



UNIVERSIDADE FEDERAL DO AMAZONAS  
FACULDADE DE TECNOLOGIA  
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

# LOCALIZAÇÃO EM AMBIENTES INTERNOS UTILIZANDO REDES IEEE 802.11

David Alan de Oliveira Ferreira

Manaus – Amazonas

Março de 2019

David Alan de Oliveira Ferreira

LOCALIZAÇÃO EM AMBIENTES INTERNOS  
UTILIZANDO REDES IEEE 802.11

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal do Amazonas, como requisito parcial para obtenção do título de Mestre em Engenharia Elétrica na área de concentração Controle e Automação de Sistemas.

Orientador: Prof. Dr. Celso Barbosa Carvalho

## Ficha Catalográfica

Ficha catalográfica elaborada automaticamente de acordo com os dados fornecidos pelo(a) autor(a).

F383I Ferreira, David Alan de Oliveira  
Localização em Ambientes Internos Utilizando Redes IEEE  
802.11 / David Alan de Oliveira Ferreira. 2019  
61 f.: il. color; 31 cm.

Orientador: Celso Barbosa Carvalho  
Dissertação (Mestrado em Engenharia Elétrica) - Universidade  
Federal do Amazonas.

1. Localização interna. 2. Algoritmo kNN. 3. Wi-Fi. 4. Análise de  
quartis. I. Carvalho, Celso Barbosa II. Universidade Federal do  
Amazonas III. Título

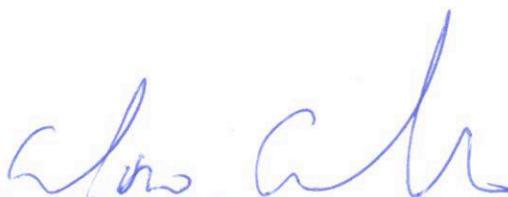
DAVID ALAN DE OLIVEIRA FERREIRA

**“LOCALIZAÇÃO EM AMBIENTES INTERNOS UTILIZANDO REDES IEEE 802.11”**

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal do Amazonas, como requisito parcial para obtenção do título de Mestre em Engenharia Elétrica na área de concentração Controle e Automação de Sistemas.

Aprovado em 27 de março de 2019.

BANCA EXAMINADORA



Prof. Dr. Celso Barbosa Carvalho

Universidade Federal do Amazonas



Prof. Dr. Thiago Brito Bezerra

Universidade Federal do Amazonas



Prof. Dr. Florindo Antonio de Carvalho Ayres Júnior

Universidade Federal do Amazonas

# Agradecimentos

Em primeiro lugar ao meu Deus, por ter me dado a vida e por me fortalecer para enfrentar os desafios.

Aos meus queridos pais, Ednaldo e Marta.

À minha esposa Carla.

Ao meu orientador Celso Carvalho, pela orientação e incentivo, fazendo com que este trabalho pudesse ser realizado.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes), pelo apoio financeiro.

Finalmente, a todos aqueles que tiveram contribuição direta ou indireta para a realização deste trabalho.

Resumo da Dissertação apresentada à UFAM como parte dos requisitos necessários para a obtenção do grau de Mestre em Engenharia Elétrica

# Localização em Ambientes Internos Utilizando Redes IEEE 802.11

David Alan de Oliveira Ferreira

Orientador: Celso Barbosa Carvalho

Programa: Pós-Graduação em Engenharia Elétrica

Este trabalho propõe um método que emprega o algoritmo de aprendizado de máquina *k-Nearest Neighbors* (*k*NN) para determinar a localização de objetos móveis em ambientes internos. No cenário de testes, o objeto móvel é representado por uma estação sem fio (*Wireless Station - WSTA*) que utiliza tecnologia Wi-Fi (*Wireless Fidelity*). Para estimar a localização da WSTA realizaram-se medições do Indicador de Intensidade do Sinal Recebido (*Received Signal Strength Indicator - RSSI*), dos sinais provenientes de pontos de acesso (*Access Points - APs*), a partir de pontos de coleta específicos denotados como pontos de referência (*Reference Points - RPs*). Neste cenário, em uma fase inicial de treinamento do algoritmo, cada RP é utilizado para coletar amostras de RSSI em um processo de varredura dos APs instalados no ambiente. Ainda na fase de treinamento, utilizam-se medidas de quartis para representar o comportamento destas amostras de RSSI. Posteriormente, na fase de testes, os dados do conjunto de treinamento, formado pelos quartis, são comparados com novos dados a fim de determinar a posição da WSTA. Na avaliação de desempenho, verificou-se que o algoritmo proposto possuiu erro nulo com apenas quatro APs e 10 leituras por amostras com 17,27 segundos de tempo de processamento. Verifica-se que os resultados com estes valores são contribuições importantes, o que assegura que utilizar o algoritmo *k*NN adotando um conjunto de dados sumarizado com medidas de quartis, é um método

promissor para localizar objetos em ambientes internos.

Palavras-chave: localização interna, algoritmo  $k$ NN, Wi-Fi, análise de quartis.

Abstract of Dissertation presented to UFAM as a partial fulfillment of the requirements for the degree of Master in Electrical Engineering

# Localização em Ambientes Internos Utilizando Redes IEEE 802.11

David Alan de Oliveira Ferreira

Advisor: Celso Barbosa Carvalho

Department: Postgraduate in Electrical Engineering

This work proposes a method that employs the  $k$ -Nearest Neighbors ( $k$ NN) machine learning algorithm to determine the location of moving objects indoors. In the test scenario, the mobile object is represented by a Wireless Station (WSTA) that uses Wi-Fi (*Wireless Fidelity*) technology. In order to estimate the location of the WSTA, measurements were made of the Received Signal Strength Indicator (RSSI), from signals from access points (APs), from specific collection points denoted as points reference points (RPs). In this scenario, in an initial phase of training the algorithm, each RP is used to collect RSSI samples in a process of scanning APs installed in the environment. Also in the training phase, quartiles measurements are used to represent the behavior of these RSSI samples. Subsequently, in the test phase, the training set data, formed by the quartiles, are compared with new data in order to determine the position of the WSTA. In the performance evaluation, it was verified that the proposed algorithm had null error with only four APs e 10 readings per sample with 17.27 seconds of processing time. It is verified that the results with these values are important contributions, which ensures that using the  $k$ NN algorithm adopting a dataset summarized with quartiles measurements is a promising method to locate objects indoors.

Keywords: indoor location,  $k$ NN algorithm, Wi-Fi, quartile analysis.

# Lista de figuras

Figura 1. Dinâmica do 3NN. ....	21
Figura 2. Modos de operação das redes IEEE 802.11. a) <i>Ad Hoc</i> e b) Infra-estruturado. ....	23
Figura 3. <i>Layout</i> do cenário de experimentação. ....	32
Figura 4. Diagrama de conexões dos elementos que compõem o sistema de localização. ....	33
Figura 5. Posicionamento dos pontos de acesso (APs) e dos pontos de referência (RPs) no ambiente interno. ....	34
Figura 6. Comparação entre as técnicas de sombreamento log-normal, aproximação quadrática e RSSI médio para obtenção da intensidade do sinal recebido (RSSI) no ponto de referência 1 (RP1). ....	37
Figura 7. Comparação entre as técnicas de sombreamento log-normal, aproximação quadrática e RSSI médio para obtenção da intensidade do sinal recebido (RSSI) no ponto de referência 2 (RP2). ....	38
Figura 8. Leituras da intensidade do sinal recebido (RSSI) do ponto de acesso 2 (AP2) coletadas nos pontos de referência 1 e 2 (RP1 e RP2). ....	40
Figura 9. Diagrama de fluxo comum aos métodos propostos. ....	43
Figura 10. Função de Distribuição Acumulada (CDF) dos erros de localização (m). ....	50
Figura 11. Erro médio das 160 estimativas de localização em função do parâmetro $k$ para cada número de pontos de acesso ( $n$ ). ....	51
Figura 12. Erro médio de localização com número reduzido de leituras de RSSI. ....	53

# Lista de tabelas

Tabela 1. Conjunto de dados e nova instância. ....	20
Tabela 2. Similaridade.....	20
Tabela 3. Características dos padrões IEEE 802.11. ....	22
Tabela 4. Valores obtidos da análise de quartis do exemplo descrito. ....	26
Tabela 5. Características dos trabalhos relacionados. ....	31
Tabela 6. Coordenadas dos pontos de acesso (APs) e dos pontos de referência (RPs) no ambiente interno.....	35
Tabela 7. Quartis do ponto de acesso 2 (AP2) com a intensidade do sinal recebido (RSSI) medida nos pontos de referência 1 e 2 (RP1 e RP2). ....	40
Tabela 8. Valores iniciais, finais e incrementos das variáveis mensuradas.....	46
Tabela 9. Características dos métodos propostos e dos métodos apresentados na literatura. ....	48
Tabela 10. Tempo de processamento para cada número de leituras de RSSI.....	53

# Lista de abreviaturas

**APs** – *Access Points*

**CDF** – *Cumulative Distribution Function*

**GPS** – *Global Positioning System*

**IDLE** – *Python's Integrated Development and Learning Environment*

**IoT** – *Internet of Things*

**IPS** – *Indoor Positioning System*

**kNN** – *k-Nearest Neighbors*

**ML** – *Machine Learning*

**MN** – *Mobile Node*

**PC** – *Personal Computer*

**PCA** – *Principal Component Analysis*

**RP**s – *Reference Points*

**RSSI** – *Received Signal Strength Indicator*

**SSID** – *Service Set Identifier*

**STA** – *Station*

**SVM** – *Support Vector Machine*

**Wi-Fi** – *Wireless Fidelity*

**WSTA** – *Wireless Station*

# Sumário

<b>1 Introdução</b> .....	<b>14</b>
1.1 Objetivo geral.....	16
1.2 Objetivos específicos .....	16
1.3 Organização do trabalho .....	16
<b>2 Fundamentação teórica</b> .....	<b>18</b>
2.1 Algoritmo <i>k-Nearest Neighbors</i> .....	18
2.2 Redes sem fio IEEE 802.11 .....	22
2.2.1 Modos de operação .....	23
2.3 Análise de quartis .....	23
2.4 Trabalhos relacionados .....	26
<b>3 Material e métodos</b> .....	<b>32</b>
3.1 Cenário de experimentação .....	32
3.2 Experimentos para a relação entre distância e RSSI.....	36
3.3 Propostas para as estimativas de localização.....	39
3.3.1 Representação dos dados utilizando quartis .....	39
3.3.2 Localização utilizando as coordenadas do ponto de referência majoritário .....	41
3.3.3 Localização utilizando as coordenadas do centróide dos pontos de referência.....	42
3.4 Experimentos para as estimativas de localização .....	43
3.4.1 Fase de treinamento.....	44
3.4.2 Fase de testes .....	45
3.4.3 Medidas de avaliação de desempenho.....	48
<b>4 Resultados e discussão</b> .....	<b>50</b>

4.1 Função de distribuição acumulada do erro médio de localização .....	50
4.2 Erro médio de localização por número de APs e valor do parâmetro $k$ .....	51
4.3 Tempo médio de processamento para localização .....	52
<b>5 Conclusão e trabalhos futuros .....</b>	<b>55</b>
<b>Referências .....</b>	<b>57</b>

# 1 Introdução

O conceito de Internet das Coisas (*Internet of Things* - IoT) surgiu com a proposta de tornar dispositivos antes meramente tradicionais em dispositivos inteligentes, que podem tomar decisões na realização de tarefas e que são capazes de informar parâmetros de seu funcionamento. Desta maneira, os dispositivos podem ser analisados no contexto de suas posições no tempo e no espaço (Al-Fuqaha et al., 2015). Embora o Sistema de Posicionamento Global (*Global Positioning System* - GPS) forneça dados aceitáveis da localização física em ambiente externo, sua precisão e utilidade ficam comprometidas à medida que o espaço habitado pelos objetos diminui, principalmente em ambiente interno (Salim et al., 2014).

Existem vários métodos propostos na literatura que podem ser utilizados para a localização de objetos em ambientes internos. Muitas destas propostas são baseadas em redes Wi-Fi (*Wireless Fidelity*), devido à grande presença dessas redes em locais públicos e em locais privados, e utilizam o Indicador de Intensidade do Sinal Recebido (*Received Signal Strength Indicator* - RSSI) como parâmetros de localização (Le et al., 2014). Os sistemas de localização podem ser baseados em técnicas geométricas como a trilateração (Bahl e Padmanabhan, 2000), que relaciona o RSSI e a distância para formar três círculos com raios de tamanho variável e encontra um ponto em comum; ou ser baseados em técnicas que utilizam padrões de RSSI como o *fingerprint*, que compara leituras previamente coletadas de RSSI a uma nova leitura e identifica a posição com o valor de RSSI correspondente.

Em ambientes internos com múltiplos obstáculos, os sistemas tradicionais apresentam dificuldades em localizar objetos, devido aos problemas de interferência de sinal e instabilidade das medições brutas, causados principalmente pelos obstáculos (Kaemarungsi e Krishnamurthy, 2004; Ghourchian et al., 2017), porém a escolha de técnicas mais adequadas pode produzir resultados com melhor precisão. Técnicas que

geram resultados de localização com alta acurácia e confiabilidade permitem implementar sistemas de localização mais precisos e que podem auxiliar nas aplicações e serviços de diversas áreas. Essas aplicações incluem assistência médica, onde possibilita a localização dos médicos, para um melhor atendimento ao paciente; rastreamento de bens (ex.: mercadorias, equipamento hospitalar, componentes de grandes dimensões, etc.); e *marketing*, pois auxilia na exibição de anúncios e ofertas com base na localização do usuário; entre outras.

