

# Análise de Redes Móveis para Predição de Mapas de Calor por meio de Técnicas de Aprendizado de Máquina

Evelim Bacury Rocha<sup>1</sup>, Márcio Éric Lamêgo Valente<sup>1</sup>, Vandermi João Da Silva<sup>1</sup>

<sup>1</sup>Instituto de Ciência e Tecnologia - Universidade Federal do Amazonas (UFAM)  
Caixa Postal 69103-128 - Manaus - AM - Brazil

{evelim.rocha, marcio.valente, vandermi}@ufam.edu.br

**Abstract.** *This project aims to use Machine Learning techniques to create heatmaps that represent signal quality in 5G and 4G LTE networks, especially in indoor and densely urban environments, contributing to the optimization of mobile networks. The goal is to generate heatmaps that help identify areas in need of improved signal coverage, particularly in complex environments. So far, signal quality data has been collected in both indoor and outdoor environments in the state of Amazonas. These data have undergone preprocessing for noise removal and handling of missing values. A prototype has been developed for visualizing the heatmaps, allowing for signal quality analysis across different locations. The next steps include applying advanced data mining techniques and building more robust predictive models, along with integrating new features into the prototype.*

**Resumo.** *Este projeto visa utilizar técnicas de Aprendizado de Máquina para criar mapas de calor que representem a qualidade do sinal em redes 5G e 4G LTE, especialmente em ambientes internos e urbanos densos, contribuindo para a otimização das redes móveis. O objetivo é gerar mapas de calor que auxiliem na identificação de áreas com necessidade de melhoria na cobertura de sinal, especialmente em ambientes complexos.*

*Até o momento, foram coletados dados de qualidade de sinal em ambientes internos e externos no Amazonas, os quais passaram por pré-processamento para remoção de ruídos e tratamento de valores ausentes. Um protótipo foi desenvolvido para a visualização dos mapas de calor, permitindo a análise da qualidade do sinal em diferentes localizações. As próximas etapas incluem a aplicação de técnicas avançadas de mineração de dados e a construção de modelos preditivos mais robustos, com a integração de novas funcionalidades ao protótipo.*

## 1. Introdução

Em ambientes externos (*outdoor*), as tecnologias de posicionamento baseadas em satélite desempenham papel crucial para serviços como navegação veicular e rastreamento de cargas. Contudo, em ambientes internos (*indoor*) e áreas urbanas densas, sua precisão é comprometida por obstáculos físicos e efeitos de propagação multipercurso, inviabilizando aplicações críticas [Mogyorósi et al. 2022].

Nesse cenário, a evolução das redes móveis — em especial a quinta geração (5G) — surge como solução promissora ao oferecer maior precisão de localização, maior lar-

gura de banda e cobertura contínua. Isso impõe novos desafios ao gerenciamento de redes, exigindo soluções baseadas em análise de dados. [Mogyorósi et al. 2022].

Técnicas de Aprendizado de Máquina (*Machine Learning* – ML) têm se destacado nesse contexto, permitindo automatizar processos, reduzir custos e melhorar a experiência do usuário [Sarkar and Debnath 2021]. Uma aplicação particularmente estratégica é a geração de mapas de calor, que representam a intensidade do sinal por área geográfica e auxiliam na identificação de zonas de baixa cobertura [Haskić and Radončić 2024].

Este estudo se justifica por sua relevância social e técnica: visa construir mapas de calor com base em dados reais coletados no Amazonas, uma região ainda pouco explorada nesse tipo de análise, empregando um fluxo metodológico estruturado e escalável.

## **2. Objetivos**

### **2.1. Objetivo Geral**

Analisar dados de redes móveis 5G e 4G LTE de diversas operadoras para construir mapas de calor baseados nas variáveis de qualidade do sinal, utilizando técnicas de aprendizado de máquina clássicas.

### **2.2. Objetivos Específicos**

- Mapear as técnicas de processamento de aprendizagem de máquina no contexto de redes móveis 5G e 4G LTE;
- Tratar os dados coletados para retirada de anomalias e dados faltantes;
- Extrair conhecimento de redes móveis localizadas em ambiente interno e externo por meio de algoritmo de aprendizagem de máquina clássicos;
- Construir modelos de predição de mapas de calor baseado nos dados coletados;
- Implantar o modelo em um protótipo inteligente em um dispositivo móvel.

