

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA E
INFORMÁTICA INDUSTRIAL

EDUARDO LUIS GOMES

**ARQUITETURA RF-MINER - UMA SOLUÇÃO PARA
LOCALIZAÇÃO EM AMBIENTES INTERNOS**

DISSERTAÇÃO

CURITIBA

2017

EDUARDO LUIS GOMES

**ARQUITETURA RF-MINER - UMA SOLUÇÃO PARA
LOCALIZAÇÃO EM AMBIENTES INTERNOS**

Dissertação apresentada ao Programa de Pós-graduação em Engenharia Elétrica e Informática Industrial da Universidade Tecnológica Federal do Paraná como requisito parcial para obtenção do grau de “Mestre em Ciências” – Área de Concentração: Telecomunicações e Redes.

Orientador: Professor Mauro Sérgio Fonseca,
PhD

Co-orientadora: Professora Anelise Munaretto Fonseca,
PhD

CURITIBA

2017

Dados Internacionais de Catalogação na Publicação

G633ar Gomes, Eduardo Luis
2017 Arquitetura RF-Miner : uma solução para localização em
 ambientes internos / Eduardo Luis Gomes.-- 2017.
 66 f.: il.; 30 cm.

Disponível também via World Wide Web.
Texto em português com resumo em inglês.
Dissertação (Mestrado) - Universidade Tecnológica
Federal do Paraná. Programa de Pós-Graduação em Engenharia
Elétrica e Informática Industrial. Área de Concentração:
Telecomunicações e Redes, Curitiba, 2017.
Bibliografia: f. 62-66.

1. Sistemas de identificação por radiofrequência. 2.
Arquitetura de computador. 3. Mineração de dados (Computação).
4. Sistemas de posicionamento interno (localização sem fio).
5. Serviços baseados em localização. 6. Radiofrequência -
Identificação. 7. Sistemas de informação geográfica. 8.
Processamento eletrônico de dados em tempo real. 9. Métodos de
simulação. 10. Engenharia elétrica - Dissertações. I. Fonseca,
Mauro Sergio Pereira, orient. II. Fonseca, Anelise Munaretto,
coorient. III. Universidade Tecnológica Federal do Paraná.
Programa de Pós-Graduação em Engenharia Elétrica e Informática
Industrial. IV. Título.

CDD: Ed. 22 -- 621.3

Biblioteca Central do Câmpus Curitiba - UTFPR
Bibliotecária: Luiza Aquemi Matsumoto CRB-9/794

TERMO DE APROVAÇÃO DE DISSERTAÇÃO Nº _____

A Dissertação de Mestrado intitulada “**Arquitetura RF-MINER- Uma Solução para Localização em Ambientes Internos**” defendida em sessão pública pelo(a) candidato(a) **Eduardo Luis Gomes**, no dia 12 de dezembro de 2017, foi julgada para a obtenção do título de Mestre em Ciências, área de concentração Telecomunicações e Redes, e aprovada em sua forma final, pelo Programa de Pós-Graduação em Engenharia Elétrica e Informática Industrial.

BANCA EXAMINADORA:

Prof(a). Dr(a). Mauro Sergio Pereira Fonseca - Presidente – (UTFPR)

Prof(a). Dr(a). Fabrício Enembreck - (PUC-PR)

Prof(a). Dr(a). Richardson Ribeiro - (UTFPR)

A via original deste documento encontra-se arquivada na Secretaria do Programa, contendo a assinatura da Coordenação após a entrega da versão corrigida do trabalho.

Curitiba, 12 de dezembro de 2017.

A superação das dificuldades é a prova que forma o caráter.
Tony Robbins

RESUMO

GOMES, Eduardo L.. ARQUITETURA RF-MINER - UMA SOLUÇÃO PARA LOCALIZAÇÃO EM AMBIENTES INTERNOS. 68 f. Dissertação – Programa de Pós-graduação em Engenharia Elétrica e Informática Industrial, Universidade Tecnológica Federal do Paraná. Curitiba, 2017.

A utilização de etiquetas RFID UHF passivas para localização *indoor* vem sendo amplamente estudada devido ao seu baixo custo. Porém ainda existe uma grande dificuldade em obter bons resultados, principalmente devido à variação de rádio frequência em ambientes que possuem materiais reflexivos, como por exemplo, metais e vidros. Esta pesquisa propõe uma arquitetura de localização para ambientes *indoor* utilizando etiquetas RFID UHF passivas e técnicas de mineração de dados. Com a aplicação da arquitetura em ambiente real foi possível identificar a posição exata de objetos com a precisão de aproximadamente cinco centímetros e em tempo real. A arquitetura se demonstrou uma alternativa para implantação de sistemas de localização *indoor*, além de apresentar uma técnica de derivação de atributos diretos que contribui efetivamente para os resultados finais.

Palavras-chave: RFID, Localização, Ambientes Internos, Etiquetas Passivas, Mineração de Dados

ABSTRACT

GOMES, Eduardo L.. RF-MINER ARCHITECTURE - AN INDOOR LOCALIZATION SOLUTION. 68 f. Dissertação – Programa de Pós-graduação em Engenharia Elétrica e Informática Industrial, Universidade Tecnológica Federal do Paraná. Curitiba, 2017.

The use of passive UHF RFID tags for indoor location has been widely studied due to its low cost. However, there is still a great difficulty to reach good results, mainly due the radio frequency variation in environments that have materials with reflective surfaces, such as metal and glass. This research proposes a localization architecture for indoor environments using passive UHF RFID tags and data mining techniques. With the application of the architecture in real environment it was possible to identify the exact position of objects with the precision of approximately five centimeters and in real time. The architecture has demonstrated an efficient alternative for the implantation of indoor localization systems, besides presenting a derivation technique of direct attributes that contributes effectively to the final results.

Keywords: RFID, Localization, Indoor, Passive Tags, Data Mining

LISTA DE FIGURAS

FIGURA 1	– Sistema RFID	16
FIGURA 2	– Etiqueta RFID	17
FIGURA 3	– Classificação Etiquetas RFID	17
FIGURA 4	– Leitores RFID	20
FIGURA 5	– Antena RFID	22
FIGURA 6	– Mineração de dados	24
FIGURA 7	– Técnicas de Mineração de Dados	25
FIGURA 8	– Tipos de Classificadores	27
FIGURA 9	– Arquitetura RF-Miner	40
FIGURA 10	– Ambiente de Implementação	41
FIGURA 11	– Distâncias das Antenas	43
FIGURA 12	– Tupla Enviada Pelo Leitor	43
FIGURA 13	– Parâmetro Janela de Intervalos de Tempo	44
FIGURA 14	– Modelagem de Processos: Definição dos Parâmetros	45
FIGURA 15	– Dados organizados após processo: Extrator de Atributos Diretos	46
FIGURA 16	– Algoritmo: Derivador de Atributos	47
FIGURA 17	– Arquivo ARFF	49
FIGURA 18	– Validação Cruzada	51
FIGURA 19	– Algoritmo: Avaliar Base de Treinamento	51
FIGURA 20	– Código Fonte: Avaliar Base de Treinamento	52
FIGURA 21	– Algoritmo: Classificação em Tempo Real	53
FIGURA 22	– Código Fonte: Classificação em Tempo Real	53
FIGURA 23	– Resultados utilizando atributos diretos	55
FIGURA 24	– Comparativo Atributos diretos x Atributos diretos e derivados	56
FIGURA 25	– Evolução de Acertos após a troca de todos os livros	58
FIGURA 26	– Evolução de Acertos após a troca de dois livros	59
FIGURA 27	– Evolução de Acertos após a troca de seis livros	60
FIGURA 28	– Desempenho em relação à quantidade de antenas	61

LISTA DE TABELAS

TABELA 1	– Comparativo entre os trabalhos relacionados	38
TABELA 2	– Comparativo entre as metodologias dos trabalhos relacionados	38
TABELA 3	– Atributos Diretos	46
TABELA 4	– Atributos Diretos e Derivados	48
TABELA 5	– Atributos não relacionados diretamente à antena	48
TABELA 6	– Algoritmos de cada tipo de classificação	50
TABELA 7	– Classificação utilizando os atributos diretos (RSSI e <i>Read Count</i>)	54
TABELA 8	– Resultados utilizando atributos diretos e derivados	55
TABELA 9	– Comparativo Atributos diretos x Atributos diretos e derivados	56
TABELA 10	– Resultados em Tempo Real: Após 1ª Janela de Intervalos de Leitura ...	57
TABELA 11	– Resultados em Tempo Real: Após 2ª Janela de Intervalos de Leitura ...	57
TABELA 12	– Resultados em Tempo Real: Após 3ª Janela de Intervalos de Leitura ...	58
TABELA 13	– Evolução de Acertos após a troca de todos os livros	58
TABELA 14	– Evolução de Taxa de Acerto após troca de dois livros	59
TABELA 15	– Evolução de Acertos após a troca de seis livros	60
TABELA 16	– Desempenho em relação à quantidade de antenas	61

LISTA DE SIGLAS

RFID	Radio Frequency Identification
UHF	Ultra High Frequency
GPS	Global Positioning System
S-CRR	Swift Communication Range Recognition
RSSI	Received Signal Strength Indication
LF	Low Frequency
KHz	Quilo Hertz
HF	High Frequency
MHz	Mega Hertz
GHz	Giga Hertz
USB	Universal Serial Bus
WI-FI	Wireless Fidelity
ID3	Induction Decision Tree
CHAID	Chi-Squared Automatic Induction
CART	Classification and Regression Trees
RNA	Rede Neural Artificial
SVM	Sequential Minimal Optimization
CM-CRR	Continuous Moving-Communication Range Recognition
CRR	Communication Range Recognition
DR	Kalman-filter drift removal
LE	Heron-bilateration
dBi	Ganho de uma antena isotrópica
RC	Read Count
WEKA	Waikato Environment for Knowledge Analysis
GPL	General Public License
ARFF	Attribute-Relation File Format
M.E	Margem de Erro

SUMÁRIO

1	INTRODUÇÃO	12
1.1	MOTIVAÇÃO	12
1.2	OBJETIVOS	14
1.2.1	Objetivo Geral	14
1.2.2	Objetivos Específicos	14
1.3	ESTRUTURA DO DOCUMENTO	14
2	IDENTIFICAÇÃO POR RADIOFREQUÊNCIA	16
2.1	ETIQUETA	16
2.1.1	Classificação por fonte de energia	17
2.1.2	Classificação por frequência	18
2.2	LEITOR	19
2.3	ANTENAS	21
2.4	SITE SURVEY	21
2.5	APLICAÇÕES	22
3	MINERAÇÃO DE DADOS	24
3.1	TÉCNICAS DE MINERAÇÃO DE DADOS	25
3.1.1	Classificadores	26
3.1.1.1	Classificadores Baseados em Árvore de Decisão	27
3.1.1.2	Classificadores Baseados em Redes Neurais Artificiais	28
3.1.1.3	Classificadores Bayesianos	29
3.1.1.4	Classificadores Baseados em Vizinho Mais Próximo	30
3.1.1.5	Classificadores Baseado em Regras	31
3.1.1.6	Classificadores Baseado em Máquina de Vetor de Suporte	32
3.2	VALIDAÇÃO CRUZADA	33
4	TRABALHOS RELACIONADOS	35
5	ARQUITETURA RF-MINER	39
5.1	MÓDULO DE LEITURA FÍSICA	39
5.1.1	Preparação do Ambiente de Implementação	41
5.1.1.1	Definição dos Parâmetros de Leitura e Distâncias das Antenas	41
5.1.2	Realização das Leituras e Base Inicial de Treinamento	43
5.2	MÓDULO EXTRAÇÃO DE CARACTERÍSTICAS	44
5.2.1	Extrator de Atributos Diretos	45
5.2.2	Derivador de Atributos	46
5.2.3	Formatação da Base de Treinamento	48
5.3	MÓDULO MINERAÇÃO DE DADOS	48
5.3.1	Avaliando a Base de Treinamento	50
5.3.2	Classificação em Tempo Real	51
6	RESULTADOS E DISCUSSÕES	54
6.1	ATRIBUTOS DIRETOS EM RELAÇÃO AOS ATRIBUTOS DIRETOS EM CON- JUNTO AOS DERIVADOS	54
6.2	RESULTADOS EM TEMPO REAL	56

6.3	REDUÇÃO NA QUANTIDADE DE ANTENAS	60
7	CONCLUSÃO	62
	REFERÊNCIAS	64

1 INTRODUÇÃO

1.1 MOTIVAÇÃO

Atualmente os sistemas de localização e navegação estão muito presentes no cotidiano das pessoas devido aos benefícios que os mesmos podem oferecer. Um sistema de localização pode ser definido como um conjunto de dispositivos, técnicas, algoritmos e aplicações que trabalhando em sincronia estimam as coordenadas absolutas ou relativas de uma pessoa ou objeto num determinado ambiente de localização. A aplicação inicial que deu origem aos sistemas de localização foi o suporte a aplicações militares em ambientes externos (*outdoor*). A infraestrutura deste tipo de sistemas é usualmente formada por estações terrestres e vários satélites, possibilitando assim a obtenção geográfica da localização (BERZ; HESSEL, 2015).

Em relação à localização em ambientes externos (*outdoor*), a tecnologia *Global Positioning System* (GPS) tem avançado cada vez mais, tornando-se na prática o padrão no que se refere a navegação e localização de pessoas, veículos e demais objetos. Entretanto, a tecnologia GPS não é adequada para ambientes internos, pois, é necessário uma comunicação com visada direta para os satélites (BRÁS, 2009).

A localização de pessoas e objetos em ambientes internos (*indoor*) vem sendo amplamente pesquisada, especialmente em razão dos benefícios que um sistema de localização em tempo real pode fornecer como segurança contra furtos de produtos com alto valor agregado, agilidade de identificação e localização de pessoas, produtos ou documentos em grandes ambientes ou de armazenamento, posicionamento e navegação de robôs em ambientes fechados, além de diversas outras aplicações. Ambientes internos tem maior complexidade em relação a ambientes externos e possuem muitos desafios para superar. Os maiores problemas encontrados na implementação dos sistemas de localização em ambientes internos são: 1) a grande densidade de itens, onde exige-se uma margem de precisão muito grande, pois, em geral os itens são dispostos muito próximos uns dos outros; 2) a grande quantidade de obstáculos como paredes, móveis, equipamentos, plantas; 3) materiais reflexivos como metais, vidros e lâmpadas o que geram muitas interferências nos sinais de rádio frequência (AKRE et al., 2014).

Diante da dificuldade de utilização da tecnologia GPS para ambientes internos, muitas pesquisas de modelos de localização *indoor* tem utilizado a tecnologia *Radio Frequency Identification* (RFID) para sua implementação. A tecnologia RFID é um método de identificação automática que utiliza sinais de rádio para identificar, rastrear e gerenciar produtos, documentos ou até mesmo animais e pessoas sem nenhum tipo de contato físico e nem mesmo campo visual. Isto só é possível devido a utilização de dispositivos conhecidos como *transponders* ou *tags*, que são etiquetas eletrônicas, passivas, semi-passivas ou ativas, classificadas do ponto de vista da fonte de alimentação.

Modelos de localização indoor utilizando RFID são criados com vários objetivos diferentes e utilizando técnicas distintas de apuração da acuracidade das etiquetas. Existem pesquisas que têm como objetivo auxiliar a movimentação indoor de robôs utilizando como técnica o atributo S-CRR (*Swift Communication Range Recognition*) (NAKAMORI et al., 2012). Outros possuem o mesmo objetivo, porém utilizam técnicas probabilísticas de apuração (HORI et al., 2008). Da mesma maneira existem trabalhos que utilizam a mesma técnica S-CRR mas para atender objetivos diferentes, como o de identificar itens em um determinado local, ao invés de movimentação de robôs (UCHITOMI et al., 2010). O principal problema encontrado em todas as pesquisas descritas se refere à propagação de ondas de rádio frequência em ambientes internos, pois muitos materiais reflexivos podem ser encontrados nestes ambientes, o que torna a apuração da localização muito complexa devido a falta de padronização das leituras dos atributos, principalmente do *Received Signal Strength Indication* (RSSI - Indicador de Força do Sinal Recebido) que é o principal atributo utilizado para mensurar a distância entre os leitores RFID e as etiquetas eletrônicas. Esta falta de padronização nas leituras de um objeto alvo em um sistema de localização utilizando técnicas simples de triangulação em conjunto com o atributo RSSI, faz com que o objeto seja identificado em movimento, mesmo permanecendo parado no mesmo local.

Devido as dificuldades encontradas nos sistemas de localização em ambientes internos, o presente trabalho apresenta uma arquitetura de localização *indoor* utilizando a tecnologia RFID com etiquetas passivas em conjunto com técnicas de mineração de dados, demonstrando passo a passo a modelagem da mesma e sua aplicação em um ambiente real. Inicialmente a metodologia utilizada para desenvolver a arquitetura se concentrou em prover independência de atributos e hardware, sem depender exclusivamente de equipamentos de um fabricante. Foi identificado o melhor tempo de leitura, permitindo que todas as etiquetas fossem identificadas em pelo menos uma das antenas respeitando as características do local de implantação como distância das antenas. Após foram apuradas as localizações utilizando apenas os atributos diretos (RSSI e *Read Count*). Em seguida foi utilizada a técnica de derivação dos atributos em

janela de intervalos de leitura onde foram calculados e identificados para os atributos diretos os valores das Médias Aritméticas, Valores Mínimos, Valores Máximos e Desvios Padrão em uma janela de intervalos de leitura. Após, foi realizado um teste em relação a influência da quantidade de antenas na arquitetura e o que cada antena contribui ou interfere na acuracidade. Em todas as apurações foram utilizadas e comparadas as seguintes técnicas de mineração de dados: algoritmos de classificação baseados em árvore de decisão, baseados em redes neurais artificiais, bayesianos, vizinho mais próximo, baseado em regras e máquina de vetor de suporte.

1.2 OBJETIVOS

1.2.1 OBJETIVO GERAL

O presente trabalho tem como principal objetivo propor uma arquitetura de localização *indoor* capaz de obter a localização de objetos com a precisão de poucos centímetros (cinco centímetros aproximadamente) utilizando a tecnologia RFID com etiquetas passivas em conjunto com técnicas de mineração de dados.

1.2.2 OBJETIVOS ESPECÍFICOS

- Utilizar equipamentos disponíveis no mercado (*off-the-shelf*), independente de fabricante.
- Identificar e padronizar os atributos que mais contribuem para a acuracidade da localização.
- Avaliar o desempenho das diferentes técnicas de mineração de dados para identificar as quais apresentam melhores resultados.
- Otimizar a quantidade de antenas utilizadas no modelo para reduzir os custos de implantação.
- Reduzir ao máximo o tempo de leitura necessário para apresentar a localização satisfatória com a finalidade de utilizar a arquitetura para localização em tempo real.

1.3 ESTRUTURA DO DOCUMENTO

O Capítulo 2 apresenta os conceitos básicos dos sistemas RFID, como características dos Leitores, Antenas, Etiquetas e suas classificações. Técnicas de *Site Survey* para melhor

instalação e configuração dos equipamentos do sistema em cada ambiente, além das principais aplicações dos sistemas RFID. Este capítulo é importante na estrutura do documento, pois, a arquitetura proposta neste trabalho utiliza a tecnologia RFID para identificar os itens no ambiente de localização.

O Capítulo 3 demonstra as metodologias de mineração de dados e descreve as diferentes técnicas de cada um dos tipos de classificadores utilizados na implementação da arquitetura. São apresentados os Classificadores Baseados em Árvore de Decisão, Classificadores Baseados em Redes Neurais Artificiais, Classificadores Bayesianos, Classificadores Baseados em Vizinho Mais Próximo, Classificadores Baseado em Regras e Classificadores Baseado em Máquina de Vetor de Suporte. Além de como funciona a técnica de Validação Cruzada.

O Capítulo 4 contém alguns trabalhos relacionados que utilizam a tecnologia RFID para localizar ou posicionar objetos em ambientes internos e um comparativo entre os trabalhos em relação à arquitetura proposta na presente pesquisa.

O Capítulo 5 apresenta de forma conceitual o modelo da arquitetura proposta neste trabalho que está dividido em Módulo Físico, Módulo Extração de Características e Módulo Mineração de Dados e descreve detalhadamente a implementação da arquitetura em um ambiente real.

O Capítulo 6 apresenta os resultados obtidos na implementação da arquitetura, desde a formação da base de treinamento, os resultados em tempo real e a investigação utilizando diversas quantidades de antenas.

Por fim o Capítulo 7 apresenta as conclusões e comentários finais da dissertação juntamente com os trabalhos futuros.

2 IDENTIFICAÇÃO POR RADIOFREQUÊNCIA

Um sistema de identificação por radiofrequência, *Radio Frequency Identification* - RFID, é formado por componentes de negócio e componentes de infraestrutura (Figura 1). Os componentes de negócio são as regras de negócio, ou seja, a aplicação em si e o *middleware* o qual precisa ser integrado ao *software* de gestão para realizar a interface entre os componentes. Os componentes de infraestrutura são as etiquetas, os leitores e as antenas, as quais estão presentes tanto nos leitores quanto nas etiquetas (HESSEL et al., 2009; ROH et al., 2009; NATH et al., 2006; WANT, 2006). A arquitetura de localização *indoor* proposta neste trabalho possui três módulos, o Módulo de Leitura Física (Seção 5.1) que faz parte dos componentes de infraestrutura, o Módulo de Extração de Características e o Módulo Mineração de Dados (Seções 5.2 e 5.3) que fazem parte dos componentes de negócio.