Novas abordagens estão sendo propostas para reduzir os erros de estimativa desses sistemas, como por exemplo, a utilização de métodos baseados em algoritmos de aprendizado de máquina (*Machine Learning* - ML) para a localização de usuários e de objetos móveis (Wu et al., 2015; Cai et al., 2015; Khullar e Dong, 2017; Kim et al., 2017; Labinghisa e Lee, 2018). Pesquisas recentes empenham-se principalmente em determinar técnicas para a construção de conjuntos de dados, que sejam adequados para algoritmos de aprendizado de máquina, como os usados para normalizar e representar dados brutos, identificar e remover dados redundantes, além de modelos matemáticos para obtenção dos dados. Contudo, muitas pesquisas tratam as leituras de RSSI de forma generalizada e não consideram aspectos comportamentais, tais como o grau de espalhamento dos dados presentes em conjuntos afetados por diferentes obstáculos. A omissão destes aspectos pode resultar em estimativas de localização discordantes com as posições reais.

Portanto, o objetivo deste trabalho foi projetar um método de localização em ambientes internos baseado em aprendizado de máquina, utilizando atributos com informações específicas que caracterizam o comportamento do RSSI, visando garantir uma melhor influência nos resultados de localização. Diante do exposto, esta dissertação contribuirá de forma significativa na melhoria da acurácia por meio de um método promissor para gerar a localização em ambientes internos.

## 1.1 Objetivo geral

O objetivo deste trabalho é projetar um método de localização em ambientes internos baseado em aprendizado de máquina capaz de melhorar a precisão de localização, suprindo as limitações dos métodos presentes na literatura, utilizando características dos sinais das redes IEEE 802.11 a fim de oportunizar a implementação do método proposto em ambientes com diferentes obstáculos.

## 1.2 Objetivos específicos

- Identificar um algoritmo de aprendizado de máquina adequado aos métodos de localização.
- Projetar uma plataforma de testes utilizando dispositivos de baixo custo.
- Executar testes práticos analisando os padrões de comportamento do indicador de intensidade do sinal recebido (RSSI) em ambientes internos.
- Executar testes práticos implementando métodos de localização presentes na literatura.
- Projetar e testar o método proposto utilizando características dos sinais das redes Wi-Fi.
- Validar o desempenho do método proposto.

## 1.3 Organização do trabalho

Esta dissertação apresenta-se dividida em cinco capítulos.

No **Capítulo 2**, apresentamos os fundamentos teóricos do algoritmo *k-Nearest Neighbors* (*k*NN), das redes sem fio IEEE 802.11 e da análise de quartis, além de métodos de localização propostos na literatura com temas relacionados à proposta deste trabalho.

No **Capítulo 3**, apresentamos o cenário de experimentação para avaliação da proposta de trabalho, uma verificação da relação entre distância e RSSI, bem como o procedimento proposto para localização em ambientes internos.

No **Capítulo 4**, discutimos os resultados experimentais do método proposto. Neste caso, comparamos os resultados com métodos existentes na literatura.

O **Capítulo 5** apresenta as conclusões gerais da dissertação e trabalhos futuros.

## 2 Fundamentação teórica

Nesta seção, são apresentados os conceitos básicos relacionados ao algoritmo *k-Nearest Neighbors* (*k*NN), às redes sem fio IEEE 802.11 e à análise de quartis que foram utilizados durante o desenvolvimento deste trabalho. E por fim, são apresentados artigos com temas relacionados à proposta deste trabalho e que podem oferecer contribuições relevantes para esta pesquisa.

### 2.1 Algoritmo *k-Nearest Neighbors*

O algoritmo dos *k* vizinhos mais próximos (*k-Nearest Neighbors* – *k*NN) foi proposto por Cover e Hart em 1967. Este é um dos métodos de classificação mais simples e eficazes dentre os algoritmos de aprendizado de máquina, e é um algoritmo não-paramétrico, ou seja, não necessita que dados apresentem distribuição específica (ex.: gaussiana ou exponencial).

Este algoritmo de classificação insere-se no âmbito da aprendizagem supervisionada, no qual armazena um conjunto de dados previamente classificados, podendo induzir a cálculos dispendiosos se o conjunto de dados for grande (Bishop, 2006). No entanto, para uma aplicação com modelo de representação de dados mais adequado à classificação, este método é bastante eficiente. Segundo Lantz (2015), este algoritmo foi utilizado com sucesso para:

- Aplicações de visão computacional, incluindo reconhecimento óptico de caracteres e reconhecimento facial em imagens fixas e vídeo.
- Prever se uma pessoa irá gostar de um filme ou de uma recomendação musical.
- Identificar padrões em dados genéticos, e utilizá-los na detecção de doenças.

O conjunto de dados pré-classificados é denominado conjunto de treinamento, onde as linhas são vetores *n*-dimensionais que representam

instâncias de treinamento com  $n$  atributos. Para estimar a classe de uma nova instância (denominada instância de teste), o algoritmo  $k$ NN tradicionalmente executa os seguintes procedimentos (Cover e Hart, 1967):

1. Calcula a similaridade entre a instância de teste e cada instância de treinamento.
2. Seleciona as  $k$  instâncias de treinamento mais similares (próximas) à instância de teste.
3. Verifica a classe majoritária, ou seja, a classe que aparece com maior frequência dentre as  $k$  instâncias selecionadas.
4. Atribui a instância de teste à classe majoritária.

A similaridade é calculada através de uma métrica de distância que é aplicada entre dois vetores. Podem-se definir várias métricas, tais como as distâncias Euclidiana, Manhattan e Minkowski, sendo comumente adotada na literatura, em geral, a distância euclidiana definida pela Equação 3.1:

$$dist(u, v) = \sqrt{\sum_{i=1}^n (u_i - v_i)^2} \quad (3.1),$$

em que  $u$  e  $v$  são vetores  $n$ -dimensionais; e  $u_i$  e  $v_i$  representam os valores do atributo  $i$  dos respectivos vetores.

O valor ideal do parâmetro  $k$  (número de vizinhos mais próximos) é determinado experimentalmente, para cada base de dados, onde o valor de melhor precisão é escolhido para definir o classificador. Pode-se testar com  $k$  variando de um até o valor da raiz quadrada do número de instâncias de treinamento. Devendo  $k$  possuir um valor ímpar para evitar possíveis empates entre duas classes (Lantz, 2015).

Para exemplificar o algoritmo  $k$ NN, na Tabela 1 são apresentados um conjunto de dados classificados e uma nova instância a ser classificada (Lantz, 2015):

**Tabela 1.** Conjunto de dados e nova instância.

Ingrediente	Doçura	Crocância	Tipo de comida
Conjunto de dados			
Bacon	1	4	Proteína
Banana	10	1	Fruta
Cenoura	7	10	Vegetal
Salsão	3	10	Vegetal
Queijo	1	1	Proteína
Uva	8	5	Fruta
Feijão verde	3	7	Vegetal
Nozes	3	6	Proteína
Laranja	7	3	Fruta
Nova instância			
Tomate	6	4	-

Para responder a questão “O tomate é uma fruta ou vegetal?”, baseado no algoritmo  $k$ NN e nas notas de um a 10 para as características “Doçura” e “Crocância”, calcula-se a similaridade entre o tomate e todos os ingredientes do conjunto de dados. Por exemplo, para calcular a similaridade entre o tomate e o queijo, pode-se usar a distância euclidiana (Equação 3.1) da seguinte forma:

$$dist(\text{tomate}, \text{queijo}) = \sqrt{(6 - 1)^2 + (4 - 1)^2} = 5,8310 \quad (3.2)$$

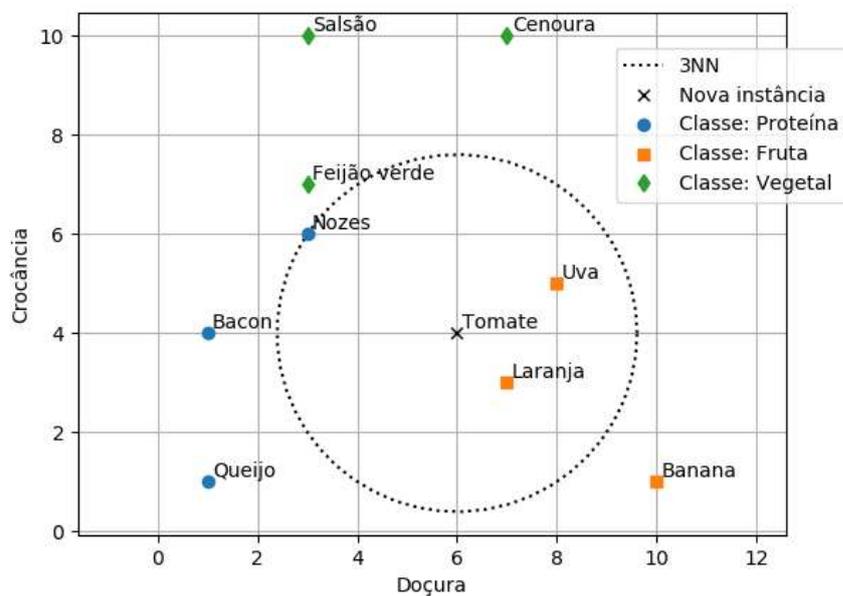
As distâncias calculadas entre o tomate e todos os ingredientes do conjunto de dados são apresentadas na Tabela 2:

**Tabela 2.** Similaridade.

Ingrediente	Tipo de comida	Distância ao Tomate
Bacon	Proteína	5,0000
Banana	Fruta	5,0000
Cenoura	Vegetal	6,0828
Salsão	Vegetal	6,7082
Queijo	Proteína	5,8310
Uva	Fruta	2,2361
Feijão verde	Vegetal	4,2426
Noz	Proteína	3,6056
Laranja	Fruta	1,4142

Por fim, usando  $k = 1$  (1-vizinho mais próximo) o algoritmo classifica o tomate como uma “Fruta”, pois a distância entre o tomate e a laranja é menor; e aplicando  $k = 3$  (3-vizinhos mais próximos) o algoritmo classifica novamente o tomate como uma “Fruta”, pois a laranja e a uva são “Frutas”, e a noz é “Proteína”, então a maioria desses vizinhos são “Frutas”.

A Figura 1 apresenta a dinâmica do classificador para 3-vizinhos mais próximos (3NN). Como o conjunto de dados possui apenas dois atributos (bidimensional), pode-se plotar os dados em um gráfico de dispersão, com o eixo x indicando a doçura do ingrediente e o eixo y, a crocância.



**Figura 1.** Dinâmica do 3NN.

O algoritmo  $k$ NN é implementado em diversas plataformas, tais como o *software* MATLAB e o *framework* Scikit-learn para a linguagem Python. Porém, para tornar o código mais compacto, no tocante a quantidade de instruções, e adaptável para suprir as particularidades de duas propostas de localização desenvolvidas nesta dissertação, optou-se por uma codificação própria do algoritmo  $k$ NN em linguagem Python.

## 2.2 Redes sem fio IEEE 802.11

As redes sem fio IEEE 802.11, também conhecidas como redes Wi-Fi (*Wireless Fidelity*), são uma alternativa às redes Ethernet cabeadas, fornecendo as mesmas funcionalidades destas de maneira flexível e móvel, facilitando a conectividade de mais dispositivos eletrônicos em ambientes internos (Silva, 1998).

O padrão IEEE 802.11 é uma série de especificações a serem adotadas no nível da camada física (PHY) e na subcamada de controle de acesso ao meio (*Media Access Control* – MAC) do modelo OSI, a fim de garantir atendimento às exigências estabelecidas para as transmissões sem fio. A interface para as camadas superiores da arquitetura é a mesma das redes cabeadas, mantendo assim a interoperabilidade (Deng et al., 2004).

Após anos de desenvolvimento, em 1997 foi lançada a primeira versão do IEEE 802.11 pelo *Institute of Electrical and Electronics Engineers* (IEEE). Originalmente, o padrão IEEE 802.11 operava com FHSS (*Frequency Hoping Spread Spectrum*) e DSSS (*Direct Sequence Spread Spectrum*) e suportava taxas de dados de 1 e 2 Mbps a banda de 2,4 GHz na camada física. Ao longo dos anos, variantes do padrão original foram oficializadas, tais como o 802.11g, que foram criados a partir de novas tecnologias de multiplexação para garantir maiores taxas de transmissão (Ni, 2005). Algumas características dos padrões mais utilizados são apresentadas na Tabela 3.

**Tabela 3.** Características dos padrões IEEE 802.11.

Padrão IEEE	802.11	802.11b	802.11a	802.11g	802.11n
Tecnologia	DSSS	DSSS/CKK	OFDM	OFDM/ DSSS/CKK	SDM/OFDM
Taxa de dados	1 e 2 Mbps	5,5 e 11 Mbps	6 até 54 Mbps	1 até 54 Mbps	6 até 600 Mbps
Frequência	2,4 GHz	2,4 GHz	5 GHz	2,4 GHz	2,4 e 5 GHz
Largura do canal	25 MHz	25 MHz	20 MHz	25 MHz	20 e 40 MHz

Além destes, outros padrões têm surgido com melhorias nos serviços de comunicação sem fio, tais como o IEEE 802.11c, d, e, f e i.

