



Programa de Pós-Graduação em

Computação Aplicada

Mestrado Acadêmico

Marcelo Barbosa Soares

**BEEBR: UMA PROPOSTA DE ARQUITETURA
COMPUTACIONAL NA APICULTURA, PARA A PREDIÇÃO
DE PROBLEMAS NA COLMEIA**

São Leopoldo, 2022

UNIVERSIDADE DO VALE DO RIO DOS SINOS — UNISINOS
UNIDADE ACADÊMICA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO APLICADA
NÍVEL MESTRADO

MARCELO BARBOSA SOARES

**BEEBR: UMA PROPOSTA DE ARQUITETURA COMPUTACIONAL NA
APICULTURA, PARA A PREDIÇÃO DE PROBLEMAS NA COLMEIA**

São Leopoldo
2022

Marcelo Barbosa Soares

**BEEBR: UMA PROPOSTA DE ARQUITETURA COMPUTACIONAL NA
APICULTURA, PARA A PREDIÇÃO DE PROBLEMAS NA COLMEIA**

Dissertação apresentada como requisito parcial
para a obtenção do título de Mestre pelo
Programa de Pós-Graduação em Computação
Aplicada da Universidade do Vale do Rio dos
Sinos — UNISINOS

Orientador:
Prof. Dr. Rafael Kunst

São Leopoldo
2022

S676b Soares, Marcelo Barbosa.
BeeBr : uma proposta de arquitetura computacional na apicultura, para a predição de problemas na colmeia / por Marcelo Barbosa Soares. – 2022.
141 f. : il. ; 30 cm.

Dissertação (mestrado) — Universidade do Vale do Rio dos Sinos, Programa de Pós-Graduação em Computação Aplicada, São Leopoldo, RS, 2022.
“Orientador: Dr. Rafael Kunst”.

1. Apicultura. 2. Abelhas. 3. Enxameação.
4. Aprendizado de máquina. 5. Modelo. 6. Internet das coisas (IoT). 7. Comunicação de dados móveis.
8. Predição. I. Título.

CDU: 004.85:638.1

(Esta folha serve somente para guardar o lugar da verdadeira folha de aprovação, que é obtida após a defesa do trabalho. Este item é obrigatório, exceto no caso de TCCs.)

Ao meu esforço e dedicação.

AGRADECIMENTOS

Primeiramente agradeço a uma energia muito especial que encontra-se lá em cima, na qual ajudou-me em todos os momentos difíceis nesta caminhada; A partir deste apoio, obtive força física e mental, bem como coincidências muito positivas, que favoreceram a seguir e concluir esta jornada.

Aos meus familiares, pela compreensão da minha ausência nestes dois anos de estudos, assim como, pelas palavras de incentivo e carinho que me foram dadas. Em especial, quero agradecer a duas pessoas do âmbito familiar, que sem elas, não conseguiria realizar este trabalho; Ao meu querido tio Zé, também conhecido como José Antônio Facchinetto e seu sócio Tônico, que ajudaram-me muito a conhecer o mundo da Apicultura.

Ao meu professor e orientador Rafael Kunst, pela sua disponibilidade e orientação em oferecer-me os caminhos para o conhecimento científico; estendo essa fala a todos os colaboradores do PPG (Programa de Pós-Graduação em Computação Aplicada) da Unisinos.

Destaco um agradecimento especial a TECNOSINOS e SENAI pela bolsa de estudos oferecida para a conclusão do curso.

Por fim, agradeço aqueles(as) que de alguma forma contribuíram para o desenvolvimento deste trabalho; na qual é a conquista de mais um sonho desta caminhada do saber.

RESUMO

A apicultura brasileira é um segmento formado principalmente por pequenas famílias, na qual faz da atividade um complemento na renda. O setor dispõe de métodos de trabalho obsoletos, empregando-se em sua totalidade, exercícios de forma manual e sem nenhuma intervenção tecnológica; como consequência, conduz a um processo exaustivo ao apicultor, assim como efeitos negativos à saúde das abelhas. Diante deste cenário, este estudo propõe elaborar um modelo de arquitetura computacional, em que objetiva a contribuir para à apicultura, minimizando as intervenções junto às colmeias e garantindo assim o bem estar dos insetos.

De maneira oposta a outras pesquisas, pretende-se expor do ponto de vista da contribuição científica a elaboração e treinamento de um novo modelo de aprendizado de máquina, em que mira na predição da enxameação, bem como expor soluções energeticamente sustentável e eficiente no que tange ao consumo energético dos equipamentos IoT. Por parte tecnológico, o BeeBr oferece uma solução completa de baixo custo para o segmento da apicultura.

Como resultado, o BeeBr possibilitou as leituras de seis variáveis da colmeia por um período de 20 dias. Por meio dos dados colhidos, permitiu-se uma análise estatística e a concepção de três experimentos para a avaliação do moderno modelo de aprendizado de máquina; em números finais, conseguiu-se atingir valores acima de 93% de acertos na predição da enxameação e ganhos de 16.67% em relação à eficiência energética.

Palavras-chave: Apicultura. Abelhas. Enxameação. Aprendizado de Máquina. Modelo. IoT. Comunicação de Dados Móveis.

ABSTRACT

Brazilian beekeeping is a segment formed mainly by small families, in which the activity is a supplement to income. The sector has obsolete working methods, using in their entirety exercises manually and without any technological intervention; as a consequence, it leads to an exhausting process for the beekeeper, as well as negative effects on the health of the bees. Given this scenario, this study proposes to develop a model of computational architecture, in which it aims to contribute to beekeeping, minimizing interventions with the hives and thus ensuring the well-being of insects.