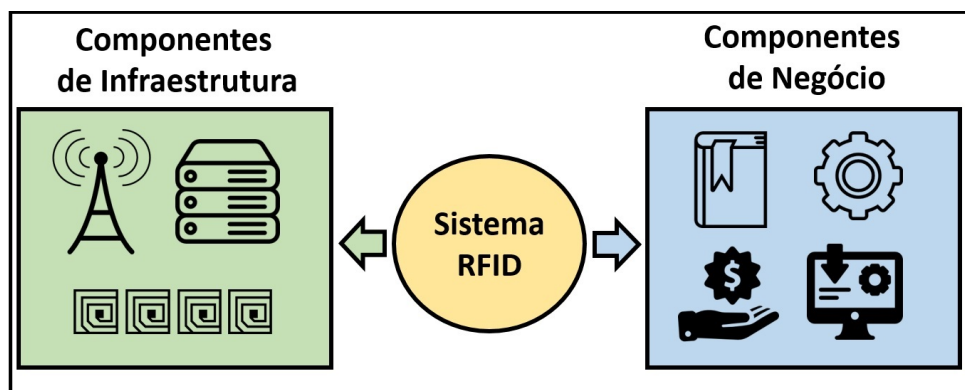


Figura 1: Sistema RFID

2.1 ETIQUETA

A etiqueta, contém dados que são transmitidos ao leitor no momento em que é lida. Em um sistema RFID a etiqueta deve ser anexada fisicamente sobre o objeto a ser localizado na regra de negócio. Cada etiqueta possui um mecanismo interno para armazenar dados e uma forma de comunicar estes dados (COONEY, 2006; PRERADOVIC et al., 2008; HESSEL et al., 2009).

Uma etiqueta, de forma geral, consiste de um circuito integrado, *microchip* que está conectado a uma antena, este *microchip* possui uma memória interna para armazenar dados e alguns processamentos lógicos, conforme demonstra a Figura 2. A memória de uma etiqueta pode ser somente de leitura, permitir uma gravação e várias leituras e de leitura e gravação. De forma geral, a capacidade de receber novos dados na memória, assim como a quantidade de dados que possam ser armazenados torna o custo das etiquetas maior. As etiquetas podem ser classificadas de acordo com a sua fonte de energia e frequência de operação, conforme demonstra a Figura 3 (CHEN; QING, 2010; HEIDRICH et al., 2010).

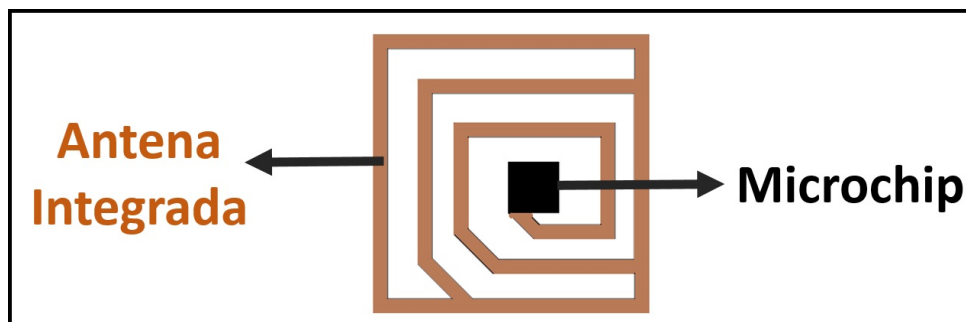


Figura 2: Etiqueta RFID

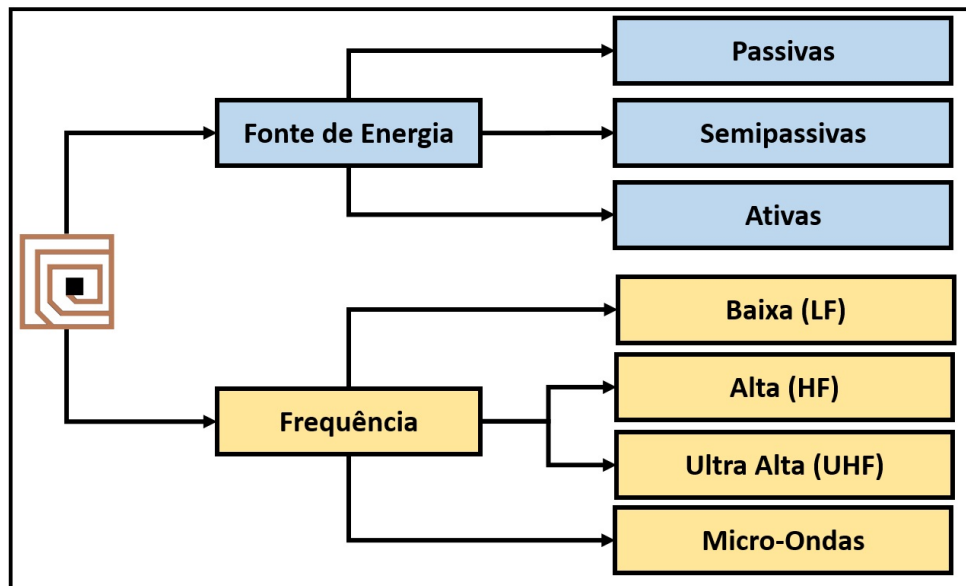


Figura 3: Classificação Etiquetas RFID

2.1.1 CLASSIFICAÇÃO POR FONTE DE ENERGIA

As etiquetas são classificadas de acordo com a fonte de alimentação existente na etiqueta, por exemplo as do tipo passivas não possuem fonte de alimentação de energia, enquanto

as do tipo ativas possuem fonte de alimentação própria. Existe ainda o tipo de etiqueta semi-passiva, que são híbridas dos outros tipos citados (HESSEL et al., 2009; CHEN; QING, 2010; KHAN et al., 2009):

- Etiquetas passivas: caracterizam por não necessitar de baterias internas, para sua ativação precisam estar na presença do campo eletromagnético produzido pelo leitor, sendo alimentadas por esta energia. A corrente elétrica que irradia na etiqueta possibilita ao *microchip* habilitar as funções de leitura, escrita e transmissão. São as mais econômicas e mais comuns na indústria de RFID, em sua maioria são pequenas e têm vida útil longa, pois não há necessidade de manutenção para o seu funcionamento. Apresentam como principal desvantagem o alcance restrito de leitura e a necessidade de alta potência do leitor. Com a finalidade de reduzir custos de implantação e viabilizar financeiramente projetos de localização *indoor*, neste trabalho é utilizado este tipo de etiqueta na implantação da arquitetura proposta.
- Etiquetas semipassivas: são híbridas das etiquetas passivas e ativas. Esse tipo de etiqueta possui uma bateria de baixo custo que é usado para alimentar os circuitos elétricos internos, porém apenas operam quando recebem uma potência do leitor, ou seja, não tomam a iniciativa de começar uma transmissão, apenas o fazem quando o leitor inicia o processo. Possuem tamanhos variados dependendo do alcance e por causa de sua bateria interna são mais espessas que as etiquetas passivas, necessitam manutenção para troca de baterias e seu custo acaba sendo mais elevado.
- Etiquetas ativas: possuem um transmissor e uma bateria interna que fornece energia para a comunicação da etiqueta. Podem significativamente melhorar o alcance da comunicação entre a etiqueta e o leitor. Possibilitam armazenar uma quantidade de informação maior do que as etiquetas passivas e semipassivas. Apresentam uma extensão de leitura longa e podem executar algumas funções na ausência de um leitor, ou ainda podem tomar a iniciativa de começar uma comunicação com algum leitor próximo para continuamente enviar informação sobre a situação de algum evento. Requerem manutenção para eventual substituição de bateria, seu custo devido suas características é o mais elevado.

2.1.2 CLASSIFICAÇÃO POR FREQUÊNCIA

Um aspecto importante para que ocorra a comunicação entre etiqueta e leitor é a frequência de operação, que pode variar com base na aplicação, nas normas e nas regulamen-

tações. Tanto etiqueta quanto leitor precisam operar na mesma frequência e as faixas mais comuns utilizadas nestes sistemas são (HESSEL et al., 2009; WARD et al., 2006):

- Baixa Frequência, *Low Frequency* (LF), em 135 KHz ou menos;
- Alta Frequência, *High Frequency* (HF), em 6,78 MHz, 13,56 MHz, 27,125 MHz e 40,680 MHz;
- Ultra Alta Frequência, *Ultra High Frequency* (UHF), em 433,920 MHz, 869 MHz e 915 MHz;
- Micro-ondas, *Micro Waves*, em 2,45 GHz, 5,8 GHz e 24,125 GHz.

Em geral a frequência define a taxa de transferência dos dados, velocidade, entre a etiqueta e o leitor, por isso quanto menor a frequência, mais lenta é a taxa de transmissão de dados. As condições ambientais e tipo de objeto em que a etiqueta será fixada precisa ser observado para determinar qual frequência será utilizada (WARD et al., 2006).

O tamanho da antena da etiqueta também é determinado pela frequência de utilização, em frequências mais altas o comprimento de onda é menor, assim as antenas são menores, resultando em etiquetas menores (WARD et al., 2006; HESSEL et al., 2009).

2.2 LEITOR

O leitor, de modo geral, é um equipamento de rádio capaz de amplificar e enviar sinais de radiofrequência por meio da utilização de uma antena, receber resposta da etiqueta e organizar os dados recebidos e armazenar por um pequeno tempo até poder enviar para um computador. Os leitores são alimentados por fonte de energia externa, ligada continuamente na tomada ou bateria (NATH et al., 2006; WANT, 2006, 2004).

As informações das etiquetas são utilizadas quando o leitor realiza a comunicação com um programa específico para a aplicação que estará instalado em um computador. Computador e leitor RFID se comunicam por meio de interfaces de conexão serial RS232, USB, Ethernet, WI-FI ou Bluetooth (HESSEL et al., 2009).

Os leitores são especificados de acordo com a aplicação e podem ser portáteis de posição fixa ou embarcados/embutidos (Figura 4), suas características são (WANT, 2004; NATH et al., 2006; ROH et al., 2009):

- Leitores portáteis: apresentam grande mobilidade e flexibilidade de uso, porém possuem antenas pequenas, o que diminui a distância máxima de leitura. São leitores de proximidade utilizados por exemplo em aplicações com brincos de animais, cartões de transporte público ou leitura e gravação de dados em produtos de prateleira e estoque. Estes leitores podem possuir tela de cristal líquido para visualização dos dados ou se comunicar com computadores por meio de interfaces *Bluetooth*.
- Leitores de posição fixa: são posicionados em locais fixos como esteiras, portais de carregamento ou locais de passagem de materiais para controle em tempo real. Permitem a ligação de até quatro antenas, o que amplia a área de cobertura de um único leitor. São conectados continuamente a computadores por meio de interfaces seriais ou ethernet para processamento dos dados coletados. Este é o tipo de leitor utilizado na arquitetura proposta neste trabalho.
- Leitores embutidos: são montados em uma mesma placa de circuito impresso de outros equipamentos ou módulos acoplados fisicamente a outros equipamentos de maneira interna de acordo com a aplicação.



Figura 4: Leitores RFID

A frequência de operação de um leitor deve ser considerada, pois sua conexão com as etiquetas depende disto. A frequência de operação pode variar de aplicação para aplicação, mas a mesma frequência deve ser utilizada tanto para o leitor quanto para as etiquetas. A frequência determina a taxa de transferência de dados entre etiqueta e leitor, então em sistemas que utilizam baixas frequências a velocidade será menor do que em sistemas que utilizam altas frequências (WANT, 2004; WARD et al., 2006). Neste trabalho tanto etiquetas quanto leitores operam em faixa de frequência UHF.

2.3 ANTENAS

Uma antena, de forma geral, é uma estrutura de transição entre uma onda guiada por um circuito e uma onda no espaço, o contrário também é válido. As antenas transmitem e/ou recebem ondas eletromagnéticas que se enquadram nas faixas de frequência para sistemas de rádio. As antenas são os condutores da comunicação de dados entre a etiqueta e o leitor e desempenham importante função, pois sua quantidade e posicionamento representam um fator decisivo na determinação da área de cobertura, alcance e desempenho na comunicação (ROH et al., 2009; HESSEL et al., 2009; NATH et al., 2006).

Do ponto de vista de aplicação, em um sistema de comunicação, uma antena opera como um dispositivo de junção entre o sistema guiado e o meio de propagação, o contrário também é verdadeiro, visando a transferência de energia entre o transmissor e o receptor. A antena deve direcionar a transmissão ou captar a recepção em direções desejadas (NATH et al., 2006).

É muito importante observar a zona de interrogação, que de forma geral, é a área em que a antena do leitor consegue ativar e se comunicar com a etiqueta. Vários fatores precisam ser considerados no ajuste da zona de interrogação para não prejudicar o desempenho garantir a eficiência do sistema RFID tais como, a quantidade de etiquetas, a potência na saída do leitor, a direção em que as antenas estão posicionadas, o tipo de antena utilizado, se são direcionais ou omni-direcionais e os tipos de materiais (e.g. reflexivos, opacos, geradores de ruídos, etc) que podem interferir na propagação do sinal de radiofrequência (WU et al., 2013; FOSTER; JAEGER, 2007; HESSEL et al., 2009).

A transmissão de radiofrequência para as etiquetas pode ser realizada por meio de uma antena e a recepção por outra antena, este tipo de configuração é conhecida como configuração biestática. Quando uma única antena realiza a transmissão e a recepção, dizemos que a configuração é monoestática. As antenas utilizadas neste trabalho operam de forma monoestática (HESSEL et al., 2009; CHEN; QING, 2010), a Figura 5 demonstra uma imagem de uma antena RFIF monoestática.

2.4 SITE SURVEY

Um *Site Survey* é uma verificação que permite analisar o ambiente em que o sistema RFID será implantado para certificar-se que a comunicação entre os dispositivos utilizados ocorra de forma adequada e sem falhas. Com isso é feita a verificação das possíveis fontes



Figura 5: Antena RFID

de interferência que possam vir a gerar algum problema no desempenho e na eficiência do sistema como um todo . É um processo muito importante na implantação de qualquer sistema RFID, pois, através do mesmo são definidos a distância das antenas em relação às etiquetas alvo, o tempo de leitura para cada interrogação e a potência que será utilizada para realizar as leituras das etiquetas (WU et al., 2013; HESSEL et al., 2009).

2.5 APLICAÇÕES

Conforme a tecnologia RFID progride e os custos de implementação diminuem, seu uso em potencial tem avanço e é incorporado nas atividades diárias das pessoas, o que proporciona sua aplicação em diversas áreas (HESSEL et al., 2009).

Dentre as variadas aplicações dos sistemas RFID algumas podem ser citadas como, rastreo e identificação animal (VOULODIMOS et al., 2010; FINKENZELLER, 2010; CHAWLA; HA, 2007), controle e rastreo de documentos (HULL et al., 2005; ABUSAFIYA; MAZUMDAR, 2004), emissão de documentos (e.g. passaportes) (STANTON, 2008; KOSCHER et al., 2009), gestão e rastreamento de containers em áreas portuárias (CHEN, 2005; GAUKLER; SEIFERT, 2007), acesso controlado de ambientes (RIEBACK et al., 2005; CHIEN, 2006), controle de estoque de materiais de alto valor agregado (e.g. construção civil, aviação, automotiva, mineração, etc.) (CHAO et al., 2010; KELEPOURIS et al., 2007, 2006; WANG et al., 2007; LU et al., 2011; SCHMITT et al., 2007; NASSAR; VIEIRA, 2014), identificação de pacientes e equipamentos hospitalares (NASSAR; VIEIRA, 2014; BACHELDOR, 2008; CHOWDHURY;

KHOSLA, 2007; SANI et al., 2010), *smart cards* (FINKENZELLER, 2010; NASSAR; VIEIRA, 2014), cancelas automáticas em praças de pedágio (NASSAR; VIEIRA, 2014; XIAO et al., 2008), segurança antifurto e inventário no comércio (e.g. roupas, livros, jogos, calçados, etc) (AYRE, 2012; ATZORI et al., 2010), gestão da cadeia de suprimentos (SARAC et al., 2010; FOSTER; JAEGER, 2007; HESSEL et al., 2009; NASSAR; VIEIRA, 2014) e localização de itens em ambientes internos (NASSAR; VIEIRA, 2014; JIN et al., 2006; SAAB; NAKAD, 2011). Outros exemplos de aplicações em que o RFID pode ser utilizado podem ser encontrados (FINE et al., 2006; SPEKMAN; SWEENEY, 2006; HESSEL et al., 2009).

Sistemas RFID apresentam como característica a versatilidade de aplicação, desde sistemas com baixo consumo de energia, em que a natureza de sua aplicação é simples e possui zona de leitura próxima, até sistemas que consomem mais energia, por armazenar maior volume de dados e necessitar maiores cuidados devido ao alcance e complexidade nas zonas de leitura (HESSEL et al., 2009; SARMA et al., 2002; SHIH et al., 2006).

Desta forma, a aplicação da tecnologia RFID tanto na indústria como na academia é um campo amplo para estudos e investigação científica, pois independente de quão avançada esteja sua utilização nos mais variados segmentos, ainda há muito potencial a ser explorado.

Para sistemas de localização, o cálculo simples de triangulação dos dados recebidos das antenas já seria suficiente para estimar a posição de uma etiqueta, mas as dificuldades encontradas em ambientes internos descritas na seção 1.1 (grande densidade de itens, obstáculos como paredes, móveis, equipamentos, plantas e materiais reflexivos como metais, vidros e lâmpadas), interferem na padronização das leituras, fazendo com que um item seja identificado em movimento mesmo permanecendo parado na sua posição original. O presente trabalho utiliza dentro dos componentes de aplicação, técnicas de mineração de dados para solucionar esta falta de padronização encontrada nas leituras das etiquetas. O seguinte capítulo (Capítulo 3) demonstra as metodologias de mineração de dados e descreve as diferentes técnicas de cada um dos tipos de classificadores utilizados na implementação da arquitetura proposta neste trabalho.

3 MINERAÇÃO DE DADOS

A mineração de dados *Data Mining* permite analisar grandes volumes de dados como forma de encontrar e estabelecer relações que possam ser compreensíveis e aplicáveis (Figura 6). Por tratar de conhecimentos multidisciplinares, permite unir técnicas de armazenamento e recuperação de dados, estatística e reconhecimento de padrões para encontrar tendências e padrões existentes nos dados. Sua aplicação está relacionada a cenários específicos de previsão, risco, recomendações, probabilidade, localizações, agrupamento, entre outros (WITTEN et al., 2016b; FAYYAD et al., 1996).

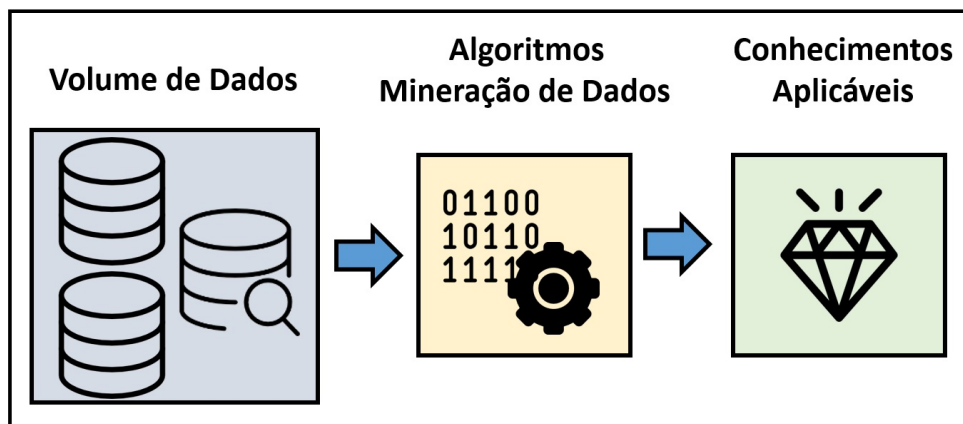


Figura 6: Mineração de dados

A mineração de dados constitui parte de um processo maior de descoberta de conhecimento em repositórios de dados. Trata-se de um processo para conversão, por meio de passos de transformação de dados brutos em informações úteis. No processo são 4 etapas (HAND et al., 2001; WITTEN et al., 2016b):

- Seleção: obtenção dos dados necessários para buscar e identificar os padrões e conhecimentos desejados de acordo com as regras de negócio;
- Pré-processamento e Transformação: tratamento dos dados para eliminar valores que não são necessários para o processamento, se houver necessidade realizar a transformação dos dados para diminuir sua dimensionalidade;

- Mineração: extração dos padrões por meio da utilização das técnicas de mineração de dados;
- Interpretação: análise dos resultados e sua aplicação do conhecimento obtido de acordo com as regras de negócio;

Na arquitetura proposta neste trabalho, os processos de Seleção e Pré-Processamento são realizados no Módulo Extração de Características (Seção 5.1) e os processos de Mineração e Interpretação no Módulo de Mineração de Dados (Seção 5.2).