### 2.2.1 Modos de operação

Existem dois modos de operação para as redes IEEE 802.11, o modo infra-estruturado e o *ad hoc*. O modo infra-estruturado é o mais utilizado em redes 802.11, e caracteriza-se pela presença de um ponto de acesso (*Access Point* – AP) como nó central para a coordenação e roteamento de pacotes na rede. Desta forma, toda a comunicação entre as estações (ex.: *notebooks*) passa pelo AP (Chen et al., 2003).

O modo *ad hoc* caracteriza-se por operar de forma descentralizada, onde as estações (*Station* – STA) comunicam-se diretamente entre si. Neste modo, as estações também podem realizar o roteamento de pacotes para conectar aquelas que se encontram fora de alcance, através de múltiplos saltos (*multi-hop*). A Figura 2 apresenta uma ilustração destes modos de operação (Chen et al., 2003).

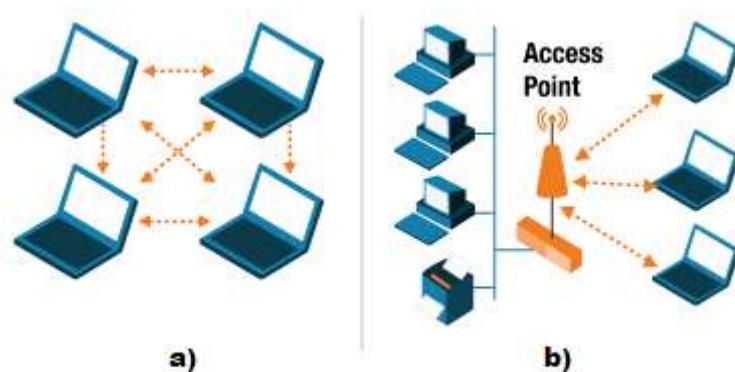


Foto: <https://www.element14.com/community/docs/DOC-90844/element14-essentials-wireless-protocol-ii-wi-fi>

**Figura 2.** Modos de operação das redes IEEE 802.11. a) *Ad Hoc* e b) Infra-estruturado.

## 2.3 Análise de quartis

Os quartis são valores que separam as observações de uma amostra ordenada em quatro segmentos iguais, ou seja, cada segmento com o mesmo número de observações. A partir desta análise, pode-se avaliar a tendência

central e a dispersão de um conjunto de dados (Joarder e Firozzaman, 2001).

Conforme Joarder e Firozzaman (2001), o primeiro quartil (Q1) ou quartil inferior corresponde ao valor que delimita as 25% menores observações, o segundo quartil (Q2) ou mediana separa as 50% menores das 50% maiores observações e o terceiro quartil (Q3) ou quartil superior é o valor que separa as 25% maiores observações. A diferença entre o terceiro e o primeiro quartil determina o intervalo interquartil (IIQ). Este é uma medida de dispersão alternativa ao desvio padrão e é dado por:

$$IIQ = Q3 - Q1 \quad (3.3)$$

Existem várias definições na literatura para calcular os valores dos quartis inferior e superior de uma amostra. Entre estas definições, o método de Tukey é o mais antigo e popular, e baseia-se em encontrar primeiro a mediana do conjunto de dados e depois encontrar a mediana da metade inferior e da metade superior dos dados. Definições diferentes também são usadas por softwares e pacotes de computação estatística (ex.: Microsoft Excel, MATLAB e NumPy) (Langford, 2006).

O NumPy é o pacote fundamental para computação científica com a linguagem Python (NumPy, 2019). Para encontrar os valores dos quartis de uma amostra com  $n$  observações ordenadas  $X = \{x_1 \leq \dots \leq x_n\}$ , o NumPy calcula os índices  $l$ ,  $m$  e  $u$ , onde o elemento  $x_l$  é o valor do quartil inferior,  $x_m$  é o valor da mediana e  $x_u$  é o valor do quartil superior.

$$l = \frac{n + 3}{4} \quad (3.4)$$

$$m = \frac{n + 1}{2} \quad (3.5)$$

$$u = \frac{3n + 1}{4} \quad (3.6)$$

Se  $l$ ,  $m$  e  $u$  não forem valores inteiros, o NumPy usa interpolação linear para calcular os valores dos quartis, conforme a seguinte equação:

$$x_i + (x_{i+1} - x_i) \times d \quad (3.7)$$

Onde:  $x_i$  é o valor da observação posicionada no índice  $i$ , sendo  $i$  a parte inteira e  $d$  a parte decimal do índice calculado.

Por exemplo, para os seguintes dados ordenados: 1, 2, 3, 4, 7, 10, usando as Equações 3.4, 3.5 e 3.6, calcula-se os índices  $l$ ,  $m$  e  $u$ , a fim de encontrar as posições dos quartis, sendo o número de dados  $n = 6$ .

$$l = \frac{6 + 3}{4} = 2,25 \quad (3.8)$$

$$m = \frac{6 + 1}{2} = 3,5 \quad (3.9)$$

$$u = \frac{3 \times 6 + 1}{4} = 4,75 \quad (3.10)$$

Como  $l$ ,  $m$  e  $u$  não são valores inteiros, realiza-se a interpolação linear usando a Equação 3.7, com o valor de  $l$  para calcular o quartil inferior, com o valor de  $m$  para calcular a mediana e como valor de  $u$  para calcular o quartil superior.

Na interpolação usando  $l$ , como a parte inteira é  $i = 2$  e a parte decimal é  $d = 0,25$ , então:

$$Q1 = x_2 + (x_3 - x_2) \times 0,25 = 2 + (3 - 2) \times 0,25 = 2,25 \quad (3.11)$$

Na interpolação usando  $m$ , como a parte inteira é  $i = 3$  e a parte decimal é  $d = 0,5$ , então:

$$Q2 = x_3 + (x_4 - x_3) \times 0,5 = 3 + (4 - 3) \times 0,5 = 3,5 \quad (3.12)$$

Na interpolação usando  $u$ , como a parte inteira é  $i = 4$  e a parte decimal é  $d = 0,75$ , então:

$$Q3 = x_4 + (x_5 - x_4) \times 0,75 = 4 + (7 - 4) \times 0,75 = 6,25 \quad (3.13)$$

Os valores obtidos da análise de quartis do exemplo descrito, onde foram utilizados os dados ordenados: 1, 2, 3, 4, 7, 10, estão apresentados na Tabela 4.

**Tabela 4.** Valores obtidos da análise de quartis do exemplo descrito.

Medidas estatísticas	Valor
Q1	2,25
Q2 (mediana)	3,5
Q3	6,25
IIQ = Q3-Q1	4,0

Como os quartis não são afetados por observações extremas (*outliers*), a mediana (Q2) é a medida de maior eficiência na análise de tendência central para dados altamente assimétricos quando comparada com a média, pois, para se utilizar a média em amostras que contenham *outliers*, ainda seria necessária, previamente, a realização de uma análise exploratória (Mosteller e Tukey, 1977).

## 2.4 Trabalhos relacionados

Este trabalho propõe um método de localização em ambientes internos utilizando redes IEEE 802.11. Para isso, foram feitos estudo na literatura de trabalhos com temas relacionados à proposta deste trabalho que abordam algoritmos de aprendizado de máquina, baseados em conjuntos de dados originados pelo indicador de intensidade do sinal recebido (RSSI) de redes Wi-Fi. Dentre estes trabalhos estudados destacam-se:

O artigo de Kuntal e Karmakar (2014) propõe um algoritmo de localização baseado no classificador  $k$  vizinho mais próximo ( $k$ NN) para redes IEEE 802.11. A localização do nó móvel (Mobile Node – MN) é obtida pelo centróide, ou seja, pela média das coordenadas dos  $k$  vizinhos mais próximos. Foi utilizado o software MATLAB (The MathWorks Inc., 2018) nas simulações, o modelo de sombreamento log-normal para obtenção dos valores de RSSI e a distância euclidiana para verificação de similaridade. No intuito de obter um melhor desempenho, variaram-se diferentes parâmetros. Entre eles, foi examinado o efeito do número de pontos de acesso (APs), o valor de  $k$  e o fator de sombreamento no cálculo da intensidade do sinal recebido (RSSI), pois este fator é alterado quando ocorre uma alteração no ambiente.

Os resultados indicam que a utilização de sete APs reflete em um melhor desempenho, visto que o erro diminuiu à medida que o número de APs aumentou, porém o erro tornou-se constante quando o número de APs foi superior a sete. O valor de  $k=4$  também obteve melhor desempenho nas simulações, apresentando um erro de posicionamento de 0,8m. O valor do fator de sombreamento foi inicialmente definido como zero e o erro de localização só tendeu a crescer com o aumento deste fator.

A metodologia utilizada por Kuntal e Karmakar (2014) pode não apresentar resultados satisfatórios em uma plataforma de testes reais, pois a representação do RSSI na fase off-line por meio do modelo de sombreamento log-normal não apresenta conformidade com o RSSI obtido em meio à ocorrência de perturbações na fase on-line, limitando-se apenas a testes simulados.

O trabalho de Torres-Sospedra et al. (2015) apresenta um estudo abrangente usando 51 métricas de distância e similaridade, 4 alternativas para representar os dados brutos e um limiar baseado nos valores de RSSI. Os testes foram realizados utilizando o banco de dados público UJIIndoorLoc. O artigo indica que sistemas de posicionamento interno (Indoor Positioning System – IPS) baseados em  $k$ NN são melhorados apenas selecionando a configuração apropriada.

Diante das possíveis configurações apresentadas, o melhor resultado foi alcançado utilizando a função de similaridade Sørensen (Sørensen, 1948) combinada com a representação de dados *powed*, que foi proposta pelos autores para tratar a natureza logarítmica dos dados brutos de RSSI, e  $k=13$ . Este valor de  $k$  foi obtido após testes com valores ímpares de 1 a 23, apresentando taxa de acertos de 95,2% e um erro de posicionamento de 6,19m. A aplicação de um método de limiar proposta para remover os sinais Wi-Fi com intensidade muito baixa não apresentou melhorias significativas, mas diminuiu o desempenho do sistema de localização quando os valores de intensidade próximos aos valores médios foram descartados e como conclusão o método não foi sugerido.

Assim como a remoção dos níveis baixos de intensidade prejudica o desempenho do sistema, o uso da representação *powed* pode não apresentar resultados satisfatórios no algoritmo de classificação. Isto porque, variáveis ou atributos discriminantes são decisivos para um algoritmo de classificação, e os valores normalizados dos níveis baixos de intensidade pela representação *powed* possuem alta similaridade.

O trabalho desenvolvido por Jedari et al. (2015) investigou o desempenho de três classificadores baseados em aprendizado de máquina, incluindo o  $k$  vizinhos mais próximos ( $k$ NN), o classificador JRip proposto em (Witten e Frank, 2005) e um terceiro que utiliza um algoritmo de floresta aleatória para estimar a localização interna. Foram realizadas medições experimentais usando 50 pontos de referência (RPs), para cada RP foram obtidas 30 amostras com RSSIs brutos de 86 pontos de acesso (APs).

Apresentando taxa de acertos de 77,40%, o classificador  $k$ NN mostrou-se superior ao JRip, porém, não apresentou melhor desempenho quando comparado ao classificador de floresta aleatória, com 91%, devido às limitações da técnica Qui-quadrado ( $\chi^2$ ) que foi aplicada para remover dados redundantes de RSSI medidos dos APs, antes que estes dados fossem utilizados pelo algoritmo  $k$ NN.

Segundo McHugh (2013) as limitações da técnica Qui-quadrado ( $\chi^2$ ) incluem os requisitos no tamanho do conjunto de dados, uma vez que

apresenta grande quantidade de erros de interpretação dos dados quando se utiliza um grande número (20 ou mais) de classes (ex.: RPs ou pontos onde foram coletados os dados) ou de atributos (APs utilizados para as medições).

O trabalho de Salamah et al. (2016) utilizou a Análise de Componentes Principais (Principal Component Analysis - PCA) para melhorar o desempenho dos sistemas de localização interna. A PCA identifica a redundância entre múltiplos atributos, transforma o conjunto dos valores de RSSI em um conjunto com novos atributos ou componentes principais (PCs) ordenados pela importância que cada um possui em termos de informação, podendo eliminar atributos com menor importância e representar o conjunto de dados com dimensão reduzida (Fang e Lin, 2012).

O desempenho do método proposto foi testado utilizando os classificadores  $k$  vizinhos mais próximos ( $k$ NN), Árvore de Decisão, Floresta Aleatória e Máquina de Vetores de Suporte (Support Vector Machine - SVM). Os experimentos foram conduzidos em um ambiente real em 45 pontos de referência (RPs) utilizando smartphones para a coleta dos valores de RSSI de seis APs. Os resultados mostram que o  $k$ NN com  $k=2$  e baseado nos três primeiros PCs obteve o melhor desempenho nos experimentos dinâmicos, apresentando um erro médio de posicionamento de 1,71m com precisão de 60% e de 3,0m com precisão de 79%.

No método abordado por Salamah et al. (2016), a localização foi calculada pela média ponderada das coordenadas dos  $k$  vizinhos mais próximos, assim como em (Kuntal e Karmakar, 2014). Este cálculo apresenta a possibilidade de localizar objetos em posições próximas aos RPs e onde não foram realizadas as coletas de RSSI.