Contrary to other research, it is intended to expose, from the point of view of scientific contribution, the elaboration and training of a new model of machine learning, which aims at the prediction of swarming, as well as exposing energy-efficient and sustainable solutions in terms of energy consumption of IoT equipment. On the technological side, BeeBr offers a complete low-cost solution for the beekeeping segment.

As a result, BeeBr enabled readings of six hive variables for a period of 20 days. Through the collected data, a statistical analysis and the design of three experiments were allowed for the evaluation of the modern machine learning model; in final numbers, it was possible to reach values above 93% of hits in the swarm prediction and gains of 16.67% in relation to energy efficiency.

Keywords: Apiculture. Bee. Swarm. Machine Learning. Model. IoT. Mobile Data Communication.

LISTA DE FIGURAS

1	RA (Reference Architecture)	34
2	WSN - Wireless Sensor Network	36
3	Camada de Serviço - Serviços na Nuvem	39
4	DataSet do Aprendizado Supervisionado	42
5	Fluxograma do Aprendizado Supervisionado	43
6	Algoritmo Árvore de decisão	45
7	Algoritmo KNN K-Nearest Neighbors	46
8	Evolução das Redes de Comunicação Móveis	48
9	String de busca usada por cada base de dados	54
10	Resultado da Pesquisa	55
11	Visão Geral do Modelo BeeBr	64
12	Camada de Detecção	66
13	Camada de Comunicação	70
14	Camada de Serviço	73
15	Fluxograma no provisionamento das Colmeias	74
16	Visão geral do Componente Aprendizado de Máquina	76
17	Fluxo de criação e treinamento do novo modelo de aprendizado de máquina	77
18	Resumo das classes propostas para o modelo	78
19	Protótipo da Camada de Detecção	82
20	Detalhes do Gabinete e seu dispositivos	83
21	Sensores usados na camada de detecção	84
22	Pré-Validação do Protótipo	91
23	API Rest Componente Back-End	93
24	Payload - Salvar os dados coletados da colmeia	93
25	Visão geral das tabelas	95
26	Fases de criação e treinamento do novo modelo de aprendizado de máquina	95
27	Coletas feitas no Apiário	96
28	Fusão dos Dados	99
29	Leitura Consumo Energético	104
30	Dados estatísticos descritivos das variáveis de Gás e Ruído	107
31	Dados estatísticos descritivos das variáveis de Umidade e Temperatura	107
32	Dados estatísticos descritivos da variável Peso	107
33	Comportamento da Colmeia em relação às variáveis AAE e Peso	108
34	Comportamento da Colmeia em relação às variáveis Ruído e Gás	109
35	Sensores Propolizados	110
36	Experimentos a partir da seleção das características	111
37	Experimento 1: Matrizes de confusão	112
38	Experimento 2: Matrizes de confusão	113
39	Experimento 3: Matrizes de confusão	114
40	Comparativo Irradiação Solar - Sul x Nordeste - Brasil	115
41	Coleta ID 1: Correlação entre Atividade Enxame X Peso	117
42	Coleta ID 1: Correlação entre Atividade Enxame X Ruído	117
43	Coleta ID 2: Correlação entre Atividade Enxame X Peso	118
44	Coleta ID 2: Correlação entre Atividade Enxame X Ruído	118
45	Pesquisa feita com o termo IoT no Google	133
46	Pesquisa feita com o termo Machine Learning no Google	135

47	Questionário de Inspeção	137
48	Resultados do Questionário de Inspeção	139
49	Custos da Camada de Detecção	141

LISTA DE TABELAS

1	Características dos Protocolos da camada de enlace	38
2	Comparação dos trabalhos relacionados	59
3	Composição Módulo de Entrada	67
4	Protocolos propostos empregados na camada de comunicação	71
5	Rótulos usados nos conjuntos de dados	98
6	Encoder - analisador de enxame e peso	100
7	Resultados do consumo energético	104
8	Resultados das coletas na colmeia	105
9	Descrições Estatísticas da Coleta ID 1 e ID 2	106
10	Variância dos dados coletados	110
11	Variância dos dados de inspeção	110
12	Variância das características inseridas	111
13	Métricas do Experimento 1	112
14	Métricas do Experimento 2	113
15	Métricas do Experimento 3	113
16	Comparativo de ganho na eficiência energética	114
17	Coefficiente de correlação de Pearson	117

LISTA DE ABREVIATURAS

ampl.	ampliado, -a
atual.	atualizado, -a
coord.	coordenador
N. T.	Novo Testamento
seg., segs.	seguinte, -s

LISTA DE SIGLAS

FAPERGS	Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
UMTS	Universal Mobile Telecommunications System
LoWPAN	Low power Wireless Personal Area Networks
SMOTE	Synthetic Minority Over-sampling Technique
CMOS	Complementary metal–oxide–semiconductor
ABNT	Associação Brasileira de Normas Técnicas
EDGE	Enhanced Data Rates for GSM Evolution
ITU	International Telecommunications Union
HSDPA	High Speed Downlink Packet Access
MQTT	Message Queue Telemetry Transport
API	Application Programming Interface
HSUPA	High Speed Uplink Packet Access
MEMS	Micro-electro-mechanical systems
ANS	Aprendizado Não-Supervisionado
COAP	Constrained Application Protocol
W-CDMA	Wideband Code Multiple Access
NFV	Network Function Virtualization
AMPS	Advanced Mobile Phone System
MIMO	Multiple Input, Multiple Output
REST	Representational State Transfer
TDMA	Time Division Multiple Access
CDMA	Code Division Multiple Access
GPRS	General Packet Radio Services
URIs	Uniform Resources Identifiers
AAE	Analisador Atividade Enxame
SoA	Service Oriented Architecture
WLAN	Wireless Local Area Network
WWAN	