A mineração de dados aplica metodologias e técnicas que possibilitam realizar a análise sistemática e exaustiva sobre um grande volume de dados. As técnicas de mineração são organizadas de forma a agir sobre grandes repositórios de dados com o intuito de descobrir padrões úteis e recentes, que de outra forma poderiam permanecer ignorados. As técnicas também fornecem capacidade de previsão do resultado de uma observação futura.

3.1 TÉCNICAS DE MINERAÇÃO DE DADOS

De uma forma geral, cinco técnicas são utilizadas para utilizar a mineração de dados, são elas: 1) análise de associação, 2) análise de agrupamentos, 3) detecção de anomalias, 4) classificação descritiva e 5) classificação preditiva, demonstradas na Figura 7.

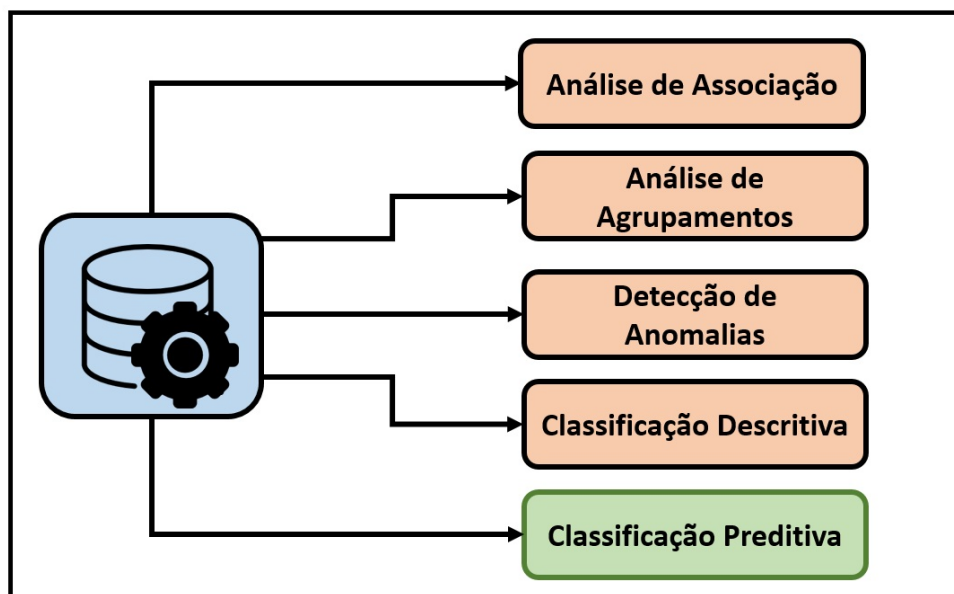


Figura 7: Técnicas de Mineração de Dados

1. **Análise de Associação:** é usada para descobrir padrões que descrevam características fortemente associadas dentro dos dados. Os padrões descobertos são normalmente repre-

sentados na forma de regras de implicação ou subconjuntos de características. Devido ao tamanho exponencial do seu espaço de busca, o objetivo da análise de associação é extrair os padrões mais interessantes de uma forma eficiente (TAN et al., 2009; LAROSE, 2014)

2. **Análise de Agrupamentos:** procura encontrar grupos de observações intimamente relacionados de modo que as observações que pertençam ao mesmo grupo sejam mais semelhantes entre si do que com as que pertençam a outros grupos. Esta técnica busca similaridades entre os dados para definir um conjunto finito de classes ou categorias que os contenha e os descreva. Por ser o objetivo dessa implementação a tarefa de encontrar agrupamentos, denomina-se como técnica de agrupamento ou segmentação (TAN et al., 2009)
3. **Detecção de Anomalias:** procura identificar observações cujas características sejam diferentes dos demais dados, são conhecidas como anomalias ou fatores estranhos. Seu objetivo é detectar as anomalias verdadeiras e evitar classificar com erro objetos normais como anômalos (TAN et al., 2009; HAND et al., 2001).
4. **Classificação Descritiva:** busca derivar padrões para evidenciar os relacionamentos entre os dados e requerem, além do pré-processamento, técnicas de pós-processamento para explicar os dados, principalmente quando sua natureza tem cunho exploratório (LAROSE, 2014; TAN et al., 2009).
5. **Classificação Preditiva:** tem como propósito prever o valor de um determinado atributo baseado nos valores de outros atributos. O atributo a ser previsto é conhecido como a variável dependente ou alvo, enquanto que os atributos usados para fazer a previsão são conhecidos como as variáveis independentes ou explicativas (LAROSE, 2014; TAN et al., 2009). No modelo preditivo emprega-se um esquema de aprendizagem de classificação, em que a partir de um conjunto de dados de treinamento, espera-se classificar amostras não conhecidas nos dados de teste (WITTEN et al., 2016b; TAN et al., 2009). Esta foi a técnica utilizada na implementação da arquitetura proposta no presente trabalho. A seção 8 apresenta detalhadamente os tipos de classificadores existentes.

3.1.1 CLASSIFICADORES

Existem seis tipos de classificadores, 1) os classificadores baseados em árvore de decisão, 2) os classificadores baseados em redes neurais artificiais, 3) os classificadores bayesianos, 4) os classificadores baseados em vetores de suporte, 5) os classificadores de vizinho mais próximo e 6) os classificadores baseados em regras, demonstrados na Figura 8 (TAN et al., 2009).

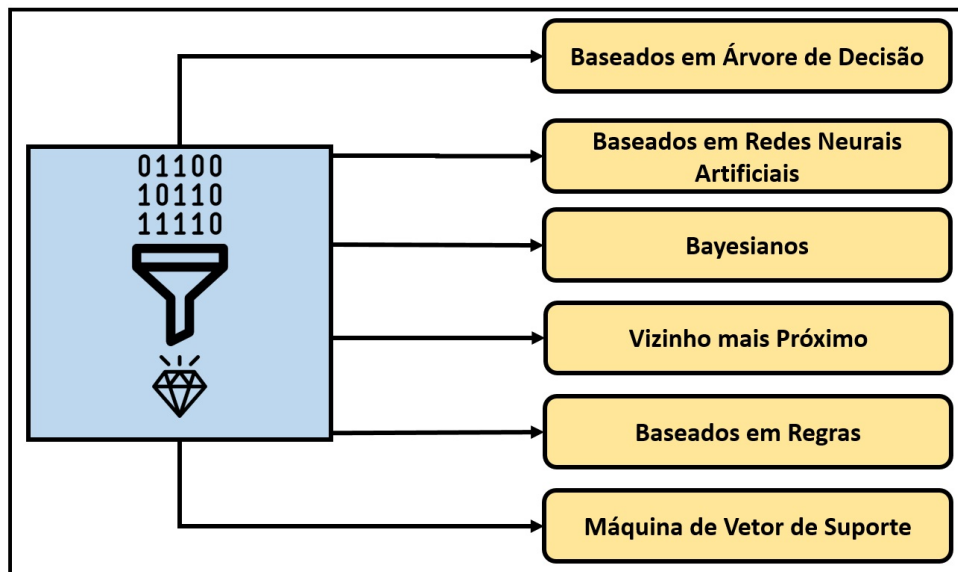


Figura 8: Tipos de Classificadores

3.1.1.1 CLASSIFICADORES BASEADOS EM ÁRVORE DE DECISÃO

Um classificador de árvore de decisão, consiste em uma técnica de classificação baseada em algoritmos de particionamento sucessivo. Também conhecida como indução de árvore de decisão, decorre do princípio de dividir para conquistar para produzir árvores de decisão. A partir da informação contida nos dados, cada ramo da árvore surge em função de um questionamento de classificação e cada folha é uma partição do conjunto de dados contendo a sua respectiva classificação. Uma árvore de decisão é formada pelo nó raiz, arestas e outros nós internos e terminais, que resultam da separação de registros com características diferentes (TAN et al., 2009).

Dado um conjunto de dados com uma das variáveis como objeto de saída, o algoritmo encontra o atributo mais fortemente relacionado à variável de saída, determinando-o como o primeiro ramo chamado de raiz. Os demais atributos são subsequentemente classificados como nós até que se chegue ao último nível, a folha. Uma vez construída a árvore de decisão, um dado novo pode ser classificado por ela em um de seus subgrupos, desde que, iniciando seu trajeto no nó raiz da árvore, este seja dirigido para a direita ou esquerda a cada nó de acordo com a regra de decisão associada ao mesmo até chegar a um nó folha da árvore (TAN et al., 2009).

Os algoritmos de árvore decisão aprendem com os exemplos, gerando uma árvore de decisão e um conjunto de regras associadas. Algumas árvores de decisão obtidas podem ser mais precisas que outras, visto que encontrar a árvore ótima seria computacionalmente inviável por causa do tamanho exponencial do espaço de pesquisa. Apesar disto, algoritmos eficientes têm sido desenvolvidos para induzir a uma árvore de decisão razoavelmente precisa, embora

não perfeita (TAN et al., 2009).

Estes algoritmos ainda incluem a poda da árvore, ou seja, o algoritmo de desenvolvimento da árvore é parado antes de gerar uma árvore totalmente desenvolvida, mas que satisfaça perfeitamente todos os dados de treinamento. Entre estes existem o algoritmo de Hunt, o algoritmo *Induction Decision Tree* - ID3, o algoritmo C4.5, o algoritmo *Chi-Squared Automatic Induction* - CHAID e o algoritmo *Classification and Regression Trees* - CART (TAN et al., 2009; WITTEN et al., 2016b; HAN et al., 2011).

Uma das otimizações encontradas no algoritmo C4.5 e versões recentes como o J48 é a introdução de um novo critério para este tipo de seleção, definido como taxa de ganho, onde uma normalização é efetuada sobre o ganho de informação para ajustar testes com muitas saídas (WITTEN et al., 2016b). O algoritmo J48 é um dos algoritmos utilizados na implementação da arquitetura proposta neste trabalho (Tabela 6).

3.1.1.2 CLASSIFICADORES BASEADOS EM REDES NEURAIAS ARTIFICIAIS

Uma Rede Neural Artificial - RNA é composta de um conjunto de nós interligado e arestas direcionadas. É um conjunto de elementos de processamento, agrupados em diversas topologias e regido por procedimentos matemáticos, como a clusterização de vetores, a otimização discreta, a minimização de erros, entre outros (HAYKIN, 2001; RUSSEL; NORVIG, 2004).

As redes neurais artificiais constituem-se em modelos computacionais paralelos baseados numa unidade atômica, o neurônio. Em geral, estes modelos possuem inspiração neurobiológica, porém, na prática, são algoritmos computacionais representando, de maneira bastante elementar, o mecanismo de funcionamento cerebral (HAYKIN, 2001).

Uma rede neural artificial é composta por várias unidades de processamento, que geralmente são conectadas por canais de comunicação associados a determinado peso. As unidades fazem operações apenas sobre seus dados locais, que são entradas recebidas pelas suas conexões. O comportamento inteligente de uma rede neural artificial vem das interações entre as unidades de processamento da rede (HAYKIN, 2001; RUSSEL; NORVIG, 2004).

A maioria dos modelos de redes neurais possui alguma regra de treinamento, onde os pesos de suas conexões são ajustados de acordo com os padrões apresentados. Em outras palavras, as RNAs aprendem por meio de exemplos. Esta capacidade de aprender e de generalizar a informação aprendida credita as RNAs características de adaptabilidade, generalização e tolerância a ruídos, dentre outras (HAYKIN, 2001; RUSSEL; NORVIG, 2004).

Outra característica importante das redes neurais artificiais é a capacidade de auto-organização, em que a rede atribui um padrão de entrada a uma classe dentre um conjunto de classes conhecidas. Este procedimento é denominado de algoritmo de treinamento e sua função principal corresponde a modificação dos pesos sinápticos visando atingir o objetivo (HAYKIN, 2001).

3.1.1.3 CLASSIFICADORES BAYESIANOS

Os Classificadores Bayesianos são classificadores estatísticos que classificam um objeto numa determinada classe, baseando-se na probabilidade deste objeto pertencer a esta classe. Quando aplicados a grandes volumes de dados, apresentam desempenho comparável aos resultados produzidos por árvores de decisão e redes neurais (HAN et al., 2011).

Para descrever como o teorema de Bayes é utilizado na classificação, torna-se necessário formalizar o problema do ponto de vista estatístico. Suponha X e Y um par de variáveis aleatórias, sua probabilidade $P(X = x, Y = y)$, se refere à probabilidade da variável X receber o valor x e a variável Y receber o valor y . Uma probabilidade condicional é a de que uma variável aleatória receba um determinado valor dado que o resultado de outra variável aleatória seja conhecido (TAN et al., 2009; HAN et al., 2011).

O teorema de Bayes pode ser utilizado para classificação de uma perspectiva estatística, onde X denota o conjunto de atributos e Y a variável de classe. Se esta tiver um relacionamento não determinístico com os atributos, então se torna possível tratar X e Y como variáveis aleatórias e obter seu relacionamento utilizando probabilisticamente $P(Y|X)$. Esta probabilidade condicional também é conhecida como probabilidade posterior de Y , em oposição a sua probabilidade anterior ou prévia ou a priori, $P(Y)$ (TAN et al., 2009; HAN et al., 2011).

Durante a fase de treinamento, torna-se necessário descobrir as probabilidades posteriores $P(Y|X)$ para cada combinação X e Y baseada em informações coletadas a partir dos dados de treinamento. Conhecendo estas probabilidades, um registro de teste X' pode ser classificado encontrando-se a classe Y' que maximize a probabilidade posterior, $P(Y'|X')$. Para avaliar as probabilidades condicionais de classe $P(X|Y)$, será apresentado o método de implementação de classificação bayesiano denominado BayesNet.(HAN et al., 2011; TAN et al., 2009).

O classificador BayesNet utilizado na implementação da arquitetura proposta neste trabalho (Tabela 6) consiste na implementação de um método de classificação bayesiano, assim chamado por assumir que os atributos são condicionalmente independentes, ou seja, a informação de um evento não interfere na informação de nenhum outro (HAN et al., 2011). Apesar desta

premissa simplista e pouco realista, o classificador BayesNet se apresenta eficaz em muitos casos práticos de aprendizado supervisionado, apresentando inclusive o melhor desempenho em várias tarefas de classificação quando comparado a outros classificadores com diferentes abordagens. Aliás, o classificador BayesNet apresenta um bom desempenho, apesar da hipótese de independência condicional ser frequentemente violada. Mas sem esse tipo de suposição em algum nível o problema se torna intratável (HAN et al., 2011).

Os classificadores bayesianos são robustos para pontos de ruídos isolados, porque calculam a média de tais pontos ao avaliar probabilidades condicionais a partir de dados. Também manipulam valores faltantes, ignorando o exemplo durante a construção e classificação do modelo (TAN et al., 2009; RUSSEL; NORVIG, 2004).

3.1.1.4 CLASSIFICADORES BASEADOS EM VIZINHO MAIS PRÓXIMO

Um classificador de vizinho mais próximo *Nearest Neighbor* - NN tem por objetivo encontrar todos os exemplos de treinamento que sejam relativamente semelhantes aos atributos do exemplo de teste. Estes exemplos são conhecidos como vizinhos mais próximos e podem ser utilizados para determinar o rótulo da classe como exemplo de teste. Uma desvantagem óbvia da abordagem relaciona-se ao fato de que alguns registros de testes podem não ser classificados por não corresponderem a nenhum exemplo de treinamento. (RUSSEL; NORVIG, 2004; TAN et al., 2009).

Um classificador de vizinho mais próximo representa cada exemplo como um ponto dado em um espaço d -dimensional, onde d é o número de atributos dado um exemplo de teste, calcula-se sua proximidade com o resto dos pontos de dados no conjunto de treinamento. Os vizinhos mais próximos K , onde K representa a quantidade de vizinhos mais próximos de um determinado exemplo de testes. Os classificadores de vizinho mais próximo fazem suas previsões baseadas em informações locais, sendo as decisões de classificação tomadas localmente, tornando-os suscetíveis a ruídos para valores pequenos de k (RUSSEL; NORVIG, 2004; TAN et al., 2009).

Portanto, assim que a lista de vizinhos mais próximos for obtida pelo algoritmo de classificação, o exemplo de teste é classificado baseado na classe majoritária dos seus vizinhos mais próximos (TAN et al., 2009).

O algoritmo *Kstar* utilizado na arquitetura proposta neste trabalho (Tabela 6) consiste em um classificador baseado em exemplos, onde a classe de um exemplo teste é baseada no treinamento de classes de exemplos similares a ele, determinado por alguma função de simi-

laridade. O *Kstar*, implementa o classificador de vizinho mais próximo, utilizando funções de distância baseadas na entropia. Classificadores baseados em exemplo tais como *Kstar*, assumem que os exemplos similares terão classes similares (HALL et al., 2009; CLEARY et al., 1995).

3.1.1.5 CLASSIFICADORES BASEADO EM REGRAS

Um classificador baseado em regras consiste naquele, cuja técnica para classificar registros baseia-se em um conjunto de regras “se...então”. A expressividade de um conjunto de regras é praticamente equivalente a uma árvore de decisão, porque pode ser representada por um conjunto de regras completas e mutuamente excludentes. Tanto os classificadores de árvores de decisão quanto os baseados em regras criam partições retilíneas do espaço de atributos e atribuem uma classe a cada partição (TAN et al., 2009).

A ordenação das regras pode ser implementada regra a regra ou classe a classe. O esquema de ordenação baseado em regras garante que cada registro de teste seja classificado pela melhor regra que o cobrir. Uma potencial desvantagem seria que regras com menor prioridade são mais difíceis de interpretar, porque supõem a negação das regras que as precedem (WITTEN et al., 2016b).

No esquema de ordenação baseado em classes, as regras que pertençam à mesma classe aparecem juntas no conjunto de regras R . As regras muitas vezes são ordenadas coletivamente com base na sua informação de classe. A ordenação relativa entre as regras da mesma classe não é importante; desde que uma das regras seja disparada, a classe será atribuída ao registro de teste. Isto torna a interpretação das regras um pouco mais fácil. No entanto, existe a possibilidade de uma regra de qualidade alta não ser percebida em favor de uma regra inferior que faça o prognóstico da classe de prioridade mais alta (TAN et al., 2009; WITTEN et al., 2016b).

Os classificadores baseados em regras mais conhecidos são o C4.5rules, JRip e RIPPER *Repeated Incremental Pruning to Produce Error Reduction* ou poda incremental repetida para produzir redução de erro. (WITTEN et al., 2016b; TAN et al., 2009)

Existe ainda um método para gerar um conjunto de regras a partir de uma árvore de decisão. No princípio, cada caminho a partir do nó raiz até o nodo folha de uma árvore de decisão pode ser expresso como uma regra de classificação. As condições de teste encontradas pelo caminho formam os conjuntos de antecedentes da regra, enquanto que o rótulo da classe no nodo folha é atribuído a sequência da regra (TAN et al., 2009; WITTEN et al., 2016b).

3.1.1.6 CLASSIFICADORES BASEADO EM MÁQUINA DE VETOR DE SUPORTE

Um classificador baseado em máquina de vetor de suporte consiste em uma técnica fundamentada na teoria de aprendizagem estatística. A máquina de vetores de suporte, do inglês *Support Vector Machine* - SVM foi desenvolvida com o intuito de resolver problemas de classificação de padrões. Esta técnica originalmente desenvolvida para classificação binária, busca a construção de um hiperplano como superfície de decisão, de tal forma que a separação entre exemplos seja máxima. Isso considerando padrões linearmente separáveis (VAPNIK, 2013).

A técnica SVM pertence a uma categoria de classificadores lineares e foi desenvolvida com o objetivo de auxiliar na solução de problemas de classificação e no reconhecimento de padrões. O seu conceito é baseado na ideia de minimizar riscos estruturais, ou seja, minimizar erros da classificação empírica e maximizar a margem geométrica entre os resultados. O SVM é um algoritmo linear que constrói hiperplanos como superfícies de decisão de maneira que a fronteira de separação entre classes positivas e negativas seja maximizada (VAPNIK, 2013).

O algoritmo processa geometricamente um hiperplano no espaço, o qual separa pontos que representam instâncias positivas da categoria dos pontos que representam instâncias negativas. O hiperplano é escolhido durante a etapa de treinamento como único e com margens máximas definidas. A margem consiste na distância do hiperplano entre o ponto mais próximo dos conjuntos positivos e negativos. Os hiperplanos são escolhidos por meio de um sub-conjunto das instâncias de treinamento que são chamados de vetores de suporte. Apenas parte dos dados de treinamento é utilizada para escolha do hiperplano, sendo que a outra parte não influencia nessa escolha. O classificador SVM tem como ponto forte a rapidez de execução, pois não depende da dimensionalidade do espaço amostral (VAPNIK, 2013).

Devido a sua eficiência em trabalhar com dados de alta dimensionalidade, tem sido reportada na literatura como uma técnica altamente robusta, muitas vezes comparada às redes neurais (SUNG; MUKKAMALA, 2003).

Um algoritmo eficiente para implementação da técnica SVM seria o de Otimização Sequencial Mínima *Sequential Minimal Optimization* - SMO, cuja utilização de memória é linear para realizar os treinamentos. Com isso, o SMO permite lidar com grande quantidade de dados para treinamento, além de prover a implementação da técnica SVM por meio do uso de polinômios ou núcleos gaussianos (PLATT, 1998).