No trabalho de Tang et al. (2016), para melhorar a precisão de posicionamento que é afetada pela flutuação do RSSI, foram propostos: um algoritmo de transformação RSSI-Probabilidade para a construção de um novo conjunto de dados formado por valores de RSSI e suas probabilidades correspondentes; e o agrupamento de pontos de referência (RPs) em clusters, calculando o peso de cada ponto de acesso (AP) para representar pequenas regiões no ambiente de localização.

Para verificar o desempenho da proposta, foi desenvolvido um sistema de posicionamento chamado LocNeedle baseado no classificador  $k$  vizinhos mais próximos ( $k$ NN) adotando a distância Bhattacharyya (Bhattacharyya, 1943) como métrica de medida de similaridade. O sistema foi comparado com os sistemas RADAR proposto por Bahl e Padmanabhan (2000) e o LocNeedle usando o conjunto de dados apenas com o RSSI médio de cada AP. O melhor resultado experimental foi alcançado utilizando o sistema LocNeedle combinado com o algoritmo de transformação RSSI-Probabilidade, apresentando um erro de posicionamento de 4,5m.

A probabilidade de um dado valor de RSSI pertencer a uma posição é uma alternativa promissora para contornar o problema com valores altamente similares de RSSI, sendo mais eficiente que a remoção dos valores baixos de RSSI adotada por Torres-Sospedra et al. (2015). Isto demonstra que apenas a utilização de novas representações de RSSI pode não ser suficiente para melhorar a precisão, porém, atributos adicionais com informações específicas, que caracterizam o comportamento do RSSI, garantem uma melhor influência nos resultados de localização.

Os artigos apresentados utilizaram algoritmos de aprendizado de máquina com a abordagem de conjuntos de dados originados pelo RSSI, a grande maioria implementou e obteve os melhores resultados com o classificador  $k$  vizinhos mais próximos ( $k$ NN) na validação da proposta de localização, apresentando diferentes definições na configuração apropriada que deve ser utilizada para obter um resultado satisfatório. Algumas características dos trabalhos relacionados são apresentadas na Tabela 5.

**Tabela 5.** Características dos trabalhos relacionados.

Artigo	Obtenção do RSSI	Obtenção das coordenadas	Acurácia (%)	Erro médio (m)
Kuntal e Karmakar (2014)	Simulação	Centróide dos RPs	100	4,00
Torres-Sospedra et al. (2015)	Prática	RP majoritário	95,20	6,19
Jedari et al. (2015)	Prática	RP majoritário	77,40	-
Salamah et al. (2016)	Prática	Centróide dos RPs	79,00	3,00
Tang et al. (2016)	Prática	RP majoritário	100	4,50

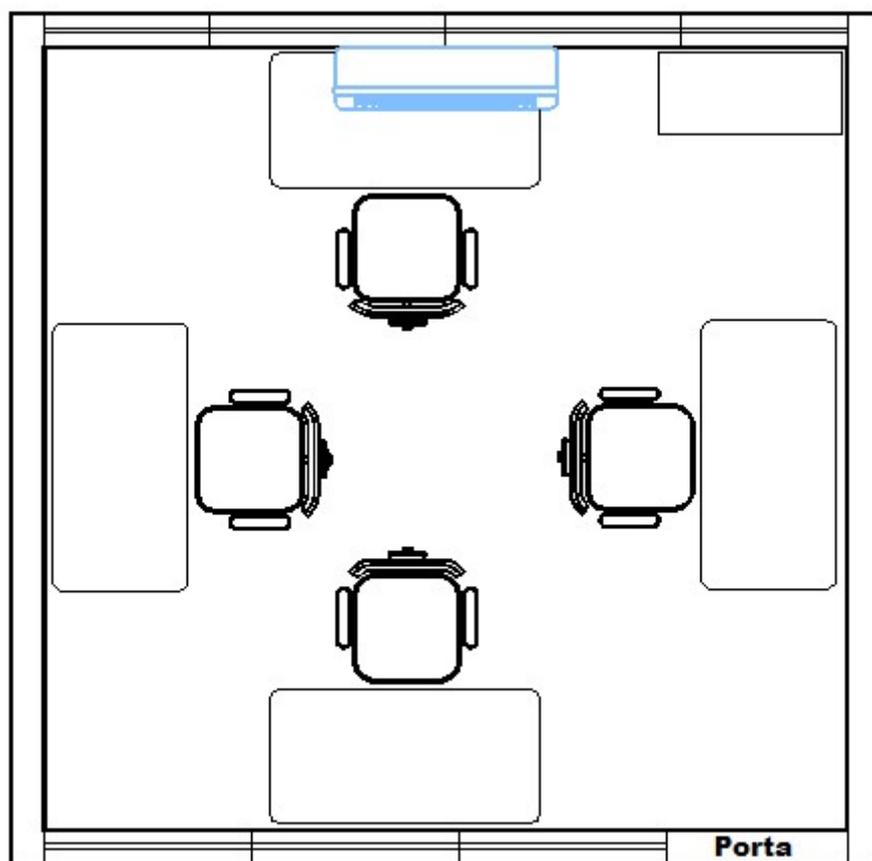
O fator decisivo para um método de localização interna baseado no algoritmo de classificação *k*NN é a existência de características discriminantes entre os diversos RPs. Para que seja possível gerar a localização com precisão, devem-se tratar primeiramente os problemas de instabilidade no RSSI do sinal Wi-Fi. Como solução, este trabalho propõe a utilização adicional de valores estatísticos calculados a partir de amostras de RSSI, tornando os dados adequados para algoritmos de classificação.

## 3 Material e métodos

### 3.1 Cenário de experimentação

Os experimentos de estimativa de localização foram realizados em um ambiente interno, tal como a sala de professor 01, bloco CETELI-2 da Faculdade de Tecnologia da Universidade Federal do Amazonas (UFAM), na cidade de Manaus, AM.

A Figura 3 mostra o *layout* do cenário de experimentação, tendo 12,46 metros quadrados, contendo um armário ilustrado no canto superior direito da figura e quatro mesas retangulares dispostas no centro de cada parede.



**Figura 3.** *Layout* do cenário de experimentação.

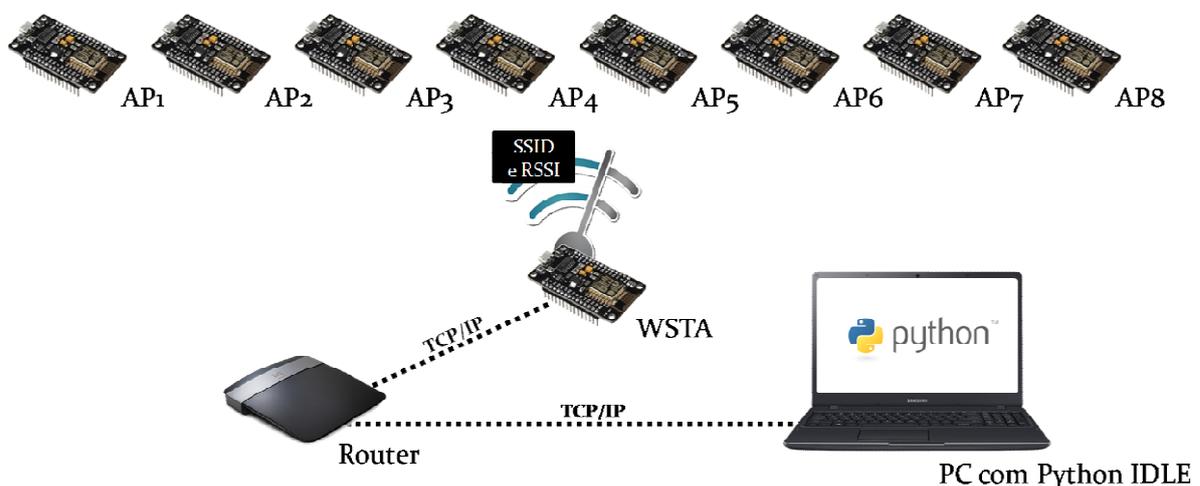
O sistema de localização foi implementado com a utilização de módulos NodeMCU. Esta placa consiste de um microcontrolador da família ESP8266, interface UART-USB e um regulador de tensão 3,3V, e é

alimentada por baterias externas de 4,5 a 9V (NodeMCU *Documentation*, 2018). A placa pode ser empregada no modo estação (STA), ponto de acesso (AP) ou em ambos (STA+AP) e pode ser configurada nos padrões 802.11b/g/n. Neste trabalho, os NodeMCUs foram pré-configurados no padrão 802.11g.

A Figura 4 ilustra o diagrama de conexões dos elementos que compõe o sistema utilizado nos experimentos, onde são apresentados os nós fixos chamados de AP1 até AP8 que foram configurados no modo AP.

Utilizando o verificador de rede “Wi-Fi Analyzer”, foi observado que os canais mais ocupados pelos pontos de acesso instalados próximos ao ambiente são o 1, 6 e 11, possivelmente por não haver sobreposição entre estes canais. Diante do exposto, foi selecionado o canal 4 para os APs do sistema de localização visando reduzir a interferência Wi-Fi.

O nó móvel a ser localizado foi configurado no modo STA e por conveniência foi chamado de WSTA (*Wireless Station*). Nenhuma antena direcional foi anexada aos módulos.



**Figura 4.** Diagrama de conexões dos elementos que compõem o sistema de localização.

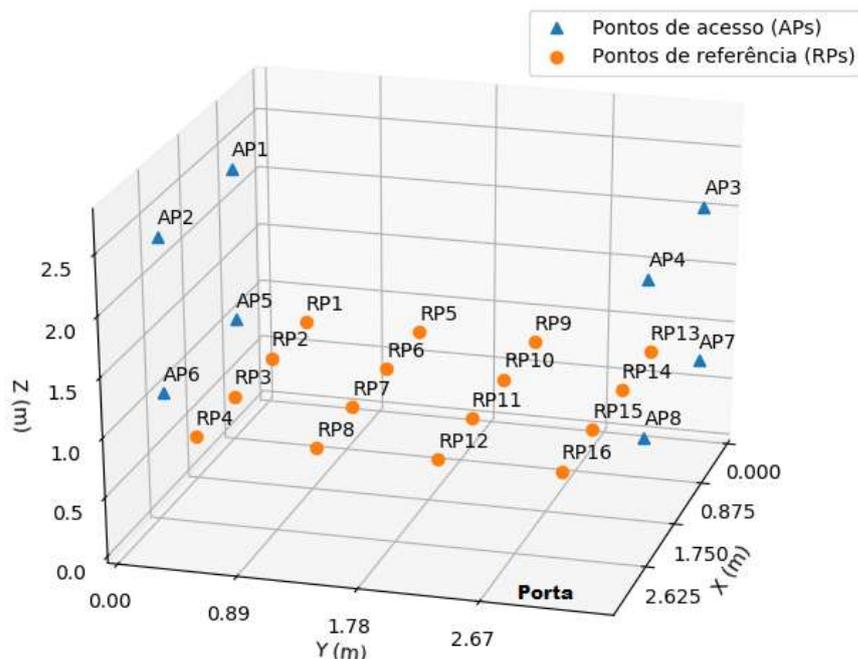
Ainda na Figura 4, uma vez que o nó WSTA é configurado no modo STA, ele varre os canais disponíveis a fim de identificar os APs existentes e coletar informações, como o identificador (*Service Set Identifier* – SSID) e o valor atual do RSSI, que são fundamentais para o sistema de localização. O computador pessoal (*Personal Computer* – PC) com Python IDLE (*Python’s*

*Integrated Development and Learning Environment*) utiliza o roteador sem fio (*Router*) para se conectar ao nó WSTA via rede Wi-Fi.

Neste trabalho foi utilizado um *notebook* Samsung RV410, Processador T4500, 2.3GHz, 32 bits, 2GB de RAM e Windows 7 Starter.

O NodeMCU utilizado como WSTA possui um servidor HTTP embarcado e configurado para, quando solicitado, enviar o RSSI coletado de cada um dos oito APs instalados no cenário de experimentação. O PC com Python IDLE utiliza a rede sem fio para enviar requisições ao nó WSTA e obter os valores de RSSI como resposta. O PC armazena e processa estas informações por um algoritmo escrito em linguagem Python, a fim de obter a localização do nó WSTA.

Para os experimentos adotou-se uma divisão do ambiente interno em 16 zonas, com Pontos de Referência (*Reference Points* - RPs) de mesma altura (0,87 metros) posicionados no centro de cada zona. Os RPs são utilizados como pontos de coleta de amostras de RSSI dos pontos de acesso (APs). A Figura 5 ilustra o posicionamento dos APs e dos RPs, onde os oito APs são representados por triângulos e os 16 RPs representados por círculos. O ambiente interno possui dimensões de 3,50m de comprimento (X), 3,56m de largura (Y) e 2,80m de altura (Z).



**Figura 5.** Posicionamento dos pontos de acesso (APs) e dos pontos de referência (RPs) no ambiente interno.

A Tabela 6 apresenta o conjunto de coordenadas dos APs e dos RPs no ambiente interno, onde foram realizados os experimentos.