## **3. Fundamentação Teórica**

### **3.1. Evolução das Redes Móveis**

Os sistemas de comunicação sem fio (*wireless*) desempenham um papel essencial na sociedade contemporânea, viabilizando aplicações nas áreas de entretenimento, negócios, comércio, saúde e segurança. Ao longo dos anos, essas redes evoluíram significativamente, desde a quarta geração (4G) *Long-Term Evolution* (LTE) até a implantação progressiva da tecnologia 5G em dispositivos móveis [Kaur et al. 2021].

A introdução do 4G revolucionou a forma como interagimos com dispositivos móveis, proporcionando conexões mais rápidas e estáveis. Inicialmente implementada na Noruega em 2009, a tecnologia 4G atingiu os requisitos de velocidade em 2011, consolidando-se como padrão global [Saari 2022]. Esse avanço possibilitou o surgimento de novos casos de uso, como *streaming* de vídeo em alta definição e jogos online em tempo real.

Contudo, o aumento exponencial do tráfego de dados móveis impulsionou o desenvolvimento do 5G, projetado para ser a espinha dorsal da infraestrutura digital moderna. Diferentemente de outras tecnologias, como Wi-Fi, Bluetooth Low Energy (BLE) e Ultra Wide Band (UWB), que possuem cobertura limitada, o 5G oferece um

serviço de localização contínuo tanto em ambientes internos quanto externos. Sua precisão depende diretamente da densidade de antenas e da qualidade do sinal de rádio [Morocho-Cayamcela et al. 2019].

À medida que surgem novas aplicações, como veículos autônomos, automação industrial, realidade virtual e monitoramento remoto de saúde, a complexidade das redes móveis aumenta consideravelmente. Nesse cenário, o Aprendizado de Máquina torna-se essencial, permitindo a análise e mineração de grandes volumes de dados para otimização dos sistemas [Zhang et al. 2019].

### 3.2. Qualidade do Sinal e Gerenciamento de Recursos

A aplicação de ML em redes sem fio não apenas aprimora o desempenho da infraestrutura, como também melhora a Qualidade de Serviço (*Quality of Service* – QoS), tornando a rede mais eficiente e responsiva. Um dos principais componentes que influenciam essa melhoria é o Gerenciamento de Recursos de Rádio (*Radio Resource Management* – RRM), responsável pela alocação eficiente do espectro de radiofrequência com o objetivo de maximizar a capacidade da rede, garantir a estabilidade dos serviços e manter a cobertura planejada [Afroz et al. 2015].

Segundo [Afroz et al. 2015], as quatro principais métricas de RRM no sistema LTE são:

- **CQI (Channel Quality Indicator)**: Indicador da qualidade do canal;
- **RSRP (Reference Signal Received Power)**: Potência recebida do sinal de referência;
- **RSRQ (Reference Signal Received Quality)**: Qualidade do sinal de referência recebido;
- **RSSI (Received Signal Strength Indicator)**: Indicador de intensidade do sinal recebido.

Além disso, a relação sinal-interferência mais ruído (SINR) é uma métrica crucial para a adaptação de link e agendamento de pacotes. A RSRP é usada para decisões de seleção e transferência de células, enquanto a RSRQ fornece critérios para classificar células candidatas com base na qualidade do sinal. Já a RSSI representa a média linear da potência total recebida.

### 3.3. Importância dos Mapas de Calor

O estudo de redes móveis para a previsão de mapas de calor apresenta impacto significativo em diversas áreas da vida moderna. A melhoria na cobertura de rede proporciona maior consistência na experiência do usuário, reduzindo falhas e interrupções — fator crítico não apenas para o conforto do consumidor, mas também para aplicações sensíveis, como comunicações de emergência e telemedicina.

Além disso, a otimização da rede contribui para a eficiência energética, permitindo que operadoras ajustem dinamicamente seus recursos com base na demanda, reduzindo o consumo desnecessário de energia e promovendo práticas mais sustentáveis.

Dessa forma, este projeto possui um caráter interdisciplinar, agregando valor tanto à pesquisa acadêmica quanto ao desenvolvimento de soluções inovadoras para desafios tecnológicos emergentes.