O algoritmo SMO escolhe a resolução dos problemas de otimização, optando pelas menores otimizações possíveis em cada passo. Nos problemas de programação quadrática em

SVM, a menor otimização possível envolve os multiplicadores de Lagrange, porque eles devem obedecer à restrição de igualdade linear. Em cada passo, o método SMO escolhe a otimização dos dois multiplicadores de Lagrange, buscando valores ótimos para eles e atualizando-os para refletir os novos valores ótimos. A vantagem está em utilizar um otimizador analítico, ao invés de ser chamada toda uma biblioteca de rotinas de programação quadrática. Além disso, não é necessário armazenamento de matrizes extras, o que permite manipular problemas com conjunto de treinamento volumoso (PLATT, 1998).

O algoritmo SMO consiste na sequência de três componentes, 1) o método analítico para solucionar os dois multiplicadores, 2) uma heurística para escolher quais multiplicadores otimizados, 3) um método para computar um dos parâmetros da margem de decisão. No funcionamento do SMO, ocorre o cômputo das restrições de apenas dois multiplicadores de Lagrange escolhidos. Em seguida, soluciona-se o problema de maximização das restrições (PLATT, 1998).

3.2 VALIDAÇÃO CRUZADA

A validação cruzada é uma abordagem utilizada para validar os modelos de classificação construídos pelos classificadores a partir da amostragem contida na base de classificação supervisionada. Uma ferramenta padrão em estatística conhecida como *cross-validation* e que se apresenta como uma alternativa a subamostragem aleatória (WITTEN et al., 2016b; ARLOT et al., 2010).

O conjunto de dados disponível é particionado aleatoriamente em um conjunto de treinamento e um conjunto de teste. O conjunto de treinamento é depois particionado em dois conjuntos disjuntos: o subconjunto de estimação e o subconjunto de validação. O subconjunto de estimação é usado para estimação do modelo, isto é, treinamento, já o subconjunto de teste é utilizado para avaliar o desempenho do modelo (WITTEN et al., 2016b; ARLOT et al., 2010).

O particionamento pode ser feito a partir de cada registro, utilizado o mesmo número de vezes para treinamento e exatamente uma vez para teste. Ao generalizar o método, a validação cruzada de várias partes, segmenta os dados em várias partições com o mesmo tamanho. Durante cada execução, uma das partições é escolhida para teste, enquanto as outras são utilizadas para treinamento. Este procedimento é repetido várias vezes de modo que cada partição seja utilizada exatamente uma vez para teste (WITTEN et al., 2016b; ARLOT et al., 2010).

Uma rodada da validação cruzada envolve o particionamento de uma amostra de dados em subconjuntos complementares, executando a análise de um subconjunto, denominado

conjunto de treinamento, e validando a análise em outro subconjunto, denominado em conjunto de teste. Para aumentar a variabilidade em múltiplas, a validação cruzada é executada diversas vezes com partições diferentes, em que o resultado da validação é a média de todas as execuções (WITTEN et al., 2016b; ARLOT et al., 2010).

Assim, a validação cruzada é uma ferramenta padrão para análise, sendo um importante recurso para ajudar a desenvolver e ajustar os modelos de mineração de dados. A validação cruzada é utilizada depois de se criar uma estrutura de mineração e os modelos de mineração relacionados para assegurar a validade do modelo. Sendo utilizada em aplicações, onde o objetivo é a predição para estimar o quão correto um modelo preditivo deverá ser executado na prática. As Tabelas 7 e 8 apresentadas no capítulo 6 demonstram duas investigações utilizando a técnica de validação cruzada. Estas investigações tiveram o objetivo de identificar o comportamento dos atributos diretos comparando com a utilização dos atributos diretos em conjunto aos atributos derivados.

4 TRABALHOS RELACIONADOS

A exemplo do que ocorre em diversas áreas, a motivação na utilização da tecnologia RFID para criação de sistemas de localização e posicionamento é muito utilizada em diversos trabalhos na literatura.

(NAKAMORI et al., 2012) apresenta um novo método de estimativa de posicionamento indoor de etiquetas RFID para sistemas de navegação contínua.

No sistema RFID, um leitor permite estimar a posição das etiquetas facilmente. O problema é como conseguir as estimativas de posição das etiquetas RFID de forma rápida e mais precisa.

Neste trabalho foi proposto uma nova forma de estimar a posição de etiquetas que utilizam a tecnologia RFID para robôs indoor, esse método é chamado CM-CRR (*Continuous Moving-Communication Range Recognition*), o qual usa intervalos longos e curtos de comunicação, com movimentação contínua de um robô móvel. O CM-CRR estima a posição de uma etiqueta em quatro etapas e só utiliza tecnologia RFID, ou seja, sem WI-FI, ultra-som ou qualquer outro sensor, o que o torna teoricamente barato e simples.

Para a realização do experimento foi investigado o erro de estimativa por meio de simulações primárias, usando o modelo de movimento direto e o modelo de curva, a fim de comprovar a diferença da utilização do CM-CRR para o S-CRR. Os resultados mostram que a estimativa precisa da posição das etiquetas em robôs internos é totalmente viável. O trabalho não realizou experimentos em ambientes reais, apenas em ambiente controlado.

(HORI et al., 2008) apresentou um método de intervalo de detecção múltipla para a estimativa de posição de etiquetas RFID.

Os sistemas RFID se tornaram uma nova fonte de identificação em ambientes omni-presentes, porque os objetos possuem etiquetas com ID exclusiva. Em funcionamento o usuário lê o ID's das etiquetas e obtém as informações sobre o objeto.

Neste trabalho houve uma comparação entre o método S-CRR e o método CRR, que

tem por finalidade propor um novo método para estimar a posição dos leitores RFID passivos, mostrando a eficácia do método proposto, para isso foi avaliado o erro de posição estimado e por fim foram obtidos alguns resultados significativos, um deles foi que quando o sistema detecta algumas etiquetas RFID, a posição do robô móvel pode ser estimada com mais precisão e utilizando esse método o sistema permite a estimativa de posição do leitor RFID em pouquíssimo tempo.

Já (AKRE et al., 2014) demonstra um sistema de localização preciso em 2-D de etiquetas RFID usando controle de energia de transmissão de antena.

Neste trabalho é proposto um método para a localização em 2-D de etiquetas RFID, envolvendo 70 etiquetas passivas UHF. Visto que a localização de objetos marcados com precisão é um problema em muitas aplicações industriais.

O método é baseado na definição de assinaturas de localização e cálculos de distâncias usando o algoritmo k-NN ele utiliza uma função agregada do RSSI para todas as possíveis potências de transmissão do leitor RFID.

O resultado obtido da precisão da localização em comparação com os resultados obtidos usando um único valor de RSSI foi muito significativo e o que o torna de extrema importância é que as principais contribuições deste trabalho são provenientes de resultados baseados em dados experimentais reais, porém em ambiente controlado.

(ZHANG et al., 2015) apresenta um modelo de localização de etiquetas RFID com base na análise experimental de RSSI.

Neste trabalho foi explorado o uso de estatísticas RSSI para caracterizar a posição de uma etiqueta RFID passiva em relação a um leitor, para localizar um objeto rotulado por uma etiqueta.

Para isso foram estudadas várias características das respostas das etiquetas e foi realizado uma campanha de medição em um ambiente real o qual foi estabelecido para uma determinada posição e orientação da etiqueta.

Como resultado, as estatísticas do RSSI para diferentes potências de transmissão podem ser sintetizadas através de um triplo (A, B, P) onde A e B são, a inclinação estimada e o P a interceptação estimada da função linear que relaciona o valor RSSI médio e a transmissão do leitor.

(TING et al., 2011) apresenta um trabalho que teve como objetivo estudar a viabilidade de empregar etiquetas passivas de RFID para a detecção de localização de objetos para fornecer

informações em tempo real para rastrear o movimento.

As conclusões são resumidas em quatro partes, onde destacam-se algumas colocações, começando pela tendência do sistema de posicionamento de RFID que tem sido amplamente aplicada para rastreamento de mercadorias. Em seguida foi analisado o desenvolvimento de um sistema de posicionamento *indoor* com baixo custo e o uso de etiquetas passivas. É proposto pelos autores como uma forma de economizar, pois o custo por unidade de etiquetas passivas é mais baixo em relação a etiquetas ativas.

Por fim, como resultado foi possível perceber que o sistema utilizando as etiquetas RFID é totalmente satisfatório, viável e é uma solução mais econômica quando comparada com outros posicionamentos tecnológicos, o que facilita a aceitação da indústria.

(HUANG et al., 2015) apresenta um sistema de posicionamento *indoor* de RFID em tempo real com base na DR (*Kalman-filter drift removal*) e na LE (*Heron-bilateration*).

O modelo de filtro Kalman é utilizado como base no desenvolvimento de dispositivos de posicionamento interno RFID, o qual é portátil, escalável e econômico. O modelo pode ser aplicado para melhorar a maioria dos métodos DR convencionais, para qualquer tipo de sistema de posicionamento interno de rádio.

Este artigo teve como finalidade propor a remoção do filtro de Kalman (DR) e a estimativa de localização de Heron-Bilateration, para reduzir o índice de erro do RSSI, complexidade computacional e custo da implantação do RFID. Para isso foram utilizados um leitor RFID portátil, como o dispositivo direcionado, e um par de etiquetas RFID ativas, como marcos implantados na borda. Os resultados revelaram que o método de filtro de Kalman é rápido e apresenta bons resultados para convergir o erro de medição de distância.

A Tabela 1 apresenta um comparativo entre os trabalhos relacionados e a arquitetura proposta no presente trabalho. É possível perceber que a maioria dos trabalhos utilizou ambiente controlado dentro de laboratórios para realizar os testes e que nenhum dos trabalhos foi implantado em ambiente real.

A Tabela 2 apresenta um comparativo entre as metodologias utilizadas na criação das arquiteturas. É possível perceber que o atributo RSSI está presente na maioria dos trabalhos relacionados.

Tabela 1: Comparativo entre os trabalhos relacionados

Trabalho	Fixação	Etiqueta	Precisão	Tempo Real	Ambiente
Arquitetura RF-Miner	Livros	Passiva	5 cm	SIM	Real
(NAKAMORI et al., 2012)	Robôs	Passiva	20 cm	SIM	Simulador
(HORI et al., 2008)	Robôs	Passiva	100 cm	SIM	Simulador
(AKRE et al., 2014)	Tecido	Passiva	20 cm	NÃO	Controlado
(ZHANG et al., 2015)	Etiqueta	Passiva	20 cm	SIM	Controlado
(TING et al., 2011)	Pessoas	Passiva	100 cm	SIM	Controlado
(HUANG et al., 2015)	Cartão Portátil	Ativa	100 cm	SIM	Controlado

Tabela 2: Comparativo entre as metodologias dos trabalhos relacionados

Trabalho	Metodologia
Arquitetura RF-Miner	Classificadores e Derivação de Atributos (RSSI e Read Count)
(NAKAMORI et al., 2012)	CM-CRR e Intervalos de Tempo
(HORI et al., 2008)	S-CRR e CRR
(AKRE et al., 2014)	KNN e RSSI
(ZHANG et al., 2015)	RSSI, Inclinação, Interceptação e Média
(TING et al., 2011)	RSSI e Distância Euclidiana
(HUANG et al., 2015)	Filtro Kalmann e Bilateralização (RSSI)

5 ARQUITETURA RF-MINER

A arquitetura de localização *indoor* utilizando etiquetas RFID UHF passivas proposta neste trabalho é denominada RF-Miner. A metodologia utilizada para desenvolver a arquitetura é baseada em prover independência de atributos e hardware, sem depender exclusivamente de equipamentos de um fabricante específico, além disso prioriza a utilização de equipamentos disponíveis no mercado (*off-the-shelf*), não necessitando de desenvolvimento de equipamentos especiais.

Conforme descrito no Capítulo 2, um sistema de identificação por rádio frequência é formado por componentes de infraestrutura e componentes de negócio, a arquitetura RF-Miner está subdividida em 3 módulos: Módulo de Leitura Física que faz parte do componente de infraestrutura, Módulo de Extração de Características e Módulo Mineração de Dados que fazem parte dos componentes de negócio. A Figura 9 demonstra o modelo conceitual da arquitetura RF-Miner.

Com o objetivo de validar e implementar a arquitetura de localização *indoor* RF-Miner em um ambiente real, foi utilizado como estudo de caso a Biblioteca Conselheiro Mafra da Universidade do Contestado – Campus Mafra. A Figura 10 demonstra o ambiente de implementação com os componentes RFID instalados. As seções 5.1, 5.2 e 5.3 apresentam detalhadamente como foram implementados cada um dos módulos existentes na arquitetura.

5.1 MÓDULO DE LEITURA FÍSICA

Os componentes de infraestrutura básicos de sistema RFID são as etiquetas, antenas e leitores. O leitor normalmente está conectado a um computador com capacidade suficiente para processar os dados recebidos da etiqueta. O Módulo de Leitura Física realiza as leituras e repassa a informação para o Módulo de Extração de Características.

Para o perfeito funcionamento do Módulo de Leitura Física o processo de preparação do ambiente é muito importante. A instalação dos equipamentos, qualidade de cabos e

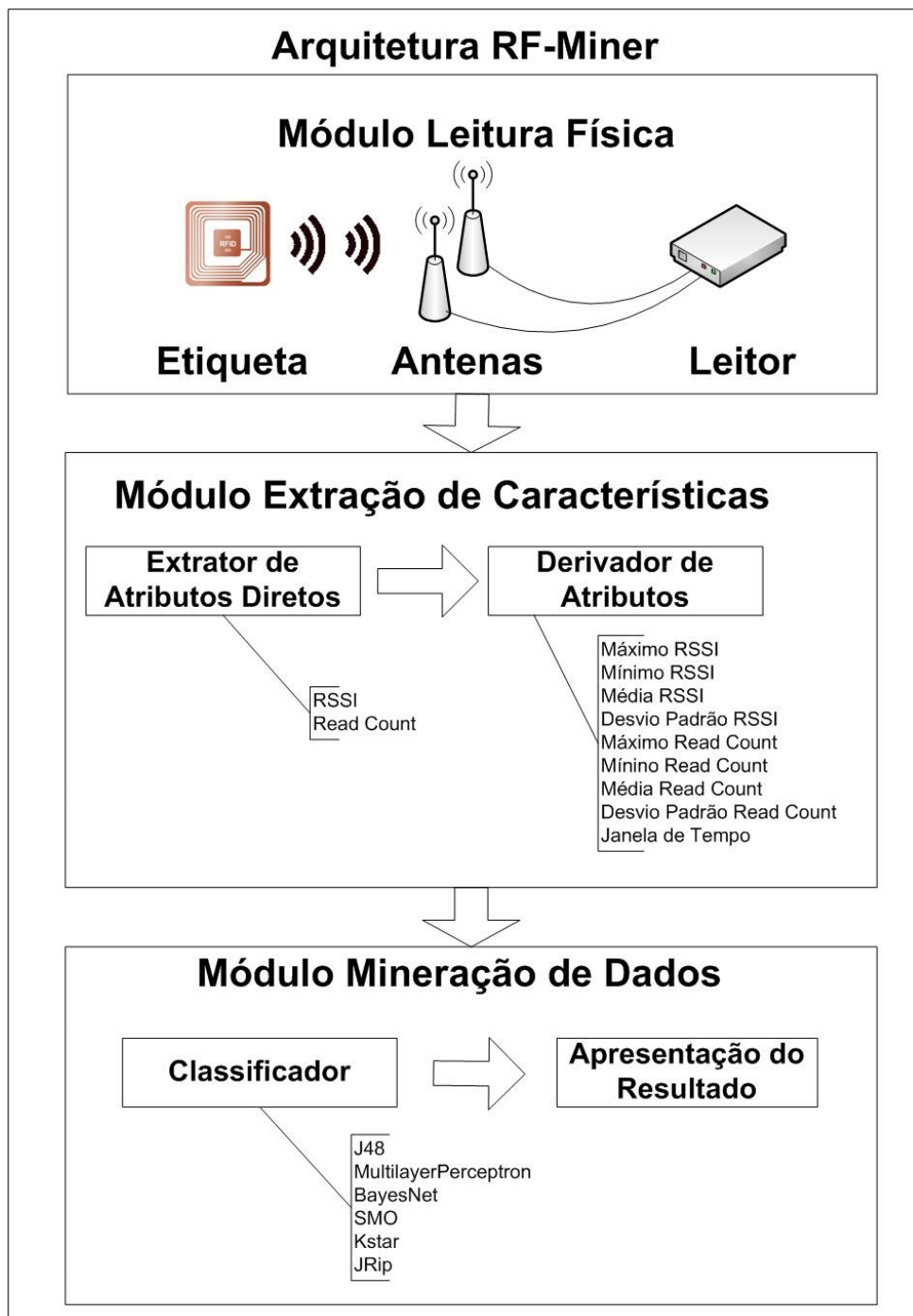


Figura 9: Arquitetura RF-Miner

fixação de antenas podem interferir na apuração dos resultados. Outro aspecto importante na implementação são as definições dos parâmetros de leitura e distância das antenas, que devem ser definidos através da realização de diversos testes até identificar os valores ideais sempre levando em consideração as dimensões do ambiente instalado. A seção 5.1.1 apresenta os equipamentos utilizados na implementação, os parâmetros de leitura, distâncias definidas e como os parâmetros foram identificados.

5.1.1 PREPARAÇÃO DO AMBIENTE DE IMPLEMENTAÇÃO

Para a preparação do ambiente foram utilizadas quatro antenas UHF mono estáticas que atuam na faixa de frequência 902 a 928 MHz com um ganho de 6 dBi. O equipamento utilizado para realização das leituras foi o *ThingMagic Mercury 6*, que é um leitor RFID UHF de alto desempenho. Suporta até 4 antenas mono estáticas, entradas e saídas digitais e conexão Wi-Fi. Ambos equipamentos comerciais e que podem ser adquiridos por qualquer pessoa em empresas especializadas na comercialização de equipamentos RFID.

As etiquetas passivas foram dispostas lado a lado em quatro prateleiras (P1, P2, P3, P4) de livros. Foram utilizadas vinte e oito etiquetas coladas no sentido vertical na lombada dos livros, a Figura 10 demonstra o ambiente de implementação montado.

Equipamentos de outros fabricantes também poderão ser utilizados na arquitetura RF-Miner, pois, por padrão todos apresentam os atributos *RSSI* e *Read Count* utilizados no Módulo Extração de Características e Mineração de Dados.