**Tabela 6.** Coordenadas dos pontos de acesso (APs) e dos pontos de referência (RPs) no ambiente interno.

Identificador	Coordenada X	Coordenada Y	Coordenada Z
Pontos de Acesso			
AP1	0,78	0,00	2,27
AP2	2,48	0,00	2,27
AP3	0,78	3,56	2,27
AP4	2,48	3,56	2,27
AP5	0,78	0,00	0,97
AP6	2,48	0,00	0,97
AP7	0,78	3,56	0,97
AP8	2,48	3,56	0,97
Pontos de Referência			
RP1	0,4375	0,445	0,87
RP2	1,3125	0,445	0,87
RP3	2,1875	0,445	0,87
RP4	3,0625	0,445	0,87
RP5	0,4375	1,335	0,87
RP6	1,3125	1,335	0,87
RP7	2,1875	1,335	0,87
RP8	3,0625	1,335	0,87
RP9	0,4375	2,225	0,87
RP10	1,3125	2,225	0,87
RP11	2,1875	2,225	0,87
RP12	3,0625	2,225	0,87
RP13	0,4375	3,115	0,87
RP14	1,3125	3,115	0,87
RP15	2,1875	3,115	0,87
RP16	3,0625	3,115	0,87

### 3.2 Experimentos para a relação entre distância e RSSI

A fim de analisar a relação entre a distância de cada um dos APs ao nó WSTA e os valores de RSSI coletados, foram executados experimentos preliminares nos pontos de referência (RPs) 1 e 2 ilustrados na Figura 5.

Foram utilizados dois RPs adjacentes para verificar as variações e o comportamento da perda de potência na propagação dos sinais de RF relacionados a RPs próximos, devido aos fenômenos causados pelas paredes e obstáculos presentes no ambiente interno.

Nestes experimentos, inicialmente foram obtidas as distâncias entre todos os APs e o nó WSTA posicionado sequencialmente nos RPs 1 e 2. Na Equação 4.1 é apresentado o modelo matemático para determinação da distância entre dois pontos em geometria espacial. A partir das coordenadas no ambiente, as distâncias são obtidas pela seguinte equação:

$$d(p, q) = \sqrt{\sum_{i=1}^3 (p_i - q_i)^2} \quad (4.1)$$

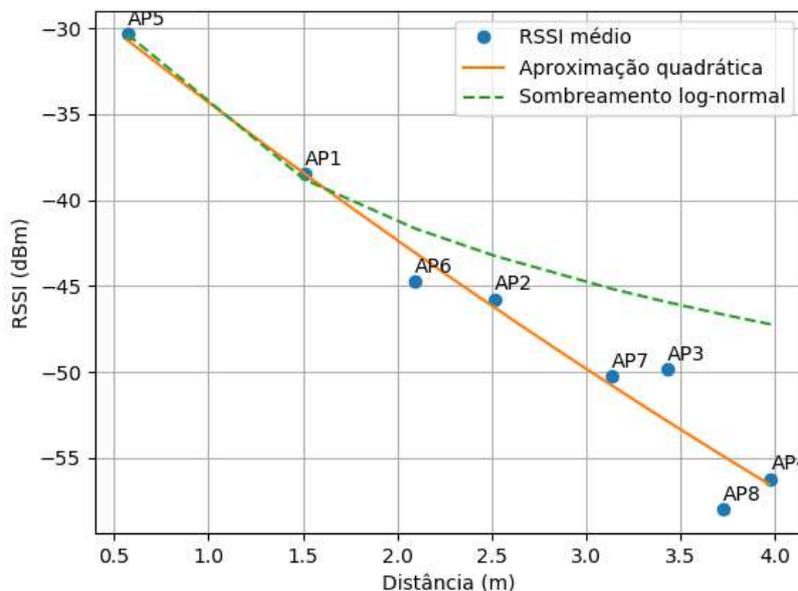
Onde:  $p_i$  é o valor da coordenada  $i$  para o ponto  $p$  e  $q_i$  é o valor da coordenada  $i$  para o ponto  $q$ .

Duas técnicas presentes na literatura foram testadas para expressar, em ambientes internos, o comportamento do RSSI em função da distância: a técnica de sombreamento log-normal, também chamado de perda de propagação log-normal, utilizado em (Feng et al., 2008; Kuntal e Karmakar, 2014; Zhao e Wang, 2017) e a técnica da aproximação quadrática utilizado em (Isokawa et al., 2013; Parwekar e Reddy, 2013).

As Figuras 6 e 7 apresentam, cada uma, duas curvas do RSSI determinadas em função da distância nos testes práticos. A curva tracejada verde corresponde à técnica de sombreamento log-normal e a curva sólida laranja corresponde à técnica de aproximação quadrática.

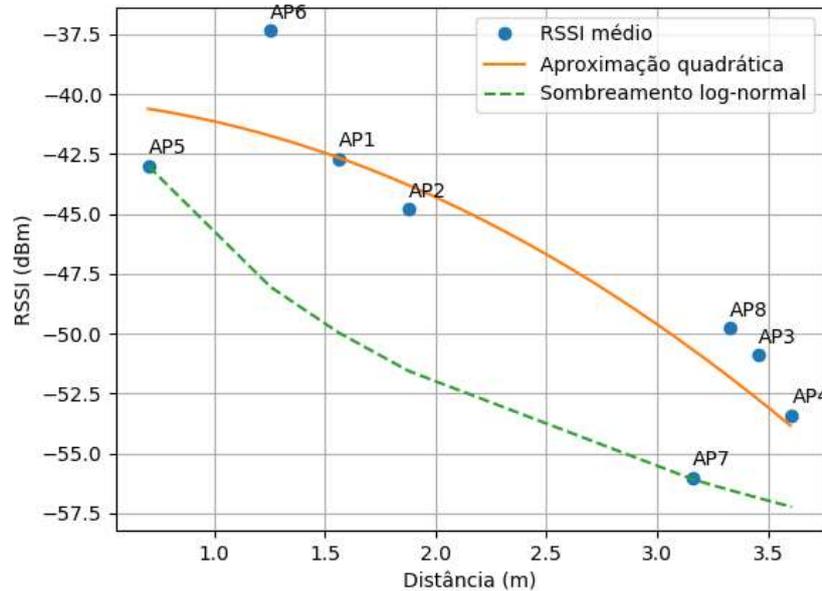
Os círculos azuis correspondem à média aritmética de 20 leituras de RSSI de todos os APs nos RPs 1 e 2, respectivamente nas Figuras 6 e 7. A estimativa do RSSI médio a partir de 20 leituras é suficiente para reduzir o erro quadrático das variações (Graciolie et al., 2011).

A Figura 6 mostra os resultados dos experimentos realizados no RP1. Verifica-se que o RSSI decresce exponencialmente com o aumento da distância nas técnicas de sombreamento log-normal e aproximação quadrática, conforme era esperado; no entanto, o RSSI médio das amostras coletadas apresentou algumas variações.



**Figura 6.** Comparação entre as técnicas de sombreamento log-normal, aproximação quadrática e RSSI médio para obtenção da intensidade do sinal recebido (RSSI) no ponto de referência 1 (RP1).

A Figura 7 apresenta os resultados das três técnicas utilizadas para obtenção do RSSI no RP2. Como o RSSI médio apresenta-se com maiores variações para distâncias próximas, não tendo uma relação estrita na proporção inversa do quadrado da distância, as técnicas de sombreamento log-normal e aproximação quadrática tendem a ficar mais afastados do valor médio do RSSI neste cenário de experimentação.



**Figura 7.** Comparação entre as técnicas de sombreamento log-normal, aproximação quadrática e RSSI médio para obtenção da intensidade do sinal recebido (RSSI) no ponto de referência 2 (RP2).

As amostras de RSSI coletadas tendem a sofrer diferentes variações no ambiente interno, decorrentes: da ausência de linha de visada entre os APs e o nó WSTA (nó coletor) por não haver antenas direcionais anexadas aos módulos; e dos fenômenos físicos de propagação das ondas eletromagnéticas causados pelas paredes e obstáculos presentes no ambiente, o que foi observado nos experimentos preliminares.

No entanto, as técnicas teóricas de sombreamento log-normal e aproximação quadrática utilizadas em alguns trabalhos não podem expressar as perdas de potência dos sinais em testes práticos. Quatro observações foram possíveis durante os experimentos:

1. Sinais coletados em um RP com a mesma distância em relação a dois APs possuíram valores médios diferentes de RSSI;
2. Sinais coletados em um RP com distâncias diferentes em relação a dois APs possuíram o mesmo valor médio de RSSI;
3. Sinais coletados em dois RPs com a mesma distância em relação a um AP possuíram valores médios diferentes de RSSI;

4. Sinais coletados em dois RPs com distâncias diferentes em relação a um AP possuíram o mesmo valor médio de RSSI.

Neste experimento, verificou-se que o RSSI médio manteve-se similar em testes conduzidos em dias diferentes. Assim, os métodos relacionados com a assinatura do RSSI médio podem disponibilizar mais informações sobre as medições de RSSI em cada RP.

Desta forma, foram utilizados métodos baseados em *fingerprint* auxiliado por características específicas do comportamento do RSSI para determinar se uma WSTA está ou não em um determinado RP, conforme poderá ser observado nas próximas seções.

### 3.3 Propostas para as estimativas de localização

#### 3.3.1 Representação dos dados utilizando quartis

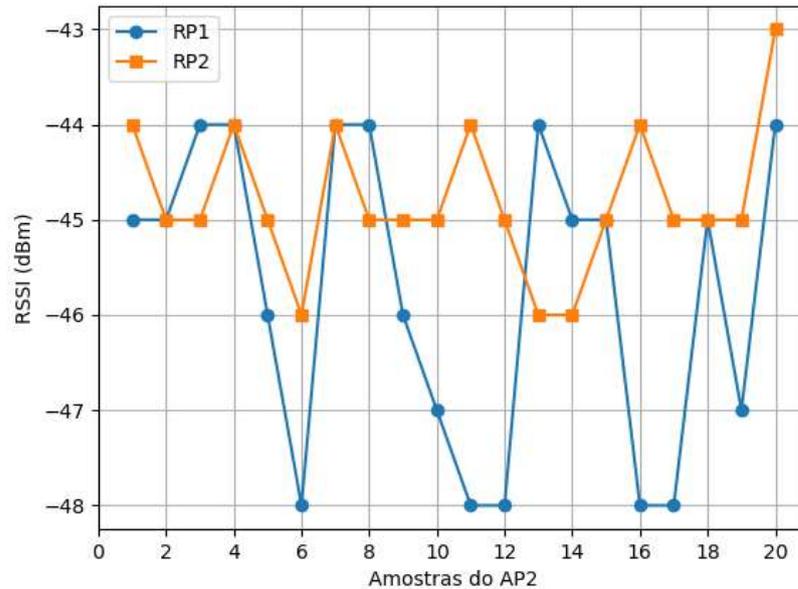
Conforme foi visto na seção anterior, o valor do RSSI médio das leituras de certo AP em um ponto de referência (RP), em ambientes internos, pode ser igual ao RSSI médio em outros RPs que estão a distâncias maiores ou menores ao mesmo AP. Por isso, são necessárias características que diferenciem os RPs de mesmo RSSI médio e com distâncias diferentes.

Após alguns estudos preliminares, uma possível solução encontrada foi o cálculo dos quartis a partir de leituras de RSSI, devido ao fato de o segundo quartil (Q2 ou mediana) ser mais robusto em relação à média aritmética, pois Q2 não sofre influência de dados com grande afastamento das demais observações, chamadas de *outliers* (Mosteller e Tukey, 1977).

Também foram utilizados o primeiro e o terceiro quartil (Q1 e Q3) para diferenciar os RPs com o mesmo valor de Q2, pois foram observados que ainda haviam valores similares da mediana para RPs com distâncias diferentes, porém estes mesmos RPs possuíam diferentes valores de Q1 e/ou Q3, conforme será visto a seguir.

A Figura 8 apresenta o comportamento das leituras de RSSI do AP2

coletadas a partir dos RPs adjacentes 1 e 2 que conforme visualizado na Figura 5, possuem diferentes distâncias para o AP2, respectivamente iguais à 2,52m e 1,88m.



**Figura 8.** Leituras da intensidade do sinal recebido (RSSI) do ponto de acesso 2 (AP2) coletadas nos pontos de referência 1 e 2 (RP1 e RP2).

A Tabela 7 mostra os valores dos quartis do AP2 analisado nos RPs 1 e 2. Embora sejam coletadas em diferentes distâncias, as 20 leituras em cada RP resultam no mesmo valor para o Q2 calculado, ou seja, o mesmo RSSI de  $-45dBm$  para o AP2 visto em dois RPs distintos. Sendo assim, entende-se a necessidade de utilizar, além da mediana (Q2), as características Q1 e/ou Q3 para diferenciar dois RPs distintos.

**Tabela 7.** Quartis do ponto de acesso 2 (AP2) com a intensidade do sinal recebido (RSSI) medida nos pontos de referência 1 e 2 (RP1 e RP2).

	AP2		
	Q1 (dBm)	Q2 (dBm)	Q3 (dBm)
RP1	-47,25	-45	-44
RP2	-45	-45	-44

Conforme visualizado na Tabela 7, pode-se verificar a diferença nos valores do Q1 para as amostras em cada RP. Isso ocorre devido as leituras de RSSI coletadas no RP1 possuírem maior variação, causada por influência dos diferentes obstáculos, que fazem com que as menores leituras de RSSI estejam mais afastadas da mediana (Q2). Esta característica pode ser observada na Figura 8, onde na curva sólida azul, observam-se leituras no RP1 com valores menores de RSSI quando comparado a curva sólida laranja, referente ao RP2. Tal comportamento também foi observado em novas amostras do AP2 nestes RPs.