## 4. Metodologia

Essa pesquisa se trata de uma pesquisa exploratória, pois busca-se entender e explorar o assunto a ser trabalhado, para finalmente abordarmos métodos e soluções, tal qual nossos objetivos. Para a concretização desta pesquisa, será utilizada a metodologia KDD (Knowledge Discovery in Databases) para o procedimento de aprendizagem contendo os processos de seleção, extração de dados e conhecimento de bases de dados robustas de redes 5G e 4G LTE. Portanto, as etapas que compõem a metodologia são:

**Revisão bibliográfica:** Essa etapa é crucial para identificar métodos existentes de aprendizado de máquina no contexto da pesquisa. Para isso, será realizada a busca por artigos científicos da área da pesquisa nas bases de dados escolhidas para realizar o mapeamento do Estado da Arte da área de pesquisa.

**Procedimento de aprendizagem implementando metodologia KDD:** O processo de KDD é um processo de extração de conhecimento em grandes bases de dados. Devido à metodologia, esta etapa é composta pelas fases de seleção de dados, pré-processamento, transformação, mineração e interpretação/avaliação.

- **Seleção de dados:** Será realizada a coleta de dados da base de dados para serem minerados. Os dados a serem coletados são as redes móveis localizadas em uma determinada posição no ambiente.
- **Pré-processamento:** Os dados coletados são organizados e as inconsistências serão tratadas.
- **Transformação:** Consiste na transformação dos dados brutos em dados transformados para aplicação da técnica de aprendizado de máquina.
- **Mineração:** Também conhecido como Data Mining, esta fase aplicará a técnica de aprendizado de máquina para extração do conhecimento.
- **Avaliação:** Nesta fase serão validados todo o conhecimento obtido da extração.

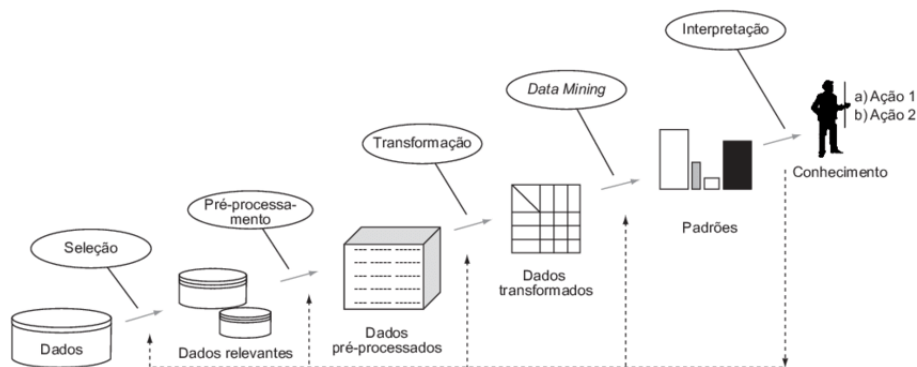
**Predição em mapa de calor:** Será realizado a predição dos conhecimentos adquiridos em mapas de calor para a melhor visualização da localização identificada das redes móveis.

Após o tratamento e transformação dos dados, pretende-se aplicar algoritmos supervisionados como K-Nearest Neighbors (KNN), Random Forest (RF) e Redes Neurais Convolucionais (CNNs), conforme identificado na revisão sistemática. A avaliação dos modelos será conduzida com métricas como acurácia, RMSE (Root Mean Square Error) e F1-score. A validação cruzada (k-fold) será utilizada para assegurar a generalização dos resultados, além de análise de importância de variáveis, quando aplicável.

## 5. Resultados e Discussão

### 5.1. Aprendizado de Máquina com Metodologia KDD

O processo seguiu a metodologia KDD (Knowledge Discovery in Databases), cujo é um fluxo contínuo de atividades que permite a extração e o compartilhamento de conhecimento a partir de bases de dados. De acordo com [Fayyad et al. 1996], esse processo é composto por cinco etapas principais: seleção dos dados, pré-processamento e limpeza, transformação, Mineração de Dados (Data Mining) e, por fim, interpretação e avaliação dos resultados. A relação entre essas etapas pode ser visualizada na Figura 1, onde as três primeiras são frequentemente associadas à análise exploratória dos dados.