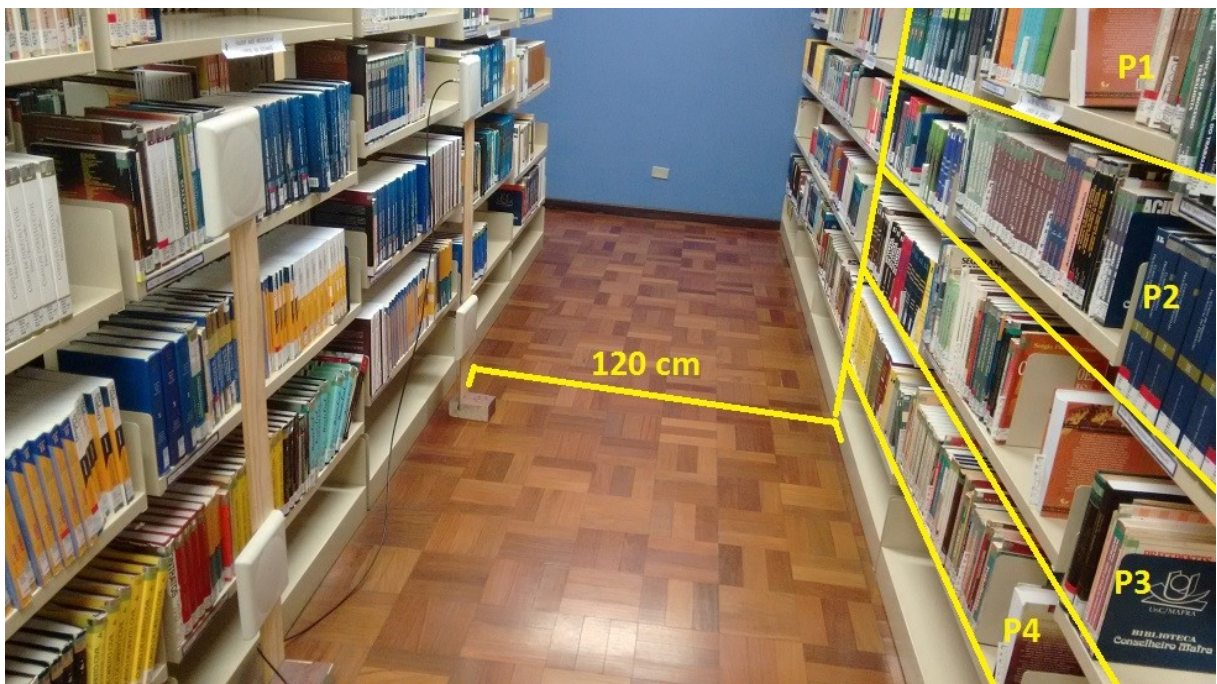


Figura 10: Ambiente de Implementação

5.1.1.1 DEFINIÇÃO DOS PARÂMETROS DE LEITURA E DISTÂNCIAS DAS ANTENAS

Na implantação de um sistema RFID o processo de definição dos parâmetros de leitura e distâncias das antenas é conhecido como *Site Survey* apresentado na seção 2.4. Para a implementação da arquitetura RF-Miner é necessário identificar os seguintes parâmetros:

1. **Distância das antenas:** Distância das antenas em relação as etiquetas, esta distância normalmente é delimitada pelo espaço existente no ambiente instalado, no caso da implementação apresentada neste trabalho, foi definida com o valor da largura do corredor **120 cm** demonstrada nas Figuras 10 e 11;
2. **Distância Entre os Grupos de Antenas:** A definição da distância entre os grupos de antenas foi identificada através da realização de diversas leituras até o limite de distância máximo onde todas as etiquetas são identificadas em pelo menos uma das quatro antenas. O valor identificado na implementação foi **210 cm** conforme demonstra a Figura 11;
3. **Intervalo de Leitura:** Este é o tempo parametrizável no leitor onde o mesmo aciona as antenas e recebe os dados das etiquetas encontradas, padroniza informações de entrada em um fluxo de saída composto pela tupla *timestamp* (Hora de leitura), *RSSI*, *TagID* (identificador único da etiqueta), *Read Count* (número de ocasiões que o leitor encontrou a etiqueta no intervalo de leitura) e *Antenna* (identificador da antena onde foi realizada a leitura), conforme demonstra a Figura 12 . Para cada Intervalo de Leitura o leitor apresenta uma tupla com os valores encontrados para cada uma das etiquetas e em cada antena. Para identificar o valor do parâmetro Intervalo de Leitura ideal, inicialmente foi utilizado o valor 1 segundo e foi sendo aumentado gradativamente até o limite mínimo em que todas as etiquetas foram identificadas pelo menos uma vez no mínimo em uma antena. O valor encontrado para a implementação foi **5 segundos**;
4. **Janela de Intervalos de Leitura:** Este parâmetro define a quantidade de Intervalos de Leitura (Parâmetro 3) que serão utilizados pelo Módulo Extração de Características (seção 5.2) para realizar a derivação dos atributos *RSSI* e *Read Count*, ou seja, a quantidade de tuplas (Figura 12) por etiqueta que serão utilizadas para calcular a média, desvio padrão e identificar os valores mínimo e o máximo. Para identificar o valor ideal deste parâmetro é necessário ter toda a arquitetura montada, realizar uma bateria de leituras começando pelo valor mínimo necessário para cálculo das médias e desvio padrão que é 2 e ir aumentando até que a arquitetura RF-Miner consiga obter 100% de acerto em pelo menos um dos classificadores utilizados. Quanto menor o valor deste parâmetro, menor será o tempo de resposta necessário após a movimentação de algum objeto, porém se o valor for insuficiente, a arquitetura poderá não identificar corretamente o posicionamento dos objetos. Na implementação apresentada neste trabalho o valor encontrado para este parâmetro após testes iniciando em 2 janelas e sendo aumentado gradativamente, foi o valor de **10 intervalos**, portanto, para obter a medição de um resultado a arquitetura implementada neste ambiente de testes necessita de 50 segundos no mínimo (10 Intervalos

de Leitura de 5 segundos), além do tempo de processamento onde são derivados os atributos e apresentadas as localizações encontradas pelos algoritmos. A Figura 13 apresenta o detalhadamente o que é o parâmetro Janela de Intervalos de Leitura;

A Figura 14 apresenta o diagrama de modelagem de processos para definição dos parâmetros necessários na arquitetura RF-Miner.

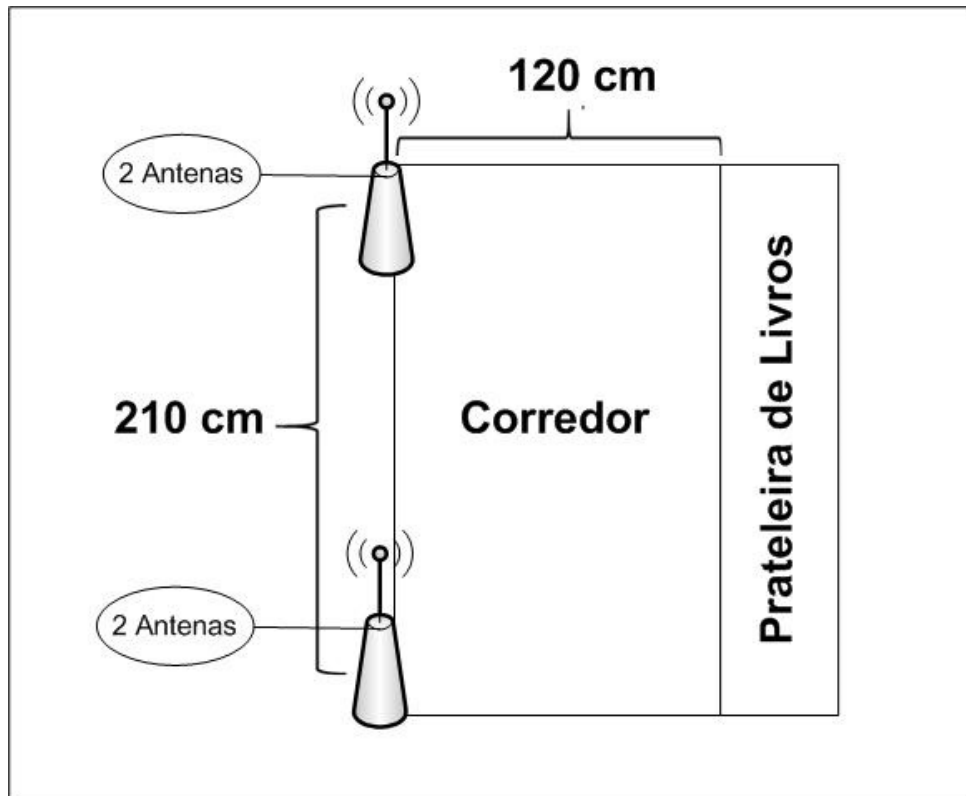


Figura 11: Distâncias das Antenas

300833B2DDD906C000000001	13:56:27:091	-59	22	1
300833B2DDD906C000000002	13:56:27:204	-65	22	1
300833B2DDD906C000000004	13:56:27:314	-69	21	1
300833B2DDD906C000000005	13:56:27:165	-65	23	1
TagID	TimeStamp (msec)	RSSI (dBm)	ReadCount	Antenna

Figura 12: Tupla Enviada Pelo Leitor

5.1.2 REALIZAÇÃO DAS LEITURAS E BASE INICIAL DE TREINAMENTO

Para que os classificadores possam identificar em tempo real a localização das etiquetas, é necessário formar uma base de treinamento, onde as etiquetas são posicionadas em

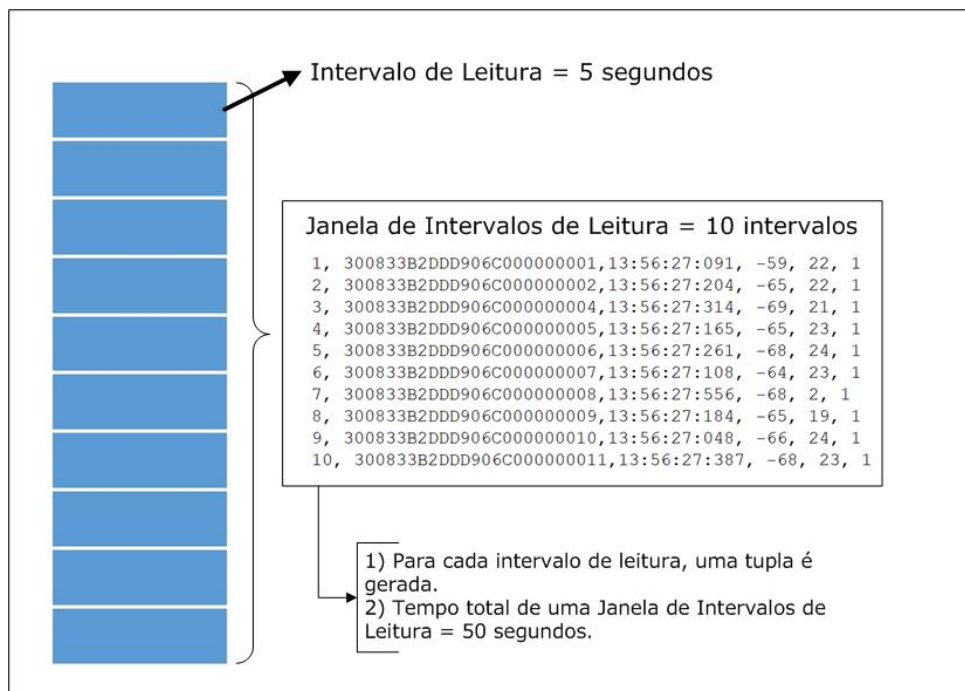


Figura 13: Parâmetro Janela de Intervalos de Tempo

localizações conhecidas e realizadas diversas leituras. Os valores dos atributos encontrados nestas localizações devem ser utilizados na formação da base inicial de treinamento. Foram realizadas em cada etiqueta alvo 100 leituras com o parâmetro Intervalo de Leitura igual a 5 segundos (definido na seção 5.1.1.1). Em cada etiqueta foi medido o RSSI em dBm e a *Read Count* para cada uma das 4 antenas, totalizando 2.800 leituras. Para esta etapa o módulo de Leitura Física faz a conexão por TCP/IP com o Leitor *ThingMagic 6*, envia os comandos para iniciar as leituras e recebe as respostas das etiquetas localizadas com seus respectivos atributos diretos, conforme padrão apresentado na Figura 12. As informações são repassadas para o Módulo de Extração de Características que é responsável por organizar o fluxo de dados e derivar os atributos.

5.2 MÓDULO EXTRAÇÃO DE CARACTERÍSTICAS

É a parte lógica da arquitetura, que faz parte dos componentes de negócio do sistema RFID (Capítulo 2). Conforme demonstrado na Figura 9, este módulo está subdividido em dois processos: **Extrator de Atributos Diretos** e **Derivador de Atributos**. As seções 5.2.1 e 5.2.2 apresentam detalhadamente como cada processo se comporta.

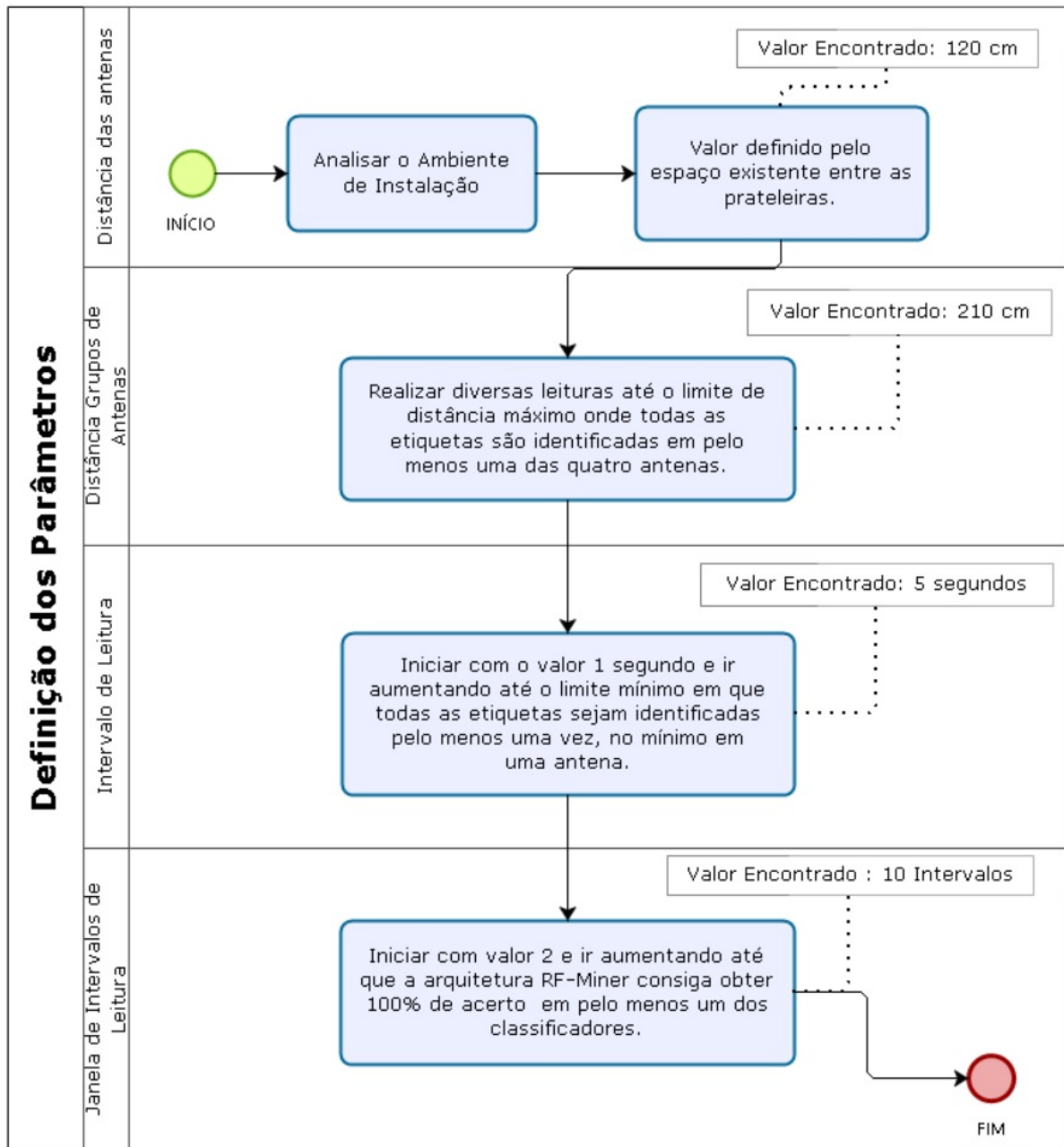


Figura 14: Modelagem de Processos: Definição dos Parâmetros

5.2.1 EXTRATOR DE ATRIBUTOS DIRETOS

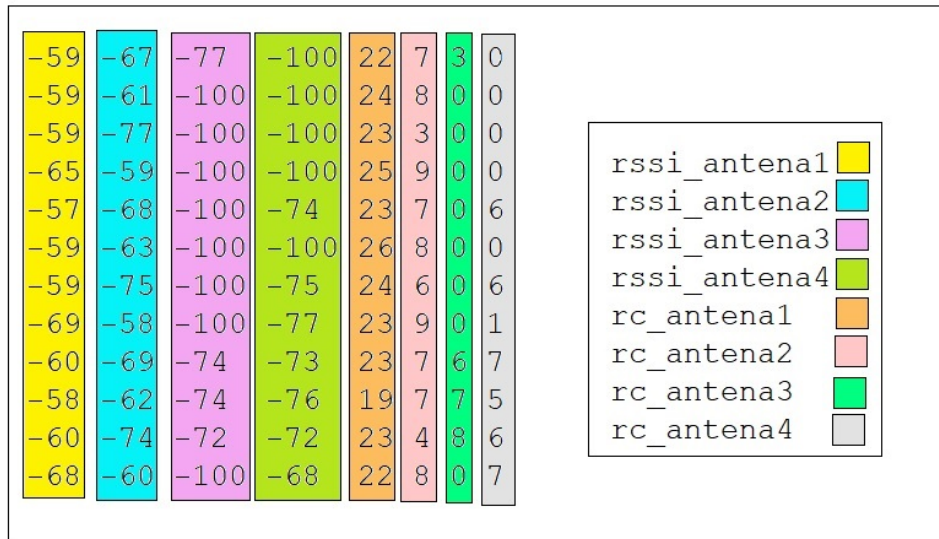
É responsável por receber o fluxo de dados do Módulo Leitura Física e organizar os valores dos atributos diretos (RSSI e *Read Count*) de cada uma das antenas. A Tabela 3 demonstra como os atributos são organizados pelo processo para cada uma das antenas X , onde X é o número da antena. Após esta etapa o total de atributos da base é oito, dois atributos (RSSI e *Read Count*) para cada antena.

A Figura 15 apresenta uma ilustração de como os dados são organizados após o pro-

Tabela 3: Atributos Diretos

Atributo	Tipo de Dado	Descrição
rss_i_antena X	Decimal	RSSI da antena
rc_antena X	Inteiro	Número de leituras da etiqueta na antena

cesso Extrator de Atributos Diretos.

**Figura 15: Dados organizados após processo: Extrator de Atributos Diretos**

5.2.2 DERIVADOR DE ATRIBUTOS

Segundo (SPIEGEL; STEPHENS, 2015), entre as técnicas de estatística descritiva existentes, quando há um conjunto de dados, a primeira a ser utilizada é a de resumo de dados. Para dados quantitativos que é o caso do presente trabalho, existem as medidas de posição (Média, Moda, Mediana, Percentis e Quartis) e as medidas de dispersão (Desvio Padrão, Amplitude, Intervalo-Interquartil, Variância e Coeficiente de Variação). Na arquitetura proposta no presente trabalho foi escolhida uma técnica de medidas de posição (Média) e uma técnica de medidas de dispersão (Desvio Padrão) para realização da derivação dos atributos diretos.

Baseado no valor do parâmetro Janela de Intervalos de Leitura, o processo Derivador de Atributos cria a identificação do número da janela para as tuplas recebidas pelo processo Extrator de Atributos Diretos, identifica os valores mínimos, máximos e calcula as médias aritméticas e desvio padrão de cada atributo direto (RSSI e *Read Count*) para cada uma das antenas X, onde X é o número da antena. A Figura 16 apresenta o fluxograma do algoritmo utilizado para derivar os atributos e a Tabela 4 demonstra a estrutura da base de dados após a finalização deste processo. O total de atributos da base aumenta para quarenta, dez atributos para cada

antena.

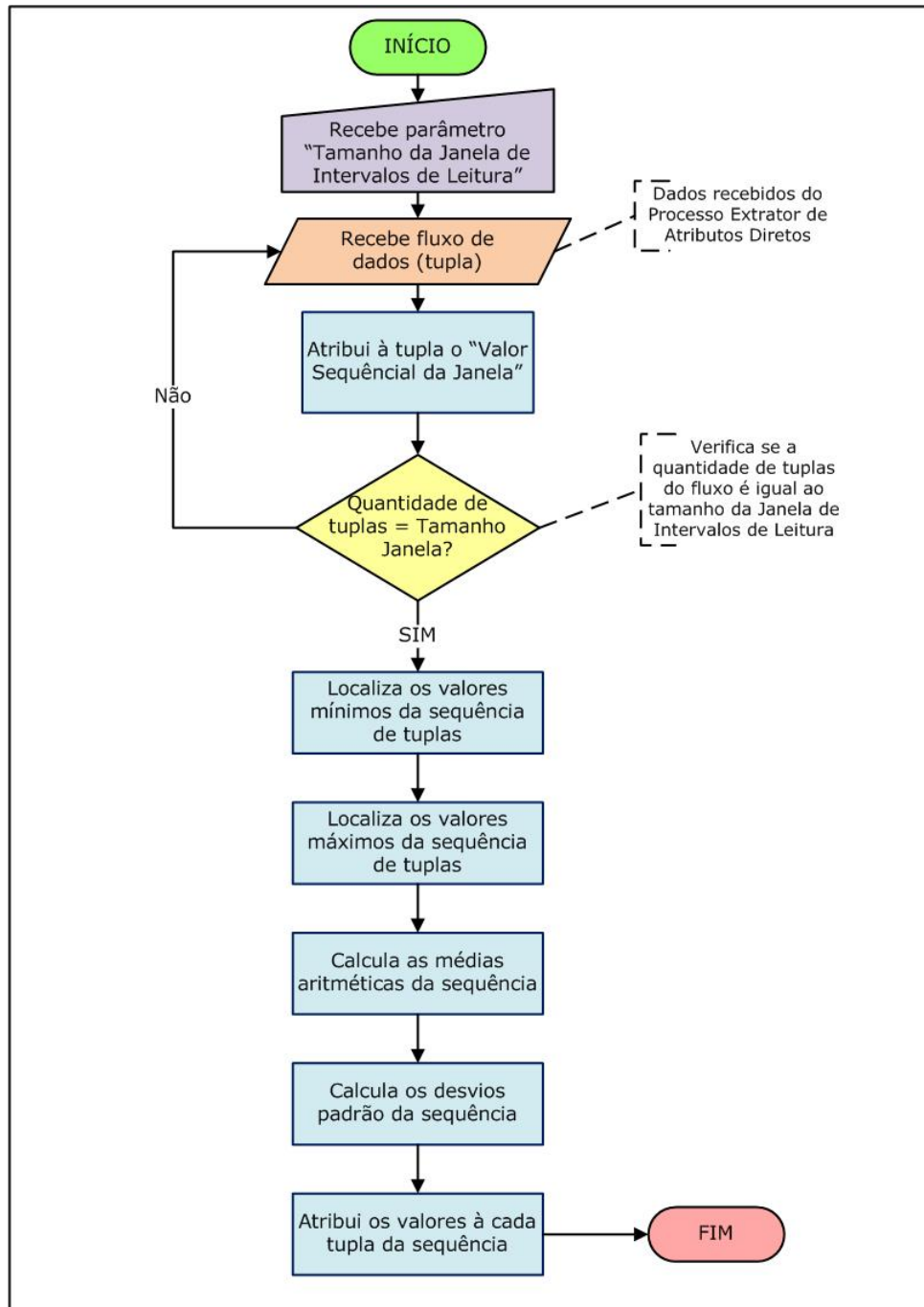


Figura 16: Algoritmo: Derivador de Atributos

Existem dois atributos que não estão relacionados diretamente à antena. O número da Janela de Intervalos de Leitura que é um número sequencial da janela definido no processo Derivador de Atributos e a posição da etiqueta na prateleira. Quando está sendo criada a base de treinamento a posição deve ser conhecida, quando está sendo realizado o teste em tempo real este valor é encontrado pelo Módulo Mineração de Dados. Os atributos são apresentados na Tabela 5.