Dado que um AP visto em diferentes RPs pode apresentar o mesmo valor de Q2 para suas amostras de RSSI, este Q2 calculado, não pode ser utilizado isoladamente para determinar, através de um algoritmo de classificação, a posição correta do nó sem fio. No entanto, um fator decisivo para um algoritmo de classificação é a existência de características com diferentes valores entre instâncias pertencentes a classes diferentes, a fim de ser possível classificar objetos em apenas uma destas classes.

Sendo assim, como alternativa para esta problemática, tem-se a utilização adicional dos demais quartis como características de cada um dos APs, vistos a partir de cada RP. Isto possibilita a diferenciação das instâncias, tornando os dados adequados para algoritmos de classificação, como o algoritmo  $k$ NN que foi implementado nos trabalhos relacionados.

### 3.3.2 Localização utilizando as coordenadas do ponto de referência majoritário

Neste método de localização, o algoritmo de classificação  $k$ NN adota a distância euclidiana como função de similaridade e a análise de quartis na representação dos dados, conforme foi proposto na Seção 4.3.1. Após transformação dos dados de RSSI em quartis e seleção dos  $k$  vizinhos mais próximos, estimou-se a localização da WSTA pela Equação 4.2.

As coordenadas  $(X_m, Y_m, Z_m)$ , correspondentes à localização estimada da WSTA, são determinadas nas coordenadas do RP majoritário, ou seja, do RP

com maior frequência de ocorrência dentre os  $k$  vizinhos mais próximos, conforme a seguinte equação:

$$f_i = \max\{f_1, \dots, f_n\} \rightarrow (X_m, Y_m, Z_m) = (X_i, Y_i, Z_i) \quad (4.2)$$

Isto é, se a frequência de ocorrência  $f_i$  é a maior frequência observada, então as coordenadas  $X_m, Y_m, Z_m$  são iguais as coordenadas  $X_i, Y_i, Z_i$  do ponto de referência  $i$ . Quando ocorre um empate nas frequências de múltiplos RPs, atribui-se as coordenadas do RP associado ao vizinho mais próximo à localização estimada.

As coordenadas do RP majoritário é uma abordagem comumente utilizada quando o algoritmo  $k$ NN é indicado para o problema de localização, como nos trabalhos feitos por Jedari et al. (2015), Torres-Sospedra et al. (2015) e Tang et al. (2016). No entanto ela possibilita a estimativa de localização apenas nas posições onde foram coletadas amostras de RSSI.

### 3.3.3 Localização utilizando as coordenadas do centróide dos pontos de referência

Assim como no método anterior, nesta proposta o algoritmo de classificação  $k$ NN adota a distância euclidiana como função de similaridade e o conjunto de instâncias baseadas em quartis. No entanto, após transformação dos dados de RSSI em quartis e seleção dos  $k$  vizinhos mais próximos, estimou-se a localização da WSTA pela Equação 4.3.

As coordenadas  $(X_c, Y_c, Z_c)$ , correspondentes à localização estimada da WSTA, foram calculadas através do centróide, ou seja, da média das coordenadas dos RPs  $(X_i, Y_i, Z_i)$  associados aos  $k$  vizinhos mais próximos, ponderadas pelas respectivas frequências das ocorrências de cada RP ( $p_i$ ), conforme a seguinte equação:

$$(X_c, Y_c, Z_c) = \left( \frac{\sum_{i=1}^n p_i X_i}{\sum_{i=1}^n p_i}, \frac{\sum_{i=1}^n p_i Y_i}{\sum_{i=1}^n p_i}, \frac{\sum_{i=1}^n p_i Z_i}{\sum_{i=1}^n p_i} \right) \quad (4.3)$$

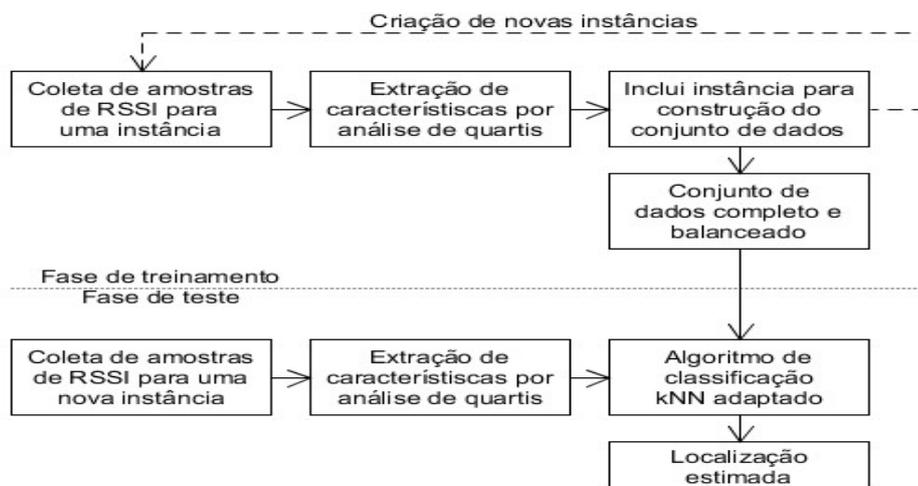
Onde:  $X_c, Y_c, Z_c$  são as coordenadas do centróide e  $p_i$  é o peso para o ponto de referência  $i$ .

As coordenadas do centróide foram abordadas nos trabalhos feitos por Kuntal e Karmakar (2014) e Salamah et al. (2016), e possibilita a estimativa de localização em qualquer posição dentro do ambiente interno, o que inclui as posições onde não foram coletadas amostras de RSSI.

### 3.4 Experimentos para as estimativas de localização

Os dois métodos propostos foram testados: o método baseado nos quartis e no ponto de referência (RP) majoritário dentre os  $k$  vizinhos mais próximos descrito na Seção 4.3.2; e o método baseado nos quartis e no centróide dos RPs dos  $k$  vizinhos mais próximos descrito na Seção 4.3.3. Estas propostas são denotadas nesta seção como Método I e Método II, respectivamente.

A Figura 9 apresenta o diagrama de fluxo de treinamento e teste comum aos métodos propostos.



**Figura 9.** Diagrama de fluxo comum aos métodos propostos.

Como mostrado na Figura 9, os métodos de localização (Método I e Método II) envolvem duas fases (ou estágios). Na fase de treinamento (estágio *off-line*), uma WSTA coleta amostras de RSSI enviadas pelos APs periodicamente e registra estes dados em posições conhecidas do espaço físico (chamados de RPs). Então, de acordo com a análise de quartis, são calculadas características mais discriminativas para estabelecer um conjunto de dados adequado à classificação. Novas instâncias são criadas até que se tenha balanceado o conjunto de dados com o mesmo número de instâncias por RPs. Na fase de teste (estágio *on-line*), o sistema também registra os valores de RSSI recebidos dos APs, mas em posições desconhecidas. Então, de acordo com o algoritmo *k*NN adaptado à cada método proposto, será estimada a localização real da WSTA.

### 3.4.1 Fase de treinamento

Nesta etapa, o treinamento ou aprendizado supervisionado do algoritmo consiste, basicamente, na criação de um conjunto de instâncias previamente classificadas em múltiplas classes. As classes do conjunto de treinamento representam os 16 pontos de referência (RPs).

Inicialmente, para criar uma instância previamente classificada em uma classe, a WSTA é posicionada manualmente no RP correspondente à classe, onde são coletadas amostras com  $m$  leituras de RSSI bruto de cada um dos  $n$  APs instalados. As amostras são organizadas numa matriz  $A_i$  ( $m \times n$ ), onde a primeira coluna contém as  $m$  leituras do AP1 e assim por diante. Nos testes iniciais, utilizou-se  $m = 20$ .

$$A_i = \begin{bmatrix} RSSI_{11} & RSSI_{12} & RSSI_{13} & \cdots & RSSI_{1n} \\ RSSI_{21} & RSSI_{22} & RSSI_{23} & \cdots & RSSI_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ RSSI_{m1} & RSSI_{m2} & RSSI_{m3} & \cdots & RSSI_{mn} \end{bmatrix} \quad (4.4)$$

Em seguida, os três quartis (Q1, Q2 e Q3) são calculados para os  $n$

APs com base nas suas respectivas amostras. O vetor  $b_i$  representa uma instância formada pelos quartis calculados a partir das amostras (colunas) que compõem a matriz  $A_i$ , resultando em um vetor linha composto por  $3n$  elementos ou atributos.

$$b_i = \overbrace{[Q1 \quad Q2 \quad Q3]}^{AP_1} \quad \overbrace{[Q1 \quad Q2 \quad Q3]}^{AP_2} \quad \dots \quad \overbrace{[Q1 \quad Q2 \quad Q3]}^{AP_n} \quad (4.5)$$

Para criar instâncias previamente classificadas em múltiplas classes, a WSTA foi posicionada sequencialmente em todos os 16 RPs e os passos supracitados foram repetidos, coletando-se novas amostras de RSSI para cada instância, até que todas as instâncias fossem criadas.

Com intuito de balancear o conjunto de treinamento utilizado neste trabalho, isto é, manter o mesmo número de instâncias entre as classes, criou-se 10 instâncias para cada uma das 16 classes.

No final desta etapa, tem-se o conjunto de treinamento representado pela matriz  $T$  de dimensões  $(160 \times 24)$ , contendo 160 instâncias representadas por  $b_i$  com 24 atributos, sendo três quartis das amostras de cada um dos oito pontos de acesso (APs), em que cada linha da matriz  $T$  representa uma instância e cada coluna um atributo.

$$T = \left[ \begin{array}{c} b_1 \\ \vdots \\ b_{10} \\ b_{11} \\ \vdots \\ b_{20} \\ \vdots \\ b_{151} \\ \vdots \\ b_{160} \end{array} \right] \left. \begin{array}{l} \} RP_1 \\ \} RP_2 \\ \vdots \\ \} RP_{16} \end{array} \right\} \quad (4.6)$$

### 3.4.2 Fase de testes

Durante a etapa de testes ou execução do algoritmo, criou-se um

conjunto de teste  $E$  com 160 novas instâncias a partir de novas amostras de RSSI, sendo 10 instâncias associadas a cada um dos 16 RPs. Porém, os RPs do conjunto de teste são omitidos no processo de classificação e utilizados somente em análise posterior para verificar a precisão dos métodos propostos.

$$E = \begin{bmatrix} b_1 \\ \vdots \\ b_{10} \\ b_{11} \\ \vdots \\ b_{20} \\ \vdots \\ b_{151} \\ \vdots \\ b_{160} \end{bmatrix} \left. \begin{array}{l} \} \\ \} \\ \vdots \\ \} \end{array} \right\} \begin{array}{l} RP_1 \\ RP_2 \\ \vdots \\ RP_{16} \end{array} \quad (4.7)$$

As 160 instâncias do conjunto de teste  $E$  são aplicadas ao processo de classificação pelo algoritmo  $k$ NN adaptado aos métodos I e II, que têm armazenado previamente o conjunto de treinamento  $T$ , matriz (3.6), criado na etapa anterior. Os métodos de localização estimam a posição real da WSTA (RP ou classe) associada a cada instância de teste e em seguida é verificada a precisão das 160 estimativas.

Durante a etapa de testes, com a finalidade de obter o menor erro de localização para cada método e com baixo custo computacional, o processo de classificação das 160 instâncias do conjunto de teste  $E$  foi submetido a 49 tratamentos quantitativos, definidos por combinações, variando-se o número de APs ( $n$ ) e o parâmetro  $k$ , conforme apresentado na Tabela 8.

**Tabela 8.** Valores iniciais, finais e incrementos das variáveis mensuradas.

Variáveis	Valor inicial	Valor final	Incremento
Número de APs ( $n$ )	2	8	1
$k$	1	13	2

A Tabela 8 mostra que o número de APs testados iniciou com valor  $n = 2$ , pois a utilização de apenas um AP evidentemente acarretaria numa baixa precisão, já que neste caso a probabilidade de repetição de RSSI seria alta; e o parâmetro  $k$  foi testado com valores ímpares de um até  $13 \cong \sqrt{160}$ .

Estes tratamentos permitem mensurar o melhor número de APs instalados e o número adequado de vizinhos mais próximos, ou seja, o melhor par  $(n, k)$ . Quando o par  $(n, k)$  de melhor precisão é definido com os menores valores possíveis, tem-se um menor consumo de hardware e um tempo de processamento reduzido, pois menos dados precisam ser coletados e tratados pelo algoritmo.

Nestes experimentos, foram realizadas  $160 \times 49 = 7840$  estimativas de localização para cada método de localização proposto (Método I e Método II). Após os testes, o desempenho das propostas deste trabalho foi comparado com os métodos 3-PCA apresentado de forma prática no trabalho feito por Salamah et al. (2016) e o método *powed*-sorensen (PS) apresentado de forma prática por J. Torres-Sospedra et al. (2015).

Devido as diferenças nos ambientes de testes utilizados pelos autores, tanto na dimensão do espaço e na quantidade de RPs, como na mobília disposta dentro do ambiente, o que dificulta compará-los. Para fins de comparação, estes métodos foram desenvolvidos e fisicamente testados no mesmo ambiente que os métodos propostos nesta dissertação.