**Figura 1. Resumo refatorado das etapas do processo KDD [Fayyad et al. 1996]**

### 5.1.1. Seleção dos Dados

A primeira etapa do processo KDD é a seleção dos dados, fundamental para garantir a qualidade e relevância das informações utilizadas na pesquisa. Seguindo essa metodologia, este estudo reuniu uma base de dados composta por medições reais da qualidade do sinal de redes móveis, coletadas diretamente de dispositivos móveis.

Para assegurar a consistência e confiabilidade dos registros, foram aplicados critérios de inclusão e exclusão, apresentado na Tabela 1, eliminando valores inconsistentes ou ausentes que pudessem comprometer a análise e os modelos preditivos.

Após a busca por bases de dados disponíveis em repositórios como Kaggle<sup>1</sup> e Google Datasets<sup>2</sup>, verificou-se que nenhuma delas atendia integralmente aos critérios estabelecidos. Diante disso, optou-se pela realização de uma coleta amostral em cidades do Amazonas, utilizando um dispositivo exclusivo para o experimento. A coleta contemplou medições de duas operadoras de rede distintas, permitindo uma maior variação dos dados e oferecendo uma visão mais abrangente da qualidade do sinal em diferentes condições.

O dispositivo utilizado registrou diversas informações relevantes, incluindo dados do sistema operacional, operadora de rede, qualidade do sinal e status da bateria. As medições abrangeram tanto redes 4G LTE quanto 5G NR, possibilitando uma análise detalhada do desempenho da rede em diferentes cenários. Para garantir a representatividade dos dados, as coletas foram realizadas em ambientes internos e externos, visando capturar variações na qualidade do sinal de acordo com o contexto da medição.

Durante a inspeção inicial dos dados, foram identificadas diversas inconsistências, como registros incompletos, valores fora dos padrões esperados e medições redundantes. Para garantir a integridade das informações, foi realizado um pré-processamento, eliminando ruídos e garantindo que os dados estivessem limpos e estruturados para as próximas etapas da análise.

<sup>1</sup><https://www.kaggle.com/>

<sup>2</sup><https://datasetsearch.research.google.com/>

**Tabela 1. Critérios de inclusão (CI) e exclusão (CE) para os filtros de dados**

Critério	Descrição
CI.1	Os dados foram coletados de dispositivos móveis conectados ativamente a redes 4G LTE ou 5G NR.
CI.2	As coordenadas geográficas (latitude e longitude) são válidas e permitem a geolocalização precisa dos pontos de coleta.
CI.3	Os parâmetros de qualidade do sinal (ex.: RSRP, RSRQ, SINR, CQI) apresentam valores dentro dos intervalos esperados para medições reais.
CI.4	Os dados devem ser oriundos do estado do Amazonas.
CI.5	Os dados incluem medições de ambientes internos e externos.
CE.1	Não atende a nenhum critério de inclusão.
CE.2	Os registros estão incompletos, com variáveis essenciais ausentes ou corrompidas.

### 5.1.2. Pré-Processamento

Como mencionado anteriormente, a etapa de pré-processamento foi essencial para a limpeza dos dados, tratamento de valores ausentes, remoção de ruídos e anomalias, além da normalização. Com o primeiro passo consistindo na remoção de colunas corrompidas ou irrelevantes para a análise, seguidos do tratamento de valores ausentes e da conversão dos tipos de dados para garantir compatibilidade nas etapas seguintes. O código do Listing 1 apresenta esse procedimento.

Todo esse processo foi feito usando Python 3.10, utilizando as bibliotecas Pandas e NumPy, por meio de um script de tratamento. Esse processo foi realizado utilizando Python 3.10 e um script de tratamento de dados, que extrai, limpa e processa um arquivo CSV contendo informações de dispositivos móveis. Como ilustrado no listing 1, colunas contendo dados repetitivos ou inválidos foram removidas, enquanto linhas com grande quantidade de valores ausentes foram descartadas. Além disso, colunas numéricas foram convertidas para o tipo float, e a coluna de data/hora foi padronizada para o formato datetime.