Tabela 4: Atributos Diretos e Derivados

Atributo	Tipo de Dado	Descrição
rss_i_antena X	Decimal	RSSI da antena
rc_antena X	Inteiro	Número de leituras da etiqueta na antena
avg_rssi_antena X	Decimal	Média aritmética do RSSI na janela
avg_rc_antena X	Decimal	Média aritmética do RC na janela
min_rssi_antena X	Decimal	Menor valor do RSSI na janela
min_rc_antena X	Decimal	Menor valor do RC na janela
max_rssi_antena X	Decimal	Maior valor do RSSI na janela
max_rc_antena X	Decimal	Maior valor do RC na janela
stddev_rssi_antena X	Decimal	Desvio padrão do RSSI na janela
stddev_rc_antena X	Decimal	Desvio padrão do RC na janela

Tabela 5: Atributos não relacionados diretamente à antena

Atributo	Tipo de Dado	Descrição
Janela de Intervalos de Leitura	Inteiro	Número Janela de Intervalos de Leitura
Posição	Classe	Nome ou número da posição na prateleira

5.2.3 FORMATAÇÃO DA BASE DE TREINAMENTO

A base de treinamento é o conjunto de dados usados nos algoritmos de mineração de dados para construir o modelo classificador. O módulo Mineração de Dados (seção 5.3) utiliza como ferramenta de classificação o software WEKA (*Waikato Environment for Knowledge Analysis*), que é um software livre para mineração de dados, do tipo open source, desenvolvido em linguagem Java dentro das especificações da GPL (*General Public License*) que se consolidou como uma das ferramentas de mineração de dados mais utilizadas por estudantes e professores em pesquisas científicas, devido a grande variedade de algoritmos e técnicas implementados no mesmo (WITTEN et al., 2016a). O software WEKA recebe a base de treinamento via banco de dados ou via arquivo no formato ARFF (*Attribute-Relation File Format*). Este formato foi utilizado para criação da base de dados da arquitetura RF-Miner, portanto, para repassar os atributos da base de treinamento para o Módulo Mineração de Dados, o Módulo Extração de Características organiza os dados no formato ARFF. A Figura 17 demonstra um trecho do arquivo com os dados organizados neste formato.

5.3 MÓDULO MINERAÇÃO DE DADOS

Usando como entrada os dados preparados e formatados pelo módulo Extração de Características, o módulo Mineração de Dados utiliza o software WEKA para realizar a classificação através das técnicas de mineração de dados e apresenta a localização encontrada por cada

```

1 @relation leituras_rfid
2 @attribute rssi_antena1 NUMERIC
3 @attribute rc_antena1 NUMERIC
4 @attribute avg_rssi_antena1 NUMERIC
5 @attribute min_rssi_antena1 NUMERIC
6 @attribute max_rssi_antena1 NUMERIC
7 @attribute stddev_rssi_antena1 NUMERIC
8 @attribute avg_rc_antena1 NUMERIC
9 @attribute min_rc_antena1 NUMERIC
10 @attribute max_rc_antena1 NUMERIC
11 @attribute stddev_rc_antena1 NUMERIC
12 @attribute janela {1,2,3,4,5,6,7,8,9,10}
13 @attribute posicao {01,02,03,04,05,06,07,08,09,
14 10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28}
15 @data
16 -59,-67,-77,-100,22,7,3,0,-62.200000,-65.700000,-81.20000
17 -59,-61,-100,-100,24,8,0,0,-62.200000,-61.400000,-100.000
18 -59,-77,-100,-100,23,3,0,0,-61.500000,-78.700000,-100.000
19 -65,-59,-100,-100,25,9,0,0,-66.200000,-58.000000,-100.000
20 -57,-68,-100,-74,23,7,0,6,-58.900000,-67.100000,-100.0000
21 -59,-63,-100,-100,26,8,0,0,-61.200000,-62.200000,-77.5000
22 -59,-75,-100,-75,24,6,0,6,-60.600000,-74.700000,-100.0000
23 -69,-58,-100,-77,23,9,0,1,-68.900000,-57.200000,-100.0000
24 -60,-69,-74,-73,23,7,6,7,-62.600000,-68.100000,-73.200000
25 -58,-62,-74,-76,19,7,7,5,-60.600000,-61.800000,-71.600000

```

Figura 17: Arquivo ARFF

um dos algoritmos.

Das técnicas de mineração de dados existentes, a que se aplica dentro das características da arquitetura RF-Miner é a classificação preditiva (seção 3.1). Pois esta técnica tem o propósito de prever o valor de um determinado atributo baseado nos valores de outros atributos. No caso da arquitetura RF-Miner, irá prever a localização da etiqueta baseada nos valores de todos os outros atributos demonstrados na Tabela 4.

Os seguintes tipos de classificadores podem ser utilizados pela técnica de classificação preditiva.

- Classificadores Baseados em Árvore de Decisão;
- Classificadores Baseados em Redes Neurais Artificiais;
- Classificadores Bayesianos;
- Classificadores Baseados em Vizinheiro Mais Próximo;
- Classificadores Baseado em Regras;
- Classificadores Baseado em Máquina de Vetor de Suporte.

Com a finalidade de testar cada tipo de classificador na arquitetura RF-Miner, foi escolhido dentre os algoritmos implementados no software *WEKA*, um de cada tipo. A tabela 6 demonstra os algoritmos escolhidos. O módulo também pode trabalhar com outros algoritmos de classificação, provando assim a generalização da arquitetura.

Tabela 6: Algoritmos de cada tipo de classificação

Tipo de Classificação	Algoritmo
Árvore de Decisão	J48
Rede Neural Artificial	MultilayerPerceptron
Bayesiano	BayesNet
Máquina de Vetores de Suporte	SMO
Vizinho Mais Próximo	Kstar
Baseado em Regras	Jrip

O módulo Mineração de Dados conforme apresentado na Figura 9 é subdividido em dois processos, o processo Classificador que utiliza os algoritmos classificadores para identificar a localização das etiquetas e a Apresentação de Resultados. O processo Classificador pode trabalhar de duas maneiras, avaliando a base de treinamento ou apurando a classificação em tempo real. As seções 5.3.1 e 5.3.2 apresentam detalhadamente como se comportam.

5.3.1 AVALIANDO A BASE DE TREINAMENTO

O processo de avaliação da base de treinamento é importante para verificar se o parâmetro: Janela de Intervalos de Leitura (seção 5.1.1.1) foi definido com um valor suficiente para que o classificador possa obter 100% de acerto no ambiente de implementação. Este processo avalia a base de treinamento usando a técnica de validação cruzada apresentado na seção 3.2. Na implementação apresentada neste trabalho o parâmetro dobra (*fold*) necessário para a técnica validação cruzada foi definido com o valor 10 (valor *default* da técnica). Neste caso o software divide a base em 10 partes iguais. Em cada iteração utiliza 9 partes para treinamento e 1 para testes. Este processo é repetido para cada uma das 10 partes sempre trocando a parte que será utilizada para testes. A Figura 18 demonstra o comportamento desta técnica.

A Figura 19 demonstra o fluxograma do algoritmo utilizado para realizar o processo de avaliação da base de treinamento e a Figura 20 apresenta o código fonte em Java da implementação do algoritmo utilizando o classificador J48.

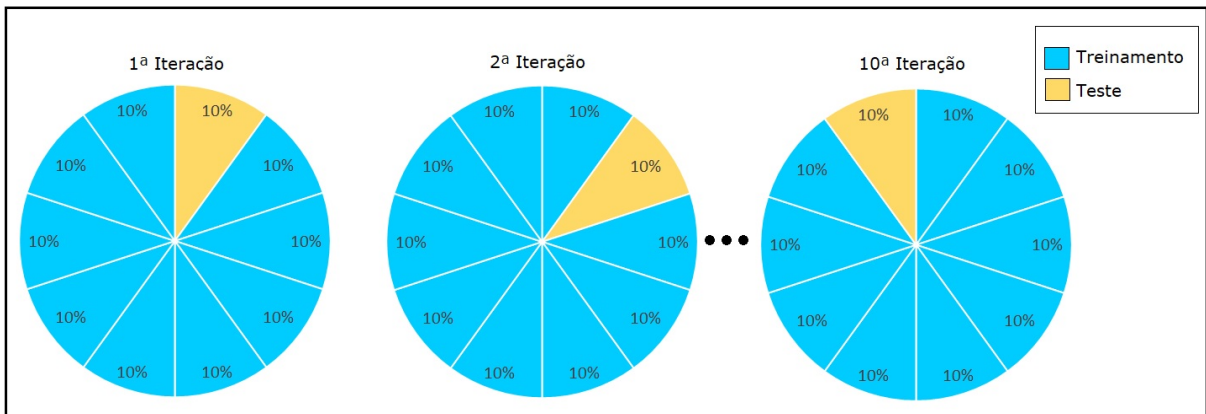


Figura 18: Validação Cruzada

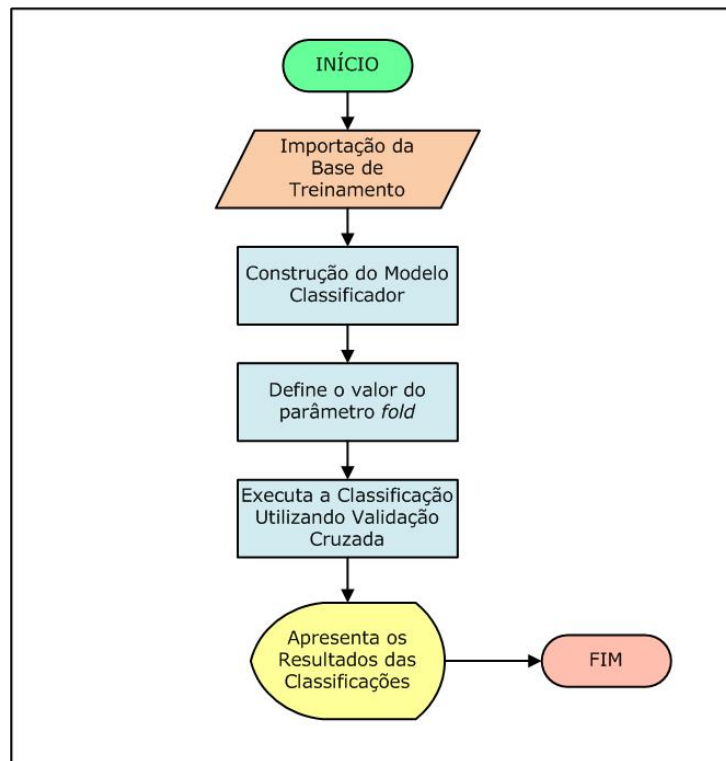


Figura 19: Algoritmo: Avaliar Base de Treinamento

5.3.2 CLASSIFICAÇÃO EM TEMPO REAL

O processo de classificação em tempo real utiliza o modelo de classificação simples, onde a base de treinamento é utilizada para a construção do modelo classificador, utilizado pelos algoritmos de classificação e o fluxo de dados que é recebido do Módulo Extração de Características é utilizado como instância de testes. A qual é classificada na posição que o algoritmo escolhido encontra. O tempo médio total necessário para a arquitetura RF-Miner apresentar o resultado na implementação demonstrada neste trabalho foi de 51,5 segundos, onde 50 segundos é o tempo da Janela de Intervalos de Leitura e 1,5 segundos é o tempo médio de

```

1  import weka.core.converters.ConverterUtils.DataSource;
2  import weka.core.Instances;
3  import weka.core.Instance;
4  import weka.classifiers.trees.J48;
5  public class ClassificaJ48 {
6      public static void main(String[] args) throws Exception {
7
8          // (1) importação da base de dados de treinamento
9          DataSource source = new DataSource("leituras_rfid.arff");
10         Instances D = source.getDataSet();
11
12         // (1.1) - especificação do atributo classe
13         if (D.classIndex() == -1)
14             D.setClassIndex(41); //Atributo 41 = Posição da Etiqueta
15
16         // (2) Construção do modelo classificador (treinamento)
17         J48 classificador = new J48(3);
18         J48.buildClassifier(D);
19
20         // (3) Define o Parâmetro Fold
21         int numFolds = 10;
22
23         // (4) Executa a Classificação Utilizando Validação Cruzada
24         MultipleEvaluation resultado;
25         resultado = J48.crossValidate(D,numFolds);
26
27         // (5) Apresenta os resultados
28         System.out.println(resultado);
29     }
30 }

```

Figura 20: Código Fonte: Avaliar Base de Treinamento

processamento do resultado. A Figura 21 demonstra o fluxograma do algoritmo utilizado para realizar o processo de classificação em tempo real e a Figura 22 apresenta o código fonte em Java da implementação do algoritmo utilizando o classificador J48.

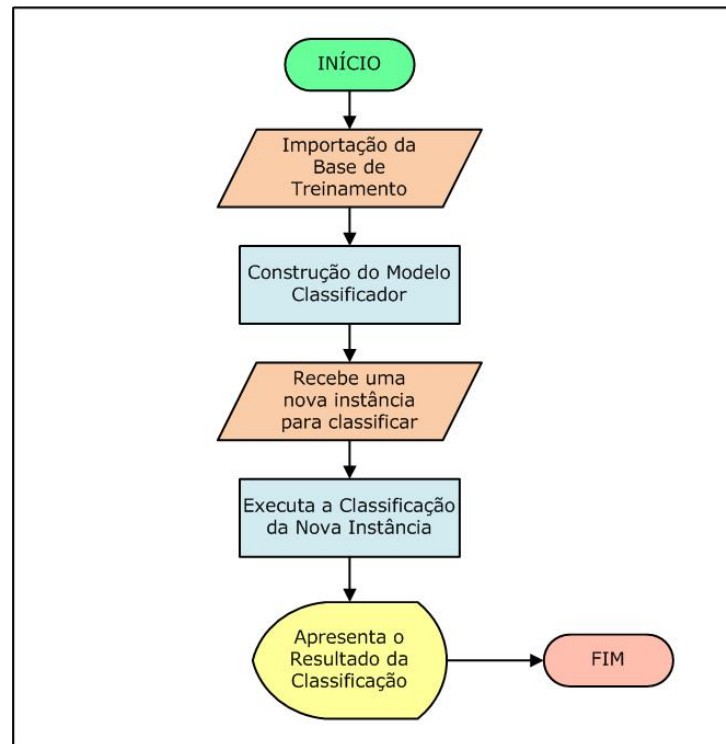


Figura 21: Algoritmo: Classificação em Tempo Real

```

1  import weka.core.converters.ConverterUtils.DataSource;
2  import weka.core.Instances;
3  import weka.core.Instance;
4  import weka.classifiers.trees.J48;
5  public class ClassificaJ48 {
6      public static void main(String[] args) throws Exception {
7          // (1) importação da base de dados de treinamento
8          DataSource source = new DataSource("leituras_rfid.arff");
9          Instances D = source.getDataSet();
10
11         // (1.1) - especificação do atributo classe
12         if (D.classIndex() == -1)
13             D.setClassIndex(41); //Atributo 41 = Posição da Etiqueta
14
15         // (2) Construção do modelo classificador (treinamento)
16         J48 classificador = new J48(3);
17         J48.buildClassifier(D);
18
19         // (3) Recebe uma nova instância para classificar
20         Instance newInst = new Instance();
21         newInst.setDataset(D);
22         String[] tupla = args; //Recebe a nova instância como argumento
23         for (i = 0; i < tupla.length; i++) //Instancia o valor dos atributos
24             newInst.setValue(i, tupla[i]);
25
26         // (4) Executa a Classificação da Nova Instância
27         String pred = J48.classifyInstance(newInst);
28
29         // (5) Apresenta o Resultado da Classificação
30         System.out.println("Predição: " + pred);
31     }
32 }
  
```

Figura 22: Código Fonte: Classificação em Tempo Real

6 RESULTADOS E DISCUSSÕES

A estratégia de análise do desempenho da arquitetura RF-Miner, se concentrou em três etapas. Inicialmente foram realizados testes com o objetivo de identificar o quanto o Módulo de Extração de Características influencia na apuração dos resultados, na segunda etapa foram realizados testes para avaliar o desempenho da arquitetura em tempo real fazendo movimentações nos livros e na terceira etapa foram analisados os testes com a redução da quantidade de antenas. As seções 6.1, 6.2 e 6.3 apresentam detalhadamente os respectivos resultados.

6.1 ATRIBUTOS DIRETOS EM RELAÇÃO AOS ATRIBUTOS DIRETOS EM CONJUNTO AOS DERIVADOS

A primeira investigação consistiu em identificar o comportamento da arquitetura utilizando apenas os atributos diretos (RSSI e *Read Count*) na apuração da localização. Para esta apuração foi criada uma base de treinamento realizando 100 leituras em todas 28 etiquetas totalizando 2.800. No Módulo de Mineração de dados, foi utilizado o algoritmo de avaliação da base de treinamento (seção 5.3.1), que utiliza a técnica de validação cruzada para realizar a classificação. Os resultados são apresentados na Tabela 7 e na Figura 23. É possível perceber que nenhum classificador conseguiu obter 100% de acerto. O algoritmo que melhor resultado apresentou foi o classificador Bayesiano *BayesNet* com 98,89% de acerto, seguido pelo classificador baseado em Vizinho Mais Próximo *KStar* com 98,86% de acerto. O pior resultado foi apresentado pelo classificador baseado em Máquina de Vetores de Suporte *SMO*. A Tabela 7 também apresenta as margens de erro (M.E) calculadas com nível de confiança de 99,00%.

Tabela 7: Classificação utilizando os atributos diretos (RSSI e *Read Count*)

Algoritmo	Corretas	Erradas	Total	% Acerto	Nível Confiança	M.E
J48	2.761	39	2.800	98,61%	99,00%	0,56
MultilayerPerceptron	2.759	41	2.800	98,54%	99,00%	0,58
BayesNet	2.769	31	2.800	98,89%	99,00%	0,50
SMO	2.635	165	2.800	94,11%	99,00%	1,15
Kstar	2.768	32	2.800	98,86%	99,00%	0,51
Jrip	2.734	66	2.800	97,64%	99,00%	0,73

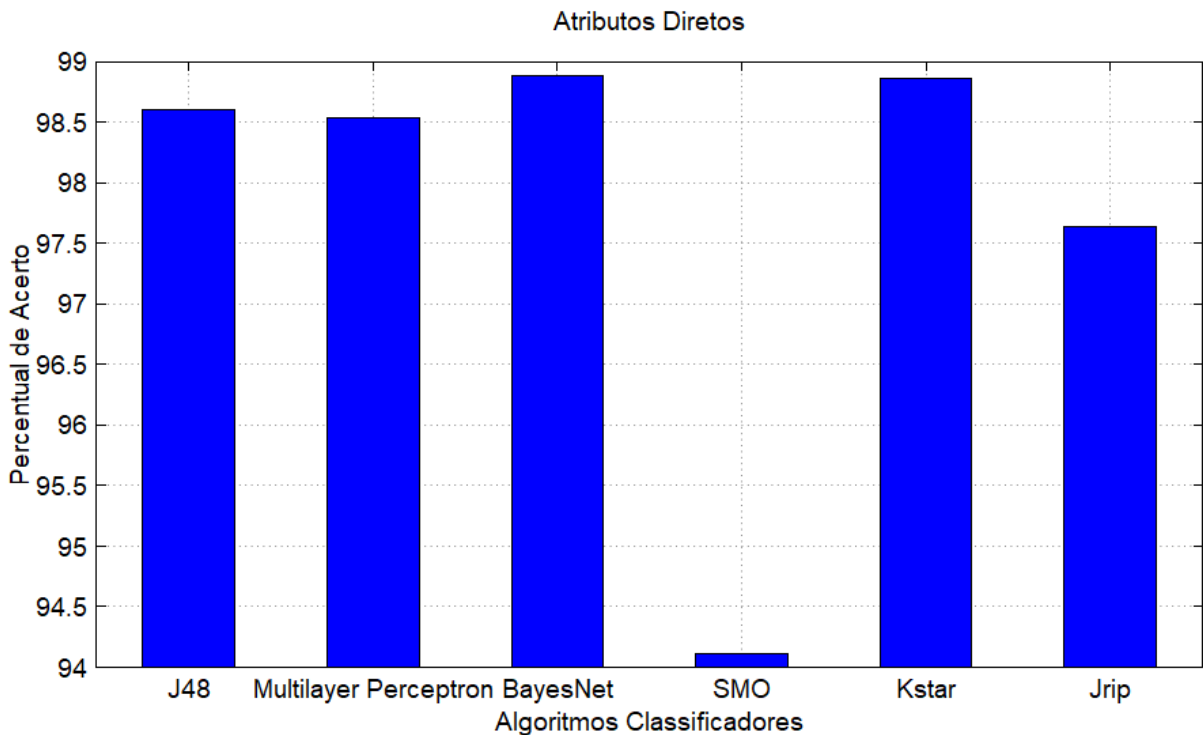


Figura 23: Resultados utilizando atributos diretos

Na segunda classificação apresentada foram considerados os atributos diretos RSSI e *Read Count* em conjunto com os atributos derivados valor máximo RSSI e *Read Count*, mínimo do RSSI e *Read Count*, média aritmética do RSSI e *Read Count* e desvio padrão do RSSI e *Read Count* para cada uma das quatro antenas, totalizando quarenta atributos. Para esta apuração foi utilizada a mesma base de dados anterior com a mesma técnica de avaliação da base de treinamento (seção 5.3.1) e com os parâmetros definidos na seção 5.1.1.1. A Tabela 8 apresenta os resultados da segunda classificação. Nesta investigação três algoritmos alcançaram 100% de acerto, o classificador baseado em árvore de decisão *J48*, o classificador Bayesiano *BayesNet* e o classificador baseado em vizinho mais próximo *KStar*. A Tabela 8 também apresenta as margens de erro (M.E) calculadas com nível de confiança de 99,00%.