Por ambos estimarem a localização nas coordenadas do RP majoritário, o método I foi comparado diante do método PS, que utiliza a função de similaridade Sørensen combinada com a representação de dados *powed*. E pelo fato de ambos estimarem a localização através do centróide dos RPs, o método II foi comparado com o método 3-PCA, que combina a distância euclidiana como função de similaridade com a PCA para representar o conjunto de dados.

As características dos métodos propostos neste trabalho e dos métodos apresentados na literatura, implementados para fins de comparação, são apresentadas na Tabela 9.

**Tabela 9.** Características dos métodos propostos e dos métodos apresentados na literatura.

Método		Função de similaridade	Representação dos dados	Obtenção das coordenadas
Propostas	I	Distância euclidiana	Análise de quartis	RP majoritário
	II	Distância euclidiana	Análise de quartis	Centróide dos RPs
Literatura	3-PCA	Distância euclidiana	PCA	Centróide dos RPs
	PS	Sørensen	<i>Powed</i>	RP majoritário

### 3.4.3 Medidas de avaliação de desempenho

Para avaliar a precisão dos métodos implementados, foi utilizado o erro médio de localização como medida de desempenho primário, tal medida auxiliará na análise do desempenho de todos os quatro métodos na Seção 5.1, quando o efeito do número de APs e do valor do parâmetro  $k$  não é considerado. E na Seção 5.2, quando o efeito do número de APs e do valor do parâmetro  $k$  é considerando.

O erro médio foi medido pela soma das distâncias entre a posição real e a posição estimada da WSTA e dividida pelo número total de testes correspondentes. Em geral, um EM mínimo é desejado. O erro médio (EM) de localização é dado em metros (m) e é calculado com base na seguinte equação:

$$EM = \frac{1}{N} \sum_{i=1}^N d(P_i, \hat{P}_i) \quad (4.8)$$

Onde: a distância entre  $P_i$  e  $\hat{P}_i$  é o erro de cada localização,  $P_i$  é a posição real da WSTA,  $\hat{P}_i$  é a posição estimada da WSTA e  $N$  é o número de testes correspondentes.

Também foi avaliado o tempo médio de processamento para localização, reduzindo o número de leituras ( $m$ ) de RSSI em cada amostra necessária para calcular os quartis dos APs ( $m = 20$ ,  $m = 15$ ,  $m = 10$  e  $m = 5$ , respectivamente).

O tempo médio (TM) é dado em segundos (s) e calculado conforme a Equação 4.9. Em que o tempo total pertinente à criação e classificação de instâncias de testes em um RP é dividido pelo número de testes correspondentes:

$$TM = \frac{1}{N} \sum_{i=1}^N t_i \quad (4.9)$$

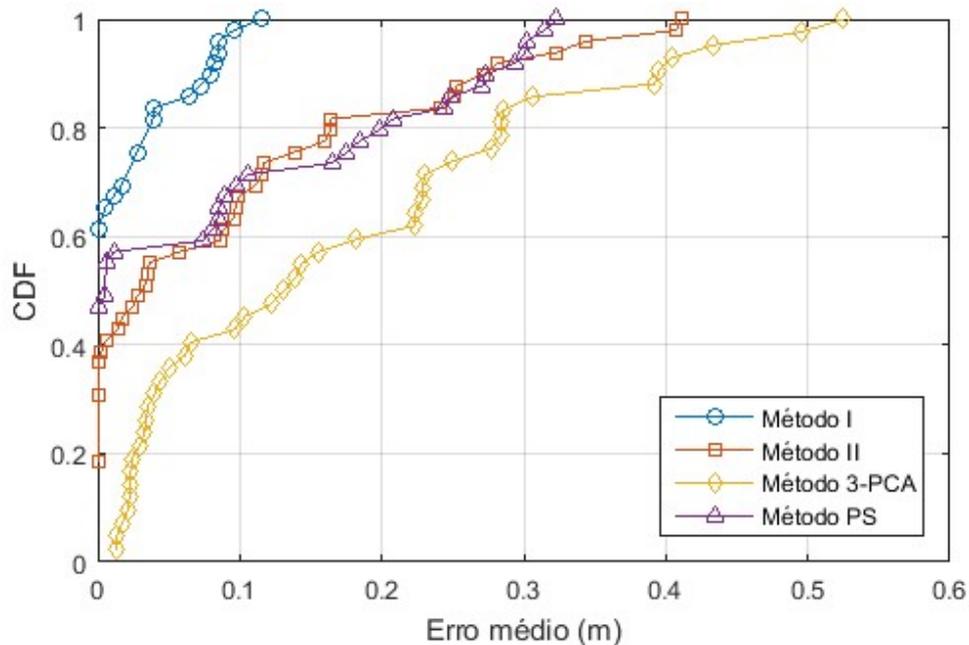
Onde:  $t_i$  é o tempo pertinente à criação e classificação de uma instância e  $N$  é o número de testes correspondentes.

Para isso, foi selecionado o esquema de localização que apresentou a melhor precisão coma melhor quantidade de APs ( $n$ ) instalados e o número adequado de vizinhos mais próximos ( $k$ ), verificados nos testes de precisão, a fim de validar um sistema de localização preciso e com obtenção de respostas em curto tempo.

## 4 Resultados e discussão

### 4.1 Função de distribuição acumulada do erro médio de localização

Primeiramente, para analisar o desempenho de todos os quatro métodos em todos os testes realizados, sem considerar o melhor número de APs instalados e o valor adequado para o parâmetro  $k$ , foi construído o gráfico da Função de Distribuição Acumulada (*Cumulative Distribution Function* – CDF) dos erros de localização.



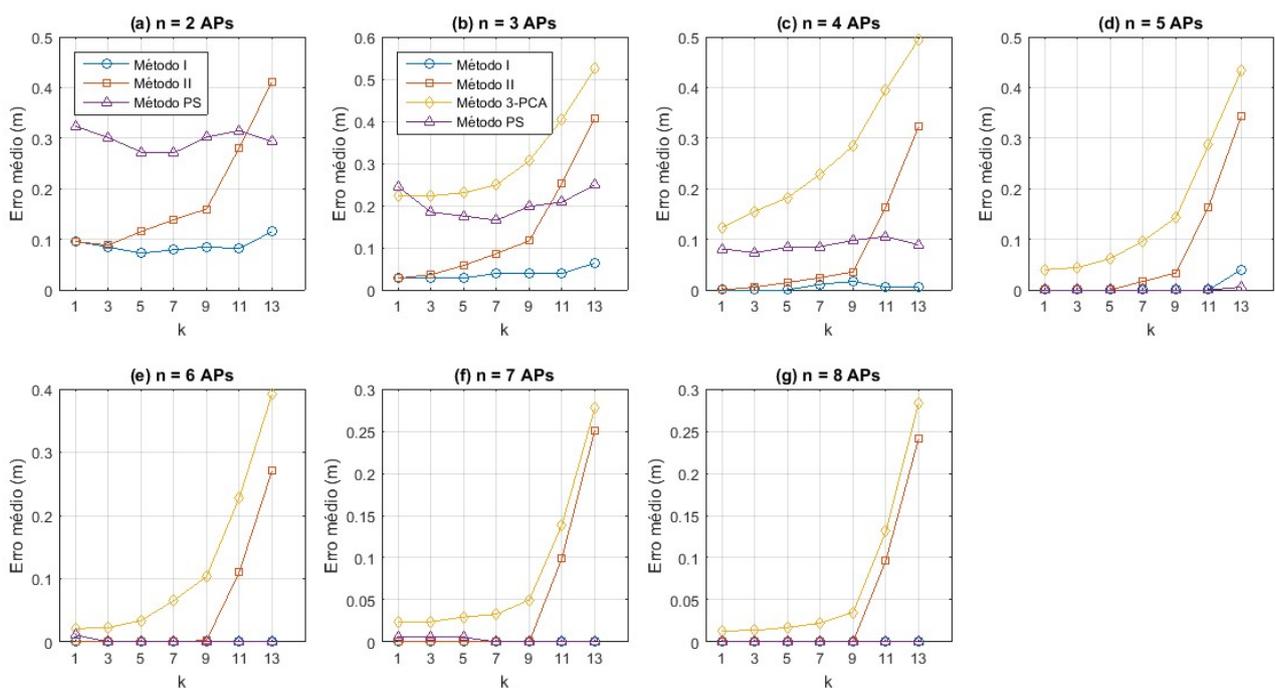
**Figura 10.** Função de Distribuição Acumulada (CDF) dos erros de localização (m).

A Figura 10 mostra a CDF dos erros de localização para os quatro métodos implementados. O método I produziu os melhores resultados, onde todas as estimativas apresentaram erro inferior a 0,12 metros, enquanto que os métodos II; 3-PCA e PS apresentaram erro menor que 0,12 metros em apenas 73,47; 45,24 e 71,43% dos testes, respectivamente. Observa-se também que o método PS alcançou um erro máximo inferior aos dos métodos II e 3-PCA, porém o erro máximo do método PS (0,3233 metros) é quase três vezes maior que o erro máximo do método I (0,1152 metros).

Ainda na Figura 10, é interessante observar que, os métodos II e 3-PCA obtiveram os piores desempenhos, tal resultado é possivelmente por ambos estimarem a localização através do centróide dos RPs. Neste caso, o erro médio de localização pode ser melhorado pela redução da distância entre RPs, utilizando-se mais RPs no ambiente. No entanto, muitos RPs implicam em mais esforços manuais para criar o conjunto de dados (Shin et al., 2012).

## 4.2 Erro médio de localização por número de APs e valor do parâmetro $k$

Em seguida, foram analisados o efeito do número de APs ( $n$ ) e a influência do número de vizinhos mais próximos ( $k$ ) nas estimativas de localização dos quatro métodos implementados. O método 3-PCA foi testado somente a partir de três APs, pois a transformação utilizada neste método necessita de no mínimo três atributos no conjunto de dados. O erro médio de localização para cada par  $(n, k)$  pode ser observado na Figura 11.



**Figura 11.** Erro médio das 160 estimativas de localização em função do parâmetro  $k$  para cada número de pontos de acesso ( $n$ ).

Na Figura 11, é possível ver que o erro médio diminui significativamente à medida que o número de APs aumenta para os métodos I e PS, mas quando o número de APs é  $n \geq 6$ , o erro médio se torna constante para o método I definido com qualquer  $k$ , enquanto que o erro médio de localização para o método PS ainda pôde ser melhorado com  $n = 8$ .

Os métodos II e 3-PCA, que utilizam a média das coordenadas dos RPs, também tiveram o erro médio reduzido à medida que o número de APs aumenta, porém há uma grande influência dos valores assumidos por  $k$ , pois à medida que este parâmetro aumenta os erros tendem a crescer.

Quanto ao menor erro médio de localização obtido e o melhor par  $(n,k)$  que define cada método, ambos os métodos propostos  $I(n = 4, k = 1)$  e  $II(n = 4, k = 1)$  obtiveram erro nulo conforme visto na Figura 11(c), o método  $PS(n = 5, k = 1)$  também obteve erro nulo conforme visto na Figura 11(d) e o método  $3-PCA(n = 8, k = 1)$  apresentou erro médio de 0,0134 metros conforme visto na Figura 11(g).

Os métodos I, II e PS também apresentaram erro nulo com muitos outros pares combinados  $(n,k)$ , porém os pares definidos anteriormente possuem os menores valores possíveis, o que implica num menor consumo de hardware e um tempo de processamento reduzido, pois menos dados precisam ser coletados e tratados pelo algoritmo.

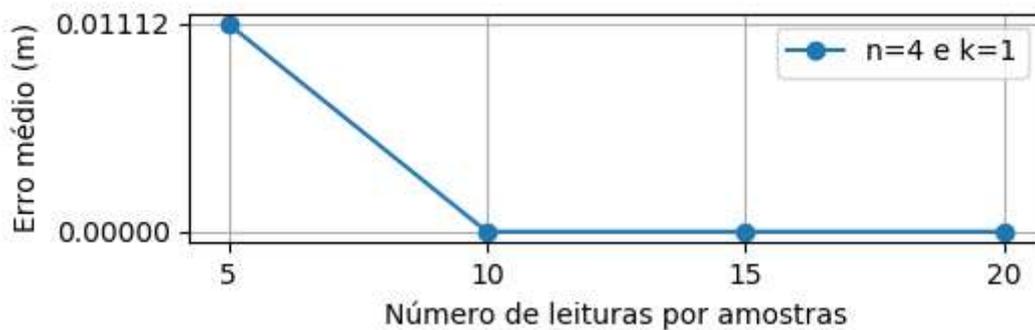
Nota-se que os métodos propostos I e II são equivalentes para qualquer valor de  $n$  com  $k = 1$ , pois a média das coordenadas de um único RP são as próprias coordenadas deste RP. Neste caso, o método I é uma simplificação do método II. Diante do exposto, o método  $I(n = 4, k = 1)$ , com abordagem da análise de quartis na representação dos dados, pode consistir em uma proposta promissora para localizar objetos em ambientes internos.

### 4.3 Tempo médio de processamento para localização

Com intuito de reduzir ainda mais o tempo de processamento mantendo o mesmo desempenho do método proposto, novos testes foram

realizados utilizando o método I com apenas quatro APs e com  $k$  igual a um. O método I ( $n = 4, k = 1$ ) implementado anteriormente utilizou  $m = 20$  leituras de RSSI para calcular os quartis de cada AP.