**Listing 1. Script de limpeza da base de dados**

```
# Retirar colunas corrompidas
base.drop(['Group', 'networkOperator', 'chipOperator', 'cqiLTE', 'earfcnNR'],
          axis=1, inplace=True)

# Retirar linhas com dados em branco
limite = len(base.columns) // 2
base.dropna(thresh=len(base.columns) - limite, inplace=True)

# Converter colunas para o tipos apropriados
for coluna in colunas_float:
    base[coluna] = pd.to_numeric(base[coluna], errors='coerce')
base['currentDateTime'] = pd.to_datetime(base['currentDateTime'])
```

Para garantir a qualidade do conjunto de dados, foi necessário um tratamento adicional de anomalias, conforme ilustrado no código de Listing 2. Neste estágio, valores anômalos foram identificados e substituídos por NaN, enquanto registros de dispositivos offline receberam a marcação "offline" na coluna networkType. Para lidar com valores ausentes remanescentes, aplicamos interpolação linear, garantindo maior consistência ao conjunto de dados.

**Listing 2. Script de tratamento de anomalias e normalização de dados**

```
# Substituindo dados estranhos por nan
base.replace(dados_estranhos, np.nan, inplace=True)
base['networkType'] = base['networkType'].replace('2147483647', value=np.nan)

# Substituindo em 0.0 para 'offline'
base = base.apply(lambda linha: linha.fillna(0.0) if linha.isna().sum() == offline
                  else linha, axis=1)
base['networkType'] = base['networkType'].replace(0.0, value='offline')
```

```
# Substituindo pela interpolacao
for coluna in columnas_float:
    base[coluna] = base[coluna].fillna(round(base[coluna].interpolate(method='linear'), 1))
```

Ao término do pré-processamento, os dados foram armazenados em um novo arquivo CSV, já limpos e estruturados. Em média, o tamanho do novo arquivo tratado foi reduzido em aproximadamente 37,4% em relação ao original. Com isso, os dados estão prontos para a etapa de transformação dentro do fluxo da metodologia KDD.

## 5.2. Predição em mapa de calor

A predição dos mapas de calor já foi iniciada, ainda que de forma parcial. Até o momento, os dados tratados já podem ser visualizados por meio de um protótipo de site, desenvolvido como uma aplicação web básica utilizando Python, HTML e CSS puro como demonstra a Figura 2. Essa ferramenta tem como principal objetivo permitir a exploração do comportamento dos dados coletados, oferecendo uma representação visual da qualidade do sinal das redes móveis analisadas em diferentes ambientes.

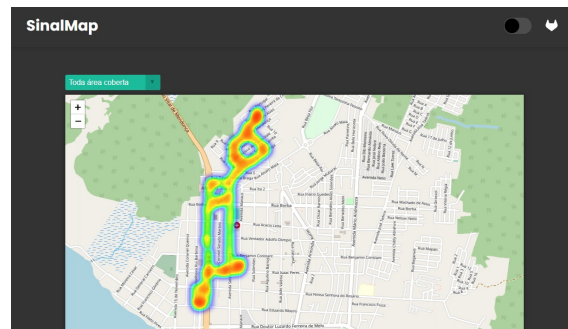


Figura 2. Site visualizador de mapa de calor

Atualmente, a aplicação possui funcionalidades essenciais, como a exibição do mapa de calor gerado a partir dos dados processados, a visualização de uma amostra dos dados de qualidade do sinal e um filtro para selecionar a cidade desejada. Isso possibilita aos usuários uma análise inicial das condições de conectividade tanto em ambientes internos quanto externos. A Figura 3 exemplifica melhor como está sendo exibido o mapa de calor.

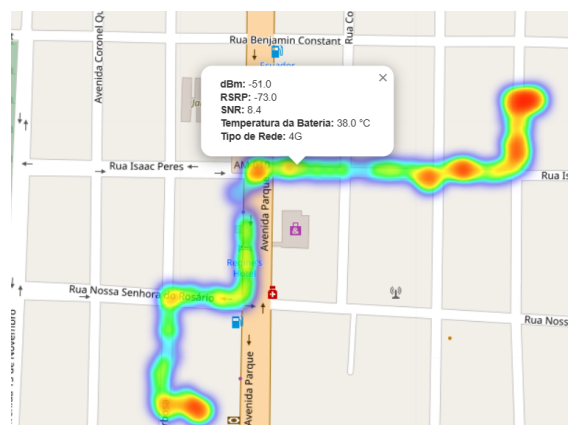


Figura 3. Mapa de calor exibindo amostra de dados em forma de pop-up

Ainda que a fase preditiva esteja em andamento, os dados tratados já permitiram a construção de mapas de calor preliminares. A visualização revela padrões de degradação do sinal em áreas urbanas mais densas, especialmente em ambientes internos com maior interferência. Essas evidências reforçam a necessidade de aplicar técnicas preditivas que considerem fatores geográficos e variáveis de rede para inferência de cobertura futura.

