), Overview of development performance for mobile phone wireless communication networks, *2021 International Conference on Electrical, Computer and Energy Technologies (ICECET)* pp. 1–11. URL : <https://doi.org/10.1109/ICECET52533.2021.9698519>
- Rolim, E. (2014), Análise da qualidade e do desempenho da cobertura celular long term evolution em curitiba, paraná, Trabalho de Conclusão de Curso (Graduação) – Universidade Federal do Paraná. 131 p.
- Salahdine, F., Han, T. e Zhang, N. (2023), 5g, 6g, and beyond: Recent advances and future challenges, *Annals of Telecommunications* . URL : <https://doi.org/10.1007/s12243-022-00938-3>
- Sathya, V., Kala, S. M. e Naidu, K. (2023), Heterogenous networks: From small cells to 5g nr-u, *Wireless Personal Communications* 128, 2779–2810. URL : <https://doi.org/10.1007/s11277-022-10070-z>
- Schmid, J., Schneider, M., HöB, A. e Schuller, B. (2019), A deep learning approach for location independent throughput prediction, pp. 1–5. URL : <https://doi.org/10.1109/ICCVE45908.2019.8965216>
- Sezgin, G., Coskun, Y., Basar, E. e Kurt, G. K. (2018), Performance evaluation of a live multi-site lte network, *IEEE Access* 6, 49690–49704. URL : <https://doi.org/10.1109/ACCESS.2018.2868385>
- Sharma, S., Dhir, S. e Punhani, R. (2022), Strategic improvement in malls for customers

- by using machine learning, Vol. 1, pp. 359–363. URL : <https://doi.org/10.1109/COM-IT-CON54601.2022.9850770>
- Souza, C. B. B. D., Arnez, J. J. A., Fernandes, T., Alves, C. A. T. e de Sousa, J. O. (2022), Analysis of power consumption in 4g volte and 5g vonr over ims network, pp. 59–64. URL : <https://doi.org/10.1109/CAMAD55695.2022.9966918>
- Stam, W. T., Ingwersen, E. W., Ali, M., Spijkerman, J. T., Kazemier, G., Bruns, E. R. J. e Daams, F. (2023), Machine learning models in clinical practice for the prediction of postoperative complications after major abdominal surgery, *Surgery Today* . URL : <https://doi.org/10.1007/s00595-023-02662-4>
- Sudarsan, S. J., Silas, S. e Kathrine, G. J. W. (2025), Performance analysis for business insights prediction using machine learning algorithms, pp. 1227–1233.
- Sutton, R. S. e Barto, A. G. (2018), *Reinforcement Learning: An Introduction*, 2 edn, The MIT Press.
- Topol, E. J. (2019), High-performance medicine: the convergence of human and artificial intelligence, *Nature Medicine* 25, 44–56. URL : <https://doi.org/10.1038/s41591-018-0300-7>
- Vats, P. e Samdani, K. (2019), Study on machine learning techniques in financial markets, *2019 IEEE International Conference on System, Computation, Automation and Networking (ICSCAN)* pp. 1–5. URL : <https://doi.org/10.1109/ICSCAN.2019.8878741>
- Wei, B., Okano, M., Kanai, K., Kawakami, W. e Katto, J. (2018), Throughput prediction using recurrent neural network model, *in 2018 IEEE 7th Global Conference on Consumer Electronics (GCCE)*, pp. 107–108. URL : [10.1109/GCCE.2018.8574877](https://doi.org/10.1109/GCCE.2018.8574877)
- Yammanuru, D. R. e Das, D. (2020), Novel data accounting procedure in 4g and 5g networks, *in 2020 IEEE 3rd 5G World Forum (5GWF)*, pp. 355–359. URL : <https://doi.org/10.1109/5GWF49715.2020.9221231>

Zhou, F. e Guo, H. (2013), Lte-a uplink antenna solution analysis, *in* 2013 8th International Conference on Communications and Networking in China (CHINACOM), pp. 296–301.
URL : <https://doi.org/10.1109/ChinaCom.2013.6694609>

Apêndice A

Descrição dos Dados Coletados

Neste apêndice, apresenta-se a [Tabela A.1](#), a qual sintetiza a descrição dos parâmetros coletados durante a execução do experimento, relacionados às características e à qualidade da rede móvel. A coleta dos dados foi realizada por meio do aplicativo desenvolvido, denominado Armadeira, projetado especificamente para capturar métricas relevantes de qualidade de serviço, bem como atributos técnicos associados ao desempenho da comunicação em redes móveis.

Tabela A.1: lista dos parâmetros coletados pelo aplicativo desenvolvido (Armadeira)

Nome do Parâmetro	Tipo	Valor	Unidade	Descrição
mobileStatus	Boolean	1 ou 0	–	Indica se o dispositivo está ligado (1) ou desligado (0).
netType	String	3G, 4G, 5G	–	Tipo de rede móvel em uso.
roaming	Boolean	–	–	Indica se está em roaming fora da rede de origem.
downstream Bandwidth	Int	Float	Kbps	Velocidade de download.
upstream Bandwidth	Int	Float	Kbps	Velocidade de upload.
rx	Long	Float	Bytes	Taxa de dados recebidos.
tx	Long	Float	Bytes	Taxa de dados transmitidos.
networkOperator	String	Ex.: Vivo	–	Operadora de rede.
chipOperator	String	Ex.: Vivo	–	Operadora do chip.
mcc	String	Ex.: 724	–	Código do país da rede.
mnc	String	Ex.: 11	–	Código da operadora.
asuLevel	Int	Ex.: 32	–	Intensidade absoluta do sinal (ASU).
cqi	Int	Float	–	Indicador de qualidade do canal (CQI).

Continua na próxima página

Nome do Parâmetro	Tipo	Valor	Unidade	Descrição
dbm	Int	Float	dBm	Potência do sinal recebido.
level	Int	Float	dBm	Indicador de intensidade do sinal (RSSI).
rsrp	Int	Float	dBm	Potência do sinal de referência.
rsrq	Int	Float	dB	Qualidade do sinal de referência.
rssi	Int	Float	dBm	Intensidade do sinal recebido.
rssnr	Int	Float	dB	Relação sinal-ruído de referência.
earfcn	Int	Float	–	Identificador do canal de frequência (EARFCN).
band	Int	Float	–	Banda de operação.
frequency	Int	Float	MHz	Frequência de operação.
ci	Int	Float	–	Identidade da célula (Cell Identity).
pci	Int	Float	–	Identificação física da célula (Physical Cell Identity).
tag	Int	Float	–	Etiqueta ou marcação do dado.
downLinkFrequency	Int	Float	MHz	Frequência de downlink.
upLinkFrequency	Int	Float	MHz	Frequência de uplink.

Fonte: elaboração própria (2025).