Tabela 8: Resultados utilizando atributos diretos e derivados

Algoritmo	Corretas	Erradas	Total	% Acerto	Nível Confiança	M.E
J48	2.800	0	2.800	100,00%	99,00%	0,00
MultilayerPerceptron	2.797	3	2.800	99,89%	99,00%	0,16
BayesNet	2.800	0	2.800	100,00%	99,00%	0,00
SMO	2.774	26	2.800	99,07%	99,00%	0,47
Kstar	2.800	0	2.800	100,00%	99,00%	0,00
Jrip	2.784	16	2.800	99,43%	99,00%	0,37

A Tabela 9 e a Figura 24 apresentam os comparativos entre a primeira classificação

utilizando apenas os atributos diretos e a segunda classificação utilizando os atributos diretos em conjunto com os derivados. É possível identificar um aumento significativo na taxa de acertos em todos os classificadores, principalmente no classificador baseado em Máquina de Vetores de Suporte *SMO* que havia obtido 165 erros na primeira classificação e reduziu para 26 erros na segunda. Estes resultados demonstram que a técnica de derivação dos atributos utilizando as Janelas de Intervalos de Leitura proposta neste trabalho contribui expressivamente na melhoria do desempenho de todos os métodos de classificação.

Tabela 9: Comparativo Atributos diretos x Atributos diretos e derivados

Algoritmo	Diretos	Derivados
J48	98,61%	100,00%
MultilayerPerceptron	98,54%	99,89%
BayesNet	98,89%	100,00%
SMO	94,11%	99,07%
Kstar	98,86%	100,00%
Jrip	97,64%	99,43%

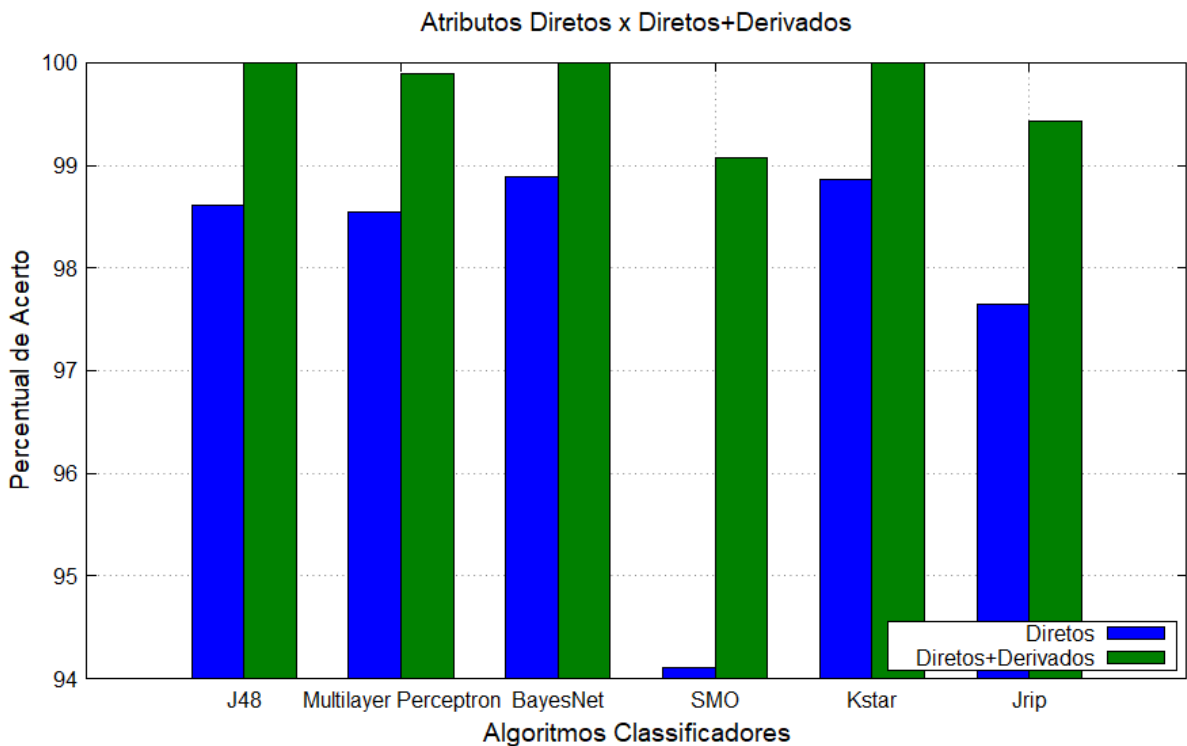


Figura 24: Comparativo Atributos diretos x Atributos diretos e derivados

6.2 RESULTADOS EM TEMPO REAL

Com o objetivo de identificar o desempenho da arquitetura proposta na apuração da localização em tempo real, foram trocados de posição todos os 28 livros e realizadas as leitu-

ras e apuração de resultados a cada Janela de Intervalos de Leitura, desta maneira foi possível identificar a quantidade de Janelas de Intervalos de Leitura necessárias para a arquitetura obter resultados semelhantes a base de treinamento (Tabela 8) após a movimentação de todas as etiquetas alvo. Neste caso a quantidade de registros testados é 280, devido a existir 28 etiquetas e 10 leituras (Tamanho da Janela de Intervalos de Leitura, parâmetro definido na seção 5.1.1.1). A Tabela 10 apresenta os resultados obtidos uma Janela de Intervalos de Leitura após a troca de todos os livros. É possível perceber que há uma redução na taxa de acerto em alguns algoritmos, esta redução é esperada devido a cada livro possuir espessura e tamanhos diferentes o que afeta sensivelmente a posição original de cada etiqueta. Mesmo assim dois algoritmos obtiveram 100% de acerto: *BayesNet* e *Kstar*.

Tabela 10: Resultados em Tempo Real: Após 1ª Janela de Intervalos de Leitura

Algoritmo	Corretas	Erradas	Total	Taxa Acerto
J48	250	30	280	89,29%
MultilayerPerceptron	276	4	280	98,57%
BayesNet	280	0	280	100,00%
SMO	233	47	280	83,21%
Kstar	280	0	280	100,00%
Jrip	238	42	280	85,00%

A Tabela 11 apresenta os resultados obtidos duas Janelas de Intervalos de Leitura após a troca de todos os livros. Existe um aumento na taxa de acerto, pois as leituras da primeira 1ª Janela de Intervalos de Leitura foram adicionadas na base de treinamento e contribuíram na melhoria do desempenho dos classificadores.

Tabela 11: Resultados em Tempo Real: Após 2ª Janela de Intervalos de Leitura

Algoritmo	Corretas	Erradas	Total	Taxa Acerto
J48	270	10	280	96,43%
MultilayerPerceptron	278	2	280	99,29%
BayesNet	280	0	280	100,00%
SMO	264	16	280	94,29%
Kstar	280	0	280	100,00%
Jrip	257	23	280	91,79%

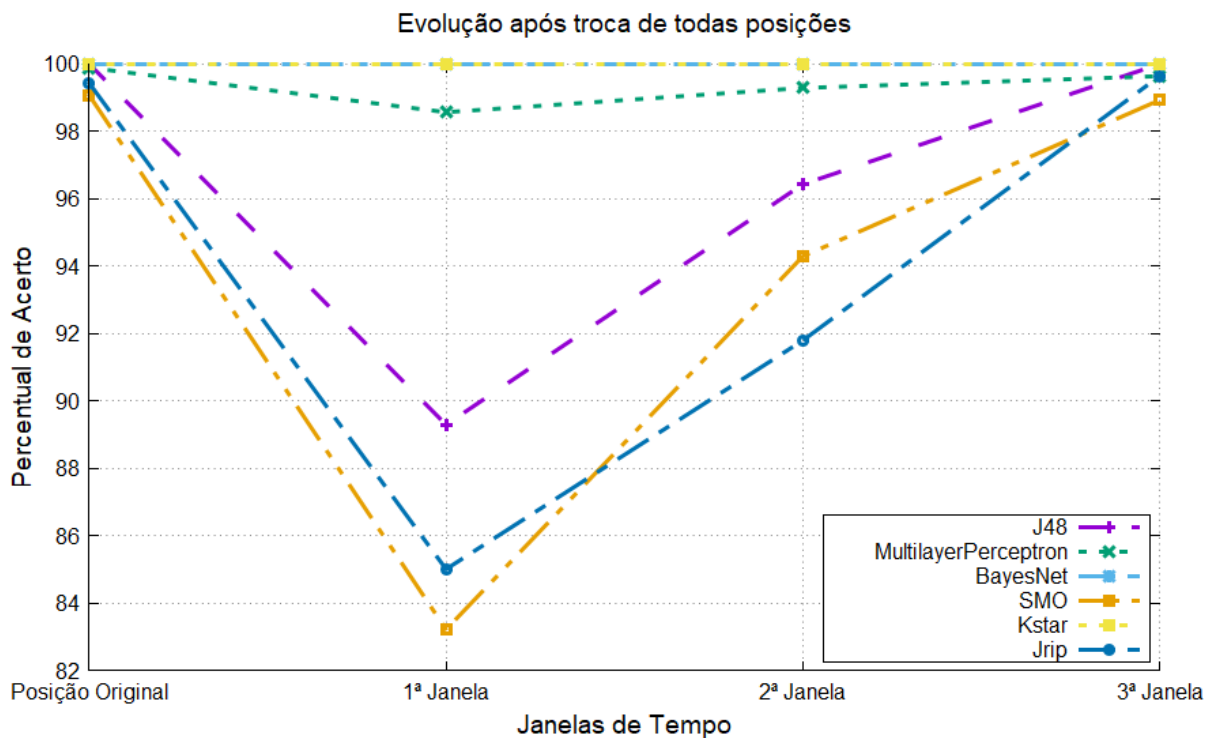
A Tabela 12 apresenta os resultados obtidos após três Janelas de Intervalos de Leitura. Nesta apuração foi possível perceber que a taxa de acerto se equiparou com a taxa de acerto da posição original da base de treinamento (Tabela 8), pois novamente os classificadores *J48*, *BayesNet* e *Kstar* obtiveram 100%. A Tabela 13 e a Figura 25 demonstram a evolução de acertos após a troca de todos os livros a cada Janela de Intervalos de Leitura em relação aos resultados obtidos na posição original.

Tabela 12: Resultados em Tempo Real: Após 3ª Janela de Intervalos de Leitura

Algoritmo	Corretas	Erradas	Total	Taxa Acerto
J48	280	0	280	100,00%
MultilayerPerceptron	279	1	280	99,64%
BayesNet	280	0	280	100,00%
SMO	277	3	280	98,93%
Kstar	280	0	280	100,00%
Jrip	279	1	280	99,64%

Tabela 13: Evolução de Acertos após a troca de todos os livros

Algoritmo	Posição Original	1ª Janela	2ª Janela	3ª Janela
J48	100,00%	89,29%	96,43%	100,00%
MultilayerPerceptron	99,89%	98,57%	99,29%	99,64%
BayesNet	100,00%	100,00%	100,00%	100,00%
SMO	99,07%	83,21%	94,29%	98,93%
Kstar	100,00%	100,00%	100,00%	100,00%
Jrip	99,43%	85,00%	91,79%	99,64%

**Figura 25: Evolução de Acertos após a troca de todos os livros**

Após a identificação do resultado da 3ª Janela de Intervalos de Leitura onde três algoritmos (*J48*, *BayesNet* e *Kstar*) conseguiram obter 100% de acerto, a investigação consistiu em trocar de posição apenas dois livros, com o objetivo de analisar o comportamento da arquitetura nesta situação. A Tabela 14 e a Figura 26 apresentam a evolução dos acertos nesta situação. Considerando como posição original o resultado obtido na 3ª Janela de Intervalos de

Leitura após a troca de todos os livros (Tabela 12). Neste caso a arquitetura precisou de duas Janela de Intervalos de Leitura para conseguir obter novamente 100% de acerto nos três melhores classificadores identificados neste ambiente.

Tabela 14: Evolução de Taxa de Acerto após troca de dois livros

Algoritmo	Posição Original	1ª Janela	2ª Janela
J48	100,00%	99,29%	100,00%
MultilayerPerceptron	99,64%	97,50%	99,29%
BayesNet	100,00%	100,00%	100,00%
SMO	98,93%	98,21%	98,25%
Kstar	100,00%	100,00%	100,00%
Jrip	99,64%	99,29%	99,64%

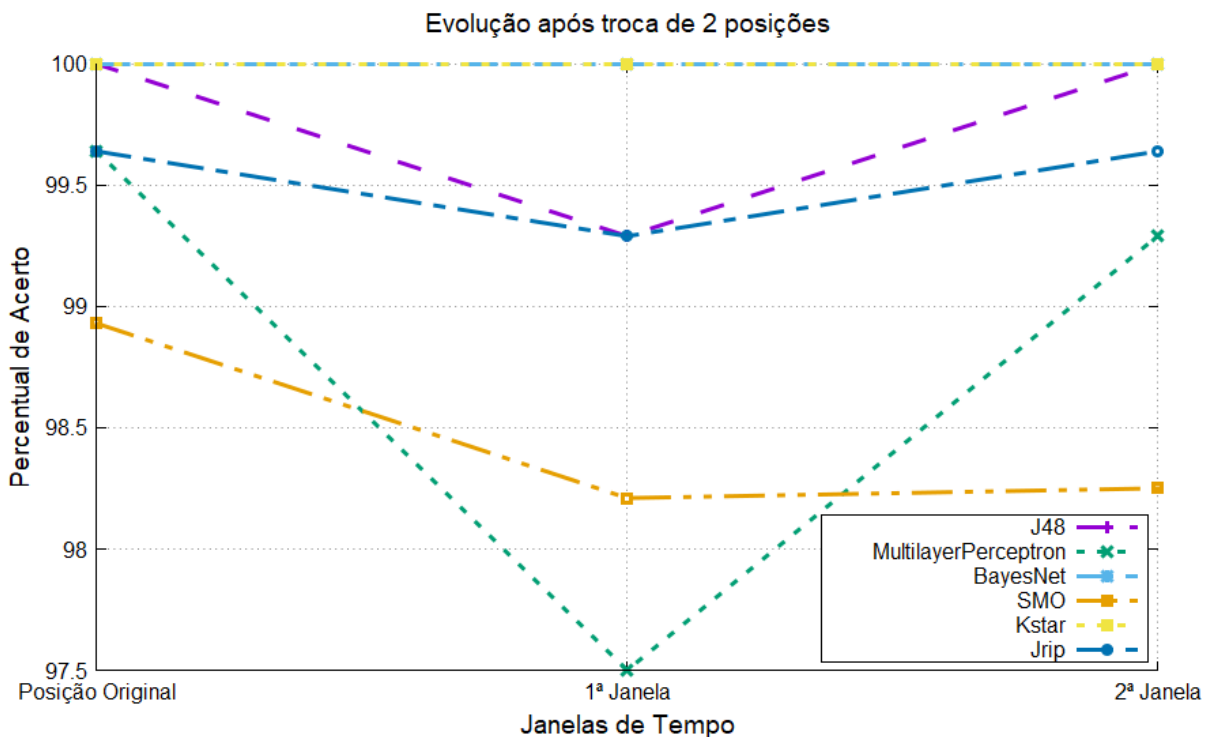


Figura 26: Evolução de Acertos após a troca de dois livros

Após esta etapa, foram realizados novos testes, onde foi trocado 6 livros de posições. A Tabela 15 e a Figura 27 apresentam a evolução dos acertos após a troca de 6 livros, considerando como posição original o resultado obtido na 2ª Janela de Intervalos de Leitura após a troca de dois livros (Tabela 14). Neste caso a arquitetura necessitou de 2 Janelas de Intervalos de Leitura para obter 100% de acerto no classificador *BayesNet* e 3 Janelas de Intervalos de Leitura para o classificador *Kstar*.

Tabela 15: Evolução de Acertos após a troca de seis livros

Algoritmo	Posição Original	1ª Janela	2ª Janela	3ª Janela
J48	100,00%	95,71%	96,07%	99,29%
MultilayerPerceptron	99,29%	97,14%	98,21%	99,29%
BayesNet	100,00%	99,64%	100,00%	100,00%
SMO	98,25%	92,86%	94,29%	98,94%
Kstar	100,00%	99,29%	99,29%	100,00%
Jrip	99,64%	96,43%	96,43%	99,29%

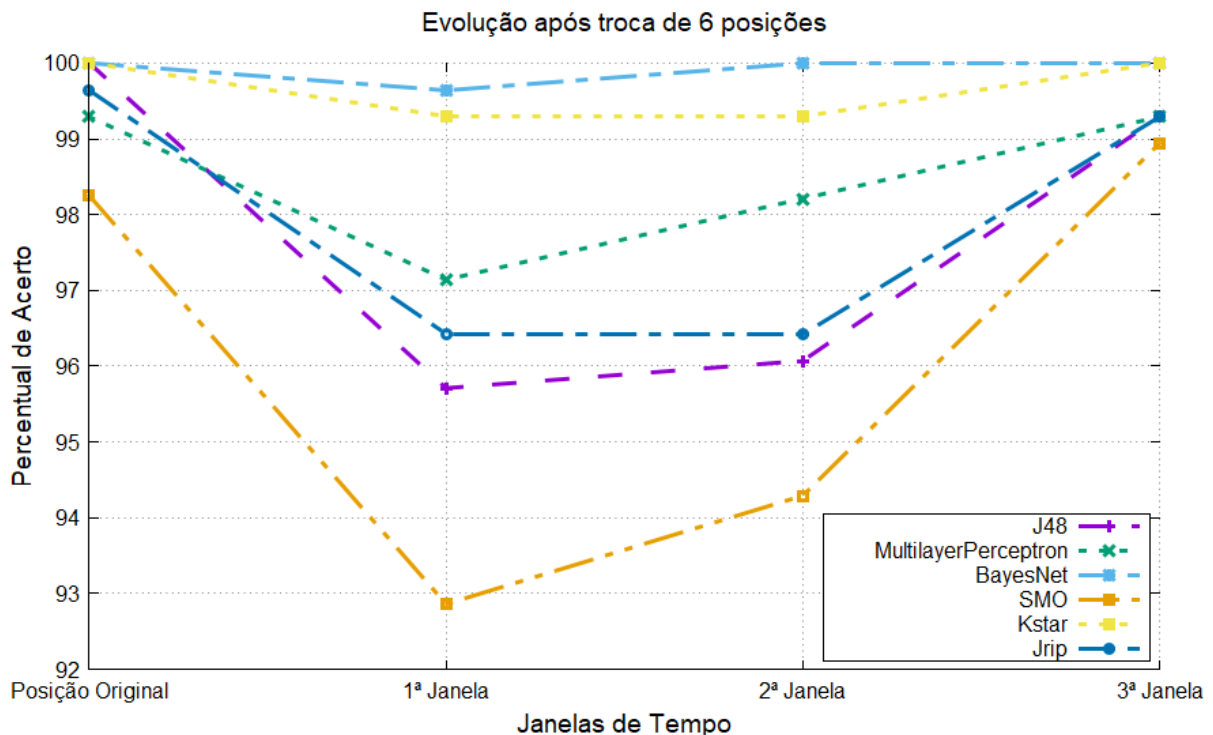


Figura 27: Evolução de Acertos após a troca de seis livros

6.3 REDUÇÃO NA QUANTIDADE DE ANTENAS

Com o objetivo de analisar como a quantidade de antenas interfere na apuração da localização, foram realizados testes considerando os atributos de uma, duas, três e quatro antenas. Para esta investigação foi utilizada a técnica de validação cruzada (Figura 18) com a base de dados de treinamento demonstrada na Tabela 8. A Tabela 16 e a Figura 28 apresentam o resultado desta investigação. Como se é esperado, é possível perceber que quanto maior a quantidade de antenas melhor é o resultado, pois cada antena possui 10 atributos em cada tupla da base. Mas deve-se considerar o resultado obtido utilizando três antenas que obteve 100% em dois classificadores (*J48* e *BayesNet*) o que pode reduzir significativamente o custo de implantação da arquitetura. O resultado obtido por duas antenas, também pode ser considerado muito bom dependendo da regra de negócio onde a arquitetura será aplicada. No caso da biblioteca

a precisão da localização dos livros é em torno de 5cm, em outros modelos de negócio onde a distância entre os objetos a serem localizados são maiores, o uso de apenas duas antenas poderá obter resultados mais significativos, o que deve ser testado caso a caso.