## **6. Aplicações e Impacto Prático**

A ferramenta proposta possui aplicabilidade direta para diversos setores. Operadoras de telefonia podem utilizá-la para otimizar a infraestrutura de antenas; órgãos reguladores, para fiscalizar a qualidade da cobertura; e desenvolvedores, para garantir a estabilidade de aplicativos que dependem de conectividade contínua.

Além das aplicações imediatas, o estudo oferece uma contribuição de grande valor com a coleta de dados inéditos no Amazonas. Essa base de dados pode ser o ponto de partida para expandir a análise a outras regiões com baixa cobertura ou topografia complexa. Futuramente, o sistema poderá evoluir para oferecer APIs públicas, permitindo sua integração com plataformas de cidades inteligentes e soluções de planejamento urbano.

## **7. Considerações Finais**

Os próximos passos incluem a finalização da transformação dos dados, seguida da fase de mineração de dados, onde serão aplicadas técnicas de aprendizado de máquina para extração de conhecimento. Além disso, será realizada a avaliação dos modelos gerados, permitindo aprimorar a predição dos mapas de calor e validar a precisão dos resultados obtidos.

No futuro, pretende-se aprimorar a aplicação com o uso do framework Django, explorando todo o ecossistema de desenvolvimento web em Python. Além disso, será implementada uma abordagem baseada em técnicas de Machine Learning para realizar a predição da qualidade do sinal em diferentes cenários. O modelo de aprendizado de máquina será definido a partir da abstração dos artigos selecionados na revisão sistemática.

Outra funcionalidade planejada é a possibilidade de carregar novos arquivos CSV contendo dados de dispositivos móveis, permitindo que os usuários visualizem essas informações diretamente na página e realizem predições utilizando os modelos de Machine Learning integrados ao sistema. Isso tornará a ferramenta mais dinâmica e flexível, facilitando a análise e a tomada de decisões em relação à qualidade do sinal das redes móveis.

## **Referências**

- Afroz, F., Subramanian, R., Heidary, R., Sandrasegaran, K., and Ahmed, S. (2015). Sinr, rsrp, rssi and rsrq measurements in long term evolution networks. *International Journal of Wireless & Mobile Networks*.
- Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P., and Uthurusamy, R. (1996). Advances in knowledge discovery and data mining. American Association for Artificial Intelligence.



- Haskić, H. and Radončić, A. (2024). The effects of 5g network on people and the environment: A machine learning approach to the comprehensive analysis. *World Journal of Advanced Engineering Technology and Sciences*, 11(1):301–309.
- Kaur, J., Khan, M. A., Iftikhar, M., Imran, M., and Haq, Q. E. U. (2021). Machine learning techniques for 5g and beyond. *IEEE Access*, 9:23472–23488.
- Mogyorósi, F., Revisnyei, P., Pašić, A., Papp, Z., Törös, I., Varga, P., and Pašić, A. (2022). Positioning in 5g and 6g networks—a survey. *Sensors*, 22(13):4757.
- Morocho-Cayamcela, M. E., Lee, H., and Lim, W. (2019). Machine learning for 5g/b5g mobile and wireless communications: Potential, limitations, and future directions. *IEEE access*, 7:137184–137206.
- Saari, O. (2022). Heatmap reporting tool: System testing.
- Sarkar, S. and Debnath, A. (2021). Machine learning for 5g and beyond: Applications and future directions. In *2021 Second International Conference on Electronics and Sustainable Communication Systems (ICESC)*, pages 1688–1693. IEEE.
- Zhang, C., Patras, P., and Haddadi, H. (2019). Deep learning in mobile and wireless networking: A survey. *IEEE Communications surveys & tutorials*, 21(3):2224–2287.