Para analisar o desempenho, três novos conjuntos de treinamentos e de testes foram criados com base em um número reduzido de leituras de RSSI ( $m$ ) por amostras, sendo estes valores: 15, 10 e 5. O resultado dos testes é apresentado na Figura 12.



**Figura 12.** Erro médio de localização com número reduzido de leituras de RSSI.

Estes resultados mostram que a utilização de amostras com apenas cinco leituras de RSSI reduz o desempenho do método. Porém a partir de amostras com tamanho  $m = 10$  não houve erros de localização, mantendo-se o erro nulo. O tempo médio foi verificado para os testes com  $m \geq 10$  e os resultados são apresentados na Tabela 10.

**Tabela 10.** Tempo de processamento para cada número de leituras de RSSI.

Número de leituras (m)	Tempo médio de processamento (s)
10	17,27
15	25,58
20	33,76

Após os testes realizados, verifica-se que o melhor tempo é alcançado com  $m = 10$  leituras por amostras, apresentando apenas 17,27 segundos, sendo este o menor tempo. Grande parte deste tempo, cerca de 95,89%,

envolve a varredura dos quatro APs e o envio dos dados pela WSTA. Portanto, o método proposto  $I(n = 4, k = 1)$ , com abordagem da análise de quartis na representação dos dados, é validado com o melhor desempenho e com obtenção de respostas em menor tempo de processamento.

## 5 Conclusão e trabalhos futuros

Nesta dissertação, foi proposto um método de localização baseado no algoritmo de aprendizado de máquina *k-Nearest Neighbors* (*k*NN), que verifica as coordenadas do ponto de referência (RP) majoritário, para estimar a localização do objeto móvel. O método projetado utiliza a análise de quartis na representação dos dados para melhorar a precisão de localização, que é comprometida pelos problemas de interferência no sinal Wi-Fi e instabilidade das medições brutas de RSSI, causados principalmente por obstáculos.

Foram realizados experimentos para avaliar o desempenho da proposta e o tempo de processamento para localização, variando-se o número de APs ( $n$ ), o parâmetro  $k$  e o tamanho de amostra ( $m$ ) necessária para o cálculo dos quartis, a fim de validar um sistema de localização de alto desempenho e rápido tempo de processamento.

O erro médio de localização da proposta deste trabalho é comparado com métodos de localização presentes na literatura para verificar a viabilidade do algoritmo proposto. Os resultados mostram que o desempenho do método proposto supera a localização interna dos métodos relatados na literatura. Durante os testes, na situação de melhor desempenho do método proposto, foi obtido erro de localização nulo e tempo de processamento de 17,27 segundos. Estes resultados foram obtidos com a utilização de quatro pontos de acesso (APs),  $k = 1$  e 10 leituras por amostras para o cálculo dos quartis.

Do ponto de vista da seleção de apenas uma instância de treinamento, definida por  $k = 1$ , os resultados revelaram que a análise de quartis representa de forma adequada os valores de RSSI, visto que, as instâncias adquirem atributos discriminantes. Verifica-se também que foi possível validar o método de localização adotando a análise de quartis a partir das amostras de apenas quatro APs.

Com base nestas observações, constata-se a possibilidade de realizar

localização de forma precisa em ambientes internos com problemas de instabilidade no RSSI do sinal Wi-Fi, o que valida o objetivo (Cap. 1.1) deste trabalho.

Como trabalhos futuros, propomos o desenvolvimento de um mecanismo de auto-adaptação para utilização de pontos de acesso (APs) já existentes no ambiente interno, de modo que o desligamento indesejado destes APs não interfira nos resultados de localização. O desligamento temporário pode resultar no aumento do tempo da localização devido ao retardo imposto por inconsistência de dados, ocasionada pela falta de aquisição de dados de RSSI dos APs durante o período de desligamento.

Adicionalmente propomos a utilização de técnicas de seleção de um único quartil com valor mais discriminante para representar o conjunto de dados com dimensão reduzida. A redução do conjunto de dados possibilita a implementação do sistema de localização, de maneira embarcada, em dispositivos com pouco espaço de armazenamento.

## Referências

- Al-Fuqaha, A., Guizani, M., Mohammadi, M., Aledhari, M. e Ayyash, M. Internet of Things: A Survey on Enabling Technologies, Protocols, and Applications. *IEEE Communication Surveys & Tutorials*, vol. 17, no. 4, 2015.
- Bahl, P. e Padmanabhan, V. N. RADAR: An In-Building RF-based User Location and Tracking System. *IEEE INFOCOM*, 2000.
- Bhattacharyya, A. On a measure of divergence between two statistical populations defined by their probability distributions. *Bull. Calcutta Math*, 1943.
- Bishop, C. M. Pattern Recognition And Machine Learning. *Springer Science+Business Media, LLC*, 2006.
- Cai, X., Li, X., Yuan, R., Hei, Y. Identification and mitigation of NLOS based on channel state information for indoor WiFi localization. *International Conference on Wireless Communications & Signal Processing (WCSP)*, Nanjing, 2015, pp. 1-5. doi: 10.1109/WCSP.2015.7341172.
- Chen, S. J., Chan G., He, J., Liew, S. C. Mixed-mode WLAN: the integration of ad hoc mode with wireless LAN infrastructure. *IEEE Global Telecommunications Conference*, San Francisco, CA, 2003, pp. 231-235 vol.1. doi: 10.1109/GLOCOM.2003.1258236.
- Cover, T. M. e Hart, P. E. Nearest Neighbor Pattern Classification. *Ieee Transactions On Information Theory*, vol. vol. 13, no. 1, 1967.
- Deng, J., Liang, B., Varshney, P. K. Tuning the Carrier Sensing Range of IEEE 802.11 MAC. *IEEE Global Telecommunications Conference, 2004 (GLOBECOM 04)*, Dallas, TX, 2004, pp. 2987-2991 vol.5. doi: 10.1109/GLOCOM.2004.1378900.
- Fang, H. S. e Lin, T. Principal component localization in indoor WLAN environments. *IEEE Trans. Mobile Comput.*, vol. 11, no. 1, pp. 100-110, 2012.

- Feng, X., Gao, Z., Yang, M. e Xiong, S. Fuzzy distance measuring based on RSSI in Wireless Sensor Network. *3rd International Conference on Intelligent System and Knowledge Engineering*, Xiamen, 2008, pp. 395-400. doi: 10.1109/ISKE.2008.4730962.
- Ghourchian, N., Allegue-Martinez, M. e Precup, D. Real-Time Indoor Localization in Smart Homes Using Semi-Supervised Learning. *Proceedings of the Twenty-Ninth AAAI Conference on Innovative Applications*, 2017.
- Graciolie, G., Fröhlich, A. A., Pires, R. P. e Wanner, L. Evaluation of an RSSI-based Location Algorithm for Wireless Sensor Networks. *Ieee Latin America Transactions*, vol. 9, no. 1, 2011.
- Isokawa, T., Motomura, S., Ohtsuka, J., Kawa, H., Nakashima, S., Kamiura, N. e Matsui, N. An Anchor-Free Localization Scheme with Kalman Filtering in ZigBee Sensor Network. *ISRN Sensor Networks*, vol. 2013, Article ID 356231, 11 pages, 2013. <https://doi.org/10.1155/2013/356231>.
- Jedari, E., Wu, Z., Rashidzadeh R. e Saif, M. Wi-Fi based indoor location positioning employing random forest classifier. *International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, Banff, AB, 2015, pp. 1-5. doi: 10.1109/IPIN.2015.7346754.
- Joarder A. H. e Firozzaman M. Quartiles for Discrete Data. *Teaching Statistics*. vol. 23, no 3, 2001.
- Kaemarungsi, K. e Krishnamurthy, P. Properties of Indoor Received Signal Strength for WLAN Location Fingerprinting. *Proceedings of the First Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services (MobiQuitous'04)*, 2004.
- Khullar, R. e Dong, Z. Indoor localization framework with WiFi fingerprinting. *26th Wireless and Optical Communication Conference (WOCC)*, Newark, NJ, 2017, pp. 1-6. doi: 10.1109/WOCC.2017.7928970.
- Kim, K. S., Wang, R., Zhong, Z., Tan, Z., Song, H., Cha, J. e Lee, S. Large-scale location-aware services in access: Hierarchical building/floor classification and location estimation using Wi-Fi fingerprinting based on deep neural networks. *International Workshop on Fiber Optics in Access*

- Network (FOAN)*, Munich, 2017, pp. 1-5. doi: 10.1109/FOAN.2017.8215259.
- Kuntal, A. e Karmakar, P. Range-Free Localization Algorithm for IEEE802.11 WLAN Based on K-NN. *Sixth International Conference on Computational Intelligence, Communication Systems and Networks*, Tetova, 2014, pp. 164-169. doi: 10.1109/CICSyN.2014.43.
- Labinghisa, B. e Lee, D. M. Indoor Localization Algorithm based on Predictive Path Learning in Wi-Fi Environment. *Proceedings of Symposium of the Korean Institute of communications and Information Sciences*. 2018.
- Langford, E. Quartiles in Elementary Statistics. *Journal of Statistics Education*, 2006, DOI: 10.1080/10691898.2006.11910589.
- Lantz, B. *Machine Learning with R*. Published by Packt Publishing Ltd., Ed. 2, 2015.
- Le, W., Wang, Z., Wang, J., Zhao, G., Miao, H. A novel WIFI indoor positioning method based on Genetic Algorithm and Twin Support Vector Regression, *The 26th Chinese Control and Decision Conference (2014 CCDC)*, Changsha, 2014, pp. 4859-4862. doi: 10.1109/CCDC.2014.6853043.
- MATLAB - The MathWorks Inc. Disponível em: <<https://www.mathworks.com/>>. Acesso em Setembro 21, 2018.
- McHugh, M. L. The Chi-square test of independence. *Biochemia Medica*, 2013; vol. 23, no 2. doi: <http://dx.doi.org/10.11613/BM.2013.018>.
- Mosteller, F. e Tukey, J. W. Data analysis and regression: a second course in statistics. *Addison-Wesley Series in Behavioral Science: Quantitative Methods*, 1977.
- Ni, Q. Performance analysis and enhancements for IEEE 802.11e wireless networks. *IEEE Network*, vol. 19, no. 4, pp. 21-27, Jul-Ago. 2005. doi: 10.1109/MNET.2005.1470679.
- NodeMCU Documentation. Disponível em: <<https://nodemcu.readthedocs.io/en/master/>>. Acesso em Outubro 01, 2018.
- NumPy. Disponível em: <<https://www.numpy.org/>>. Acesso em Janeiro 31,

2019.

Parwekar, P. e Reddy, R. An efficient fuzzy localization approach in Wireless Sensor Networks. *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, Hyderabad, 2013, pp. 1-6. doi: 10.1109/FUZZ-IEEE.2013.6622548.

Salamah, A. H., Tamazin, M., Sharkase, M. A. e Khedr, M. An enhanced WiFi indoor localization system based on machine learning, *International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, Alcalá de Henares, 2016, pp. 1-8. doi: 10.1109/IPIN.2016.7743586.

Salim, F., Williams, M., Sony, N., Pena, M. D., Petrov, Y., Saad, A. A., e Wu, B. Visualization of Wireless Sensor Networks using Zigbee's Received Signal Strength Indicator (RSSI) for Indoor Localization and Tracking. *The Second IEEE International Workshop on Social and Community Intelligence*, 2014.

Shin, B., Lee, J. H., Lee, T. e Kim, H. S. Enhanced weighted K-nearest neighbor algorithm for indoor Wi-Fi positioning systems. *8th International Conference on Computing Technology and Information Management (NCM and ICNIT)*, 2012.

Silva, A. J. S. As Tecnologias de Redes Wireless. *RNP News Generation*, vol. 2, no. 5, 1998.

Sørensen, T. A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on Danish commons. *K. Dan. Vidensk. Selsk. Biol. Skr.* 5: 1-34, 1948.

Tang, P., Huang, Z., Lei, J. e Guo, Y. Wi-Fi Fingerprint Localization Using RSSI-Probability Radio Map and AP Weight Clustering. *Journal of Advances in Computer Networks*. vol. 4, no. 2, pp. 121-124, 2016.

Torres-Sospedra, J., Montoliu, R., Trilles, S., Belmonte, O. e Huerta, J. Comprehensive analysis of distance and similarity measures for Wi-Fi fingerprinting indoor positioning systems. *Expert Systems With Applications*, 42, 9263–9278, 2015. doi: <https://doi.org/10.1016/j.eswa.2015.08.013>.

Witten, I. H. e Frank, E. *Data Mining: Practical Machine Learning Tools and*

Techniques with Java Implementations. *Morgan Kaufmann Publishers*, Ed. 2, San Francisco, California, 2005.

Wu, Z., Jedari, E., Liu, B., Rahidzadeh, R., Ahmadi, M. Particle Filter and Extreme Learning Machine Based Indoor Localization System. *IPSN*, Seattle, WA, 2015.

Zhao, J. e Wang, J. WiFi indoor positioning algorithm based on machine learning. *7th IEEE International Conference on Electronics Information and Emergency Communication (ICEIEC)*, Macau, 2017, pp. 279-283. doi: 10.1109/ICEIEC.2017.8076562.