Tabela 16: Desempenho em relação à quantidade de antenas

Algoritmo	Uma Antena	Duas Antenas	Três Antenas	Quatro Antenas
J48	73,37%	88,75%	100,00%	100,00%
MultilayerPerceptron	72,32%	87,04%	99,53%	99,89%
BayesNet	70,60%	87,14%	100,00%	100,00%
SMO	50,71%	76,21%	93,86%	99,07%
Kstar	74,32%	87,57%	99,92%	100,00%
Jrip	72,25%	83,85%	99,12%	99,43%

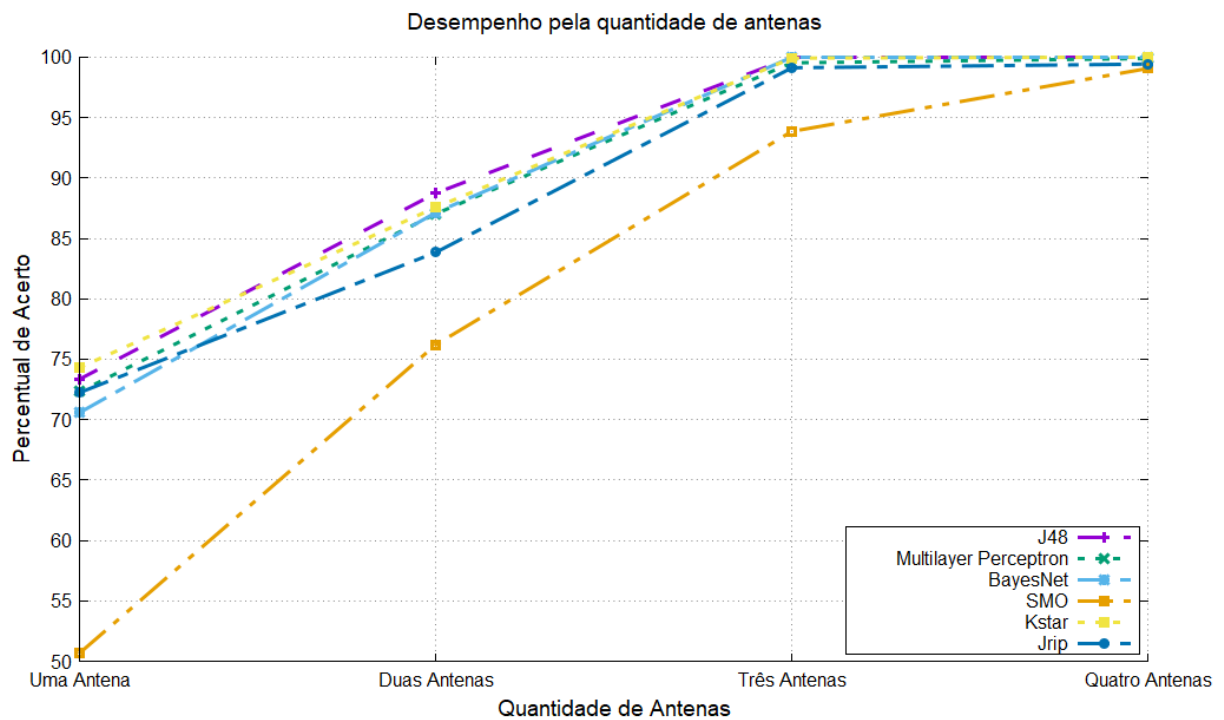


Figura 28: Desempenho em relação à quantidade de antenas

7 CONCLUSÃO

A tecnologia RFID continua sendo muito utilizada em diversos setores, principalmente em sistemas de controle de processos, estocagem, rastreamento e antifurto. Utilizar esta versatilidade da tecnologia para implementar sistemas de localização *indoor* pode contribuir ainda mais para a viabilidade econômica de implantação do sistema.

A utilização de etiquetas passivas contribui para a redução dos custos de implantação, pois além de ter um preço menor em relação às etiquetas ativas, a tendência mundial é que todos os itens já possuam etiquetas desde a fabricação, percorrendo toda a cadeia de suprimentos com a mesma etiqueta que poderá ser utilizada na arquitetura proposta. Entretanto deve-se analisar sempre os custos das antenas e leitores, pois devido ao fato da utilização das etiquetas passivas a distância de leitura é menor. Diante disso o processo de análise prévia do ambiente onde será implementado o sistema de localização *indoor* e o processo de *Site Survey* são extremamente importantes para qualquer sistema RFID.

A arquitetura RF-Miner proposta se demonstrou uma eficiente solução para o problema de localização *indoor*. Mesmo em ambientes reais onde podem existir prateleiras de metais e outros materiais reflexivos que reduzem significativamente a acuracidade, a implementação da arquitetura apresentou excelentes resultados. Nos testes de desempenho em tempo real, foi possível identificar que após uma Janela de Intervalo de Leitura o modelo já apresenta resultado de 100% de acerto (algoritmos *Bayesnet* e *KStar*) após a troca de todas as etiquetas alvo.

O módulo que vale destacar na arquitetura RF-Miner e que pode ser considerado uma das principais contribuições deste estudo, é o de Extração de Características, pois, as técnicas de derivação dos atributos contribuíram significativamente para a precisão da implementação. A técnica da derivação dos atributos pela Janela de Intervalos de Leitura pode ser utilizada em outras aplicações que necessitem trabalhar com classificadores e que possuam poucos atributos diretos, independente da utilização de sistemas RFID.

O tempo de apuração dos resultados também deve ser considerado na implementação de acordo com o modelo de negócio, pois o tempo de uma Janela de Intervalos de Leitura

definido foi de 50 segundos (10 Intervalos de Leitura de 5 segundos) somado ao tempo de processamento, pode ser muito alto e inviabilizar a aplicação. Neste caso o tamanho da Janela de Intervalo de Leitura deve ser ajustado para cada cenário, não esquecendo que se deve ter a leitura de todas as etiquetas em pelo menos uma antena. Deve-se destacar que quanto maior o valor do parâmetro Janela de Intervalos de Tempo, melhor será o resultado na apuração, mas consequentemente o tempo necessário para apuração da localização também será maior. No modelo de negócio onde foi implementada a arquitetura, (biblioteca) o tempo de 50 segundos somado ao tempo de processamento médio de 1,5 segundos são razoáveis e não interferem na aplicação.

Em relação à quantidade de antenas, como esperado, foi possível perceber que quanto maior a quantidade de antenas melhor o resultado, porém como o custo de implementação deve ser sempre considerado, com duas antenas já é possível obter 88% de acerto e com três antenas 100% de acerto na arquitetura RF-Miner. Dependendo do ambiente, estas taxas de acerto já são satisfatórias considerando a margem de precisão de 5 cm que foi utilizada nesta pesquisa, pois, em ambientes onde a margem pode ser maior devido ao tamanho dos objetos a serem identificados, consequentemente a taxa de acerto será maior. A redução da quantidade de antenas de quatro para duas, além de reduzir em 50% o custo das antenas, também reduz o custo dos leitores, já que leitores com capacidade para receber 2 antenas custam em torno de 30% a menos que leitores com capacidade para 4 antenas. A economia aproximada é de 40% na aquisição dos conjuntos de leitores e antenas além do menor consumo de energia para a operação do sistema.

Foi possível perceber que a qualidade da base de dados é imprescindível para o sucesso da implementação, e diante disso é necessário refinar todos os processos existentes na arquitetura, desde a qualidade de instalação dos equipamentos do Módulo de Leitura Física até o Módulo de Mineração de Dados para testes em outras aplicações.

Como sugestão para trabalhos futuros, vários aspectos poderão ser testados e incluídos na arquitetura. No módulo de Extração de Características outras técnicas estatísticas poderão ser utilizadas na derivação dos atributos, como Moda, Mediana, Percentis, Quartis, Amplitude, Intervalo-Interquartil, Variância e Coeficiente de Variação. No Módulo de Mineração da Dados outros algoritmos classificadores poderão ser implementados ou outras técnicas como aprendizagem de máquina utilizando *Deep Learning*.

REFERÊNCIAS

- ABUSAFIYA, M.; MAZUMDAR, S. Accommodating paper in document databases. In: ACM. **Proceedings of the 2004 ACM symposium on Document engineering**. [S.l.], 2004. p. 155–162.
- AKRE, J.-M. et al. Accurate 2-D localization of RFID tags using antenna transmission power control. **2014 IFIP Wireless Days (WD)**, p. 1–6, 2014. Disponível em: <<http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7020802>>.
- ARLOT, S.; CELISSE, A. et al. A survey of cross-validation procedures for model selection. **Statistics surveys**, The author, under a Creative Commons Attribution License, v. 4, p. 40–79, 2010.
- ATZORI, L.; IERA, A.; MORABITO, G. The internet of things: A survey. **Computer networks**, Elsevier, v. 54, n. 15, p. 2787–2805, 2010.
- AYRE, L. B. Library rfid systems for identification, security, and materials handling. **Library Technology Reports**, v. 48, n. 5, p. 9–16, 2012.
- BACHELDOR, B. Bluetag patient-tracking comes to north america. **RFID Journal**, 2008.
- BERZ, E. L.; HESSEL, F. P. Sistema híbrido de localização indoor baseado em rfid e análise visual. **Tese (Doutorado) – Faculdade de Informática, PUCRS. Porto Alegre-RS**, p. 155, 2015.
- BRÁS, L. P. M. d. A. L. **DESENVOLVIMENTO DE SISTEMA DE LOCALIZAÇÃO INDOOR DE BAIXO CONSUMO**. Tese (Doutorado), 2009.
- CHAO, Y. et al. Identification of barcode beacon and its application in underground mining. In: IEEE. **Advanced Computer Theory and Engineering (ICACTE), 2010 3rd International Conference on**. [S.l.], 2010. v. 1, p. V1–128.
- CHAWLA, V.; HA, D. S. An overview of passive rfid. **IEEE Communications Magazine**, IEEE, v. 45, n. 9, 2007.
- CHEN, T. C. Rfid and sensor-based container content visibility and seaport security monitoring system. In: **Proceedings of the International Society for Optical Engineering (SPIE)**. [S.l.: s.n.], 2005.
- CHEN, Z. N.; QING, X. Antennas for rfid applications. In: IEEE. **Antenna Technology (iWAT), 2010 International Workshop on**. [S.l.], 2010. p. 1–4.
- CHIEN, H.-Y. Secure access control schemes for rfid systems with anonymity. In: IEEE. **Mobile Data Management, 2006. MDM 2006. 7th International Conference on**. [S.l.], 2006. p. 96–96.

CHOWDHURY, B.; KHOSLA, R. Rfid-based hospital real-time patient management system. In: **IEEE. Computer and Information Science, 2007. ICIS 2007. 6th IEEE/ACIS International Conference on**. [S.l.], 2007. p. 363–368.

CLEARY, J. G.; TRIGG, L. E. et al. K*: An instance-based learner using an entropic distance measure. In: **Proceedings of the 12th International Conference on Machine learning**. [S.l.: s.n.], 1995. v. 5, p. 108–114.

COONEY, E. M. **Rfid+: The complete review of radio frequency identification**. [S.l.]: Thomson Delmar Learning, 2006.

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From data mining to knowledge discovery in databases. **AI magazine**, v. 17, n. 3, p. 37, 1996.

FINE, C. et al. The evolution of rfid networks: The potential for disruptive innovation. **MIT Sloan**, v. 8, n. 1, p. 1–6, 2006.

FINKENZELLER, K. **RFID handbook: fundamentals and applications in contactless smart cards, radio frequency identification and near-field communication**. [S.l.]: John Wiley & Sons, 2010.

FOSTER, K. R.; JAEGER, J. Rfid inside. **IEEE Spectrum**, IEEE, v. 44, n. 3, p. 24–29, 2007.

GAUKLER, G.; SEIFERT, R. Applications of rfid in supply chains. **Trends in supply chain design and management**, Springer, p. 29–48, 2007.

HALL, M. et al. The weka data mining software: an update. **ACM SIGKDD explorations newsletter**, ACM, v. 11, n. 1, p. 10–18, 2009.

HAN, J.; PEI, J.; KAMBER, M. **Data mining: concepts and techniques**. [S.l.]: Elsevier, 2011.

HAND, D. J.; MANNILA, H.; SMYTH, P. **Principles of data mining**. [S.l.]: MIT press, 2001.

HAYKIN, S. S. **Redes neurais**. [S.l.]: Bookman, 2001.

HEIDRICH, J. et al. The roots, rules, and rise of rfid. **IEEE Microwave Magazine**, IEEE, v. 11, n. 3, p. 78–86, 2010.

HESSEL, F. et al. **Implementando RFID na Cadeia de Negócios: Tecnologia a serviço da excelência**. 1. ed. [S.l.]: ediPUCRS, 2009. ISBN 978-85-7430-897-5.

HORI, T. et al. A multi-sensing-range method for position estimation of passive RFID tags. **Proceedings - 4th IEEE International Conference on Wireless and Mobile Computing, Networking and Communication, WiMob 2008**, p. 208–213, 2008.

HUANG, C. H. et al. Real-time RFID indoor positioning system based on kalman-filter drift removal and heron-bilateration location estimation. **IEEE Transactions on Instrumentation and Measurement**, v. 64, n. 3, p. 728–739, 2015. ISSN 00189456.

HULL, J. J. et al. **Method and apparatus for tracking documents in a workflow**. [S.l.]: Google Patents, mar. 1 2005. US Patent 6,860,422.

- JIN, G.-y.; LU, X.-y.; PARK, M.-S. An indoor localization mechanism using active rfid tag. In: IEEE. **Sensor Networks, Ubiquitous, and Trustworthy Computing, 2006. IEEE International Conference on**. [S.l.], 2006. v. 1, p. 4–pp.
- KELEPOURIS, T.; SILVA, S. B. D.; MCFARLANE, D. Automatic id systems: enablers for track and trace performance. **Aerospace-ID Technologies White Paper Series**, 2006.
- KELEPOURIS, T.; SILVA, S. B. D.; MCFARLANE, D. Tracking system evaluation and performance measurement: Embraer case study. **Aerospace-ID Technologies White Paper Series**, 2007.
- KHAN, M. A.; SHARMA, M.; PRABHU, H. A survey of rfid tags. Citeseer, 2009.
- KOSCHER, K. et al. Epc rfid tag security weaknesses and defenses: passport cards, enhanced drivers licenses, and beyond. In: ACM. **Proceedings of the 16th ACM conference on Computer and communications security**. [S.l.], 2009. p. 33–42.
- LAROSE, D. T. **Discovering knowledge in data: an introduction to data mining**. [S.l.]: John Wiley & Sons, 2014.
- LU, W.; HUANG, G. Q.; LI, H. Scenarios for applying rfid technology in construction project management. **Automation in Construction**, Elsevier, v. 20, n. 2, p. 101–106, 2011.
- NAKAMORI, E. et al. A new indoor position estimation method of RFID tags for continuous moving navigation systems. **2012 International Conference on Indoor Positioning and Indoor Navigation, IPIN 2012 - Conference Proceedings**, n. November, 2012.
- NASSAR, V.; VIEIRA, M. L. H. A internet das coisas com as tecnologias rfid e nfc. **Blucher Design Proceedings**, v. 1, n. 4, p. 3238–3250, 2014.
- NATH, B.; REYNOLDS, F.; WANT, R. Rfid technology and applications. **IEEE Pervasive Computing**, IEEE, v. 5, n. 1, p. 22–24, 2006.
- PLATT, J. Sequential minimal optimization: A fast algorithm for training support vector machines. 1998.
- PRERADOVIC, S.; KARMAKAR, N. C.; BALBIN, I. Rfid transponders. **IEEE microwave magazine**, IEEE, v. 9, n. 5, 2008.
- RIEBACK, M. R.; CRISPO, B.; TANENBAUM, A. S. Keep on blockin'in the free world: Personal access control for low-cost rfid tags. In: SPRINGER. **International Workshop on Security Protocols**. [S.l.], 2005. p. 51–59.
- ROH, J. J.; KUNNATHUR, A.; TARAFDAR, M. Classification of rfid adoption: An expected benefits approach. **Information & Management**, Elsevier, v. 46, n. 6, p. 357–363, 2009.
- RUSSEL, S.; NORVIG, P. Inteligência artificial. **Editora Campus**, p. 26, 2004.
- SAAB, S. S.; NAKAD, Z. S. A standalone rfid indoor positioning system using passive tags. **IEEE Transactions on Industrial Electronics**, IEEE, v. 58, n. 5, p. 1961–1970, 2011.
- SANI, A. et al. Antennas and propagation of implanted rfids for pervasive healthcare applications. **Proceedings of the IEEE**, IEEE, v. 98, n. 9, p. 1648–1655, 2010.

- SARAC, A.; ABSI, N.; DAUZÈRE-PÉRÈS, S. A literature review on the impact of rfid technologies on supply chain management. **International Journal of Production Economics**, Elsevier, v. 128, n. 1, p. 77–95, 2010.
- SARMA, S. E. et al. Rfid systems and security and privacy implications. In: SPRINGER. **CHES**. [S.l.], 2002. v. 2, p. 454–469.
- SCHMITT, P.; THIESSE, F.; FLEISCH, E. Adoption and diffusion of rfid technology in the automotive industry. In: **Proceedings of the 15th European Conference on Information Systems**. [S.l.: s.n.], 2007.
- SHIH, D.-H. et al. Taxonomy and survey of rfid anti-collision protocols. **Computer communications**, Elsevier, v. 29, n. 11, p. 2150–2166, 2006.
- SPEKMAN, R. E.; SWEENEY, P. J. Rfid: from concept to implementation. **International Journal of Physical Distribution & Logistics Management**, Emerald Group Publishing Limited, v. 36, n. 10, p. 736–754, 2006.
- SPIEGEL, M. R.; STEPHENS, L. J. **Estatística: Coleção Schaum**. 4rd. ed. Porto Alegre, RS, Brasil: Bookman Companhia Editora Ltda., 2015.
- STANTON, J. Icao and the biometric rfid passport: history and analysis. **Playing the identity card: Surveillance, security and identification in global perspective**, Routledge London, p. 253–67, 2008.
- SUNG, A. H.; MUKKAMALA, S. Identifying important features for intrusion detection using support vector machines and neural networks. In: IEEE. **Applications and the Internet, 2003. Proceedings. 2003 Symposium on**. [S.l.], 2003. p. 209–216.
- TAN, P.-N.; STEINBACH, M.; KUMAR, V. **Introdução ao datamining: mineração de dados**. [S.l.]: Ciência Moderna, 2009.
- TING, S. L. et al. The Study on Using Passive RFID Tags for Indoor Positioning. **International Journal of ...**, v. 3, n. 1, p. 9–15, 2011. ISSN 1847-9790. Disponível em: <<http://hrcak.srce.hr/65290>>.
- UCHITOMI, N. et al. Accurate indoor position estimation by Swift-Communication Range Recognition (S-CRR) method in passive RFID systems. **2010 International Conference on Indoor Positioning and Indoor Navigation, IPIN 2010 - Conference Proceedings**, n. September, p. 15–17, 2010.
- VAPNIK, V. **The nature of statistical learning theory**. [S.l.]: Springer science & business media, 2013.
- VOULODIMOS, A. S. et al. A complete farm management system based on animal identification using rfid technology. **Computers and Electronics in Agriculture**, Elsevier, v. 70, n. 2, p. 380–388, 2010.
- WANG, L.-C.; LIN, Y.-C.; LIN, P. H. Dynamic mobile rfid-based supply chain control and management system in construction. **Advanced Engineering Informatics**, Elsevier, v. 21, n. 4, p. 377–390, 2007.
- WANT, R. Enabling ubiquitous sensing with rfid. **Computer**, IEEE, v. 37, n. 4, p. 84–86, 2004.

- WANT, R. An introduction to rfid technology. **IEEE pervasive computing**, IEEE, v. 5, n. 1, p. 25–33, 2006.
- WARD, M.; KRANENBURG, R. V.; BACKHOUSE, G. Rfid: Frequency, standards, adoption and innovation. **JISC Technology and Standards Watch**, v. 5, 2006.
- WITTEN, I. H.; FRANK, E.; HALL, M. A. **Data Mining: Practical Machine Learning Tools and Techniques**. 3rd. ed. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2016. ISBN 0123748569, 9780123748560.
- WITTEN, I. H. et al. **Data Mining: Practical machine learning tools and techniques**. [S.l.]: Morgan Kaufmann, 2016.
- WU, C. et al. Will: Wireless indoor localization without site survey. **IEEE Transactions on Parallel and Distributed Systems**, IEEE, v. 24, n. 4, p. 839–848, 2013.
- XIAO, Z.-h.; GUAN, Z.-q.; ZHENG, Z.-h. The research and development of the highway's electronic toll collection system. In: IEEE. **Knowledge Discovery and Data Mining, 2008. WKDD 2008. First International Workshop on**. [S.l.], 2008. p. 359–362.
- ZHANG, X. b. et al. Towards localization of RFID tags based on experimental analysis of RSSI. **IFIP Wireless Days**, v. 2015-Janua, n. January, 2015. Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-84937803645&partnerID=40&md5=cfe2840474bcbb0d33b10d0e579bb78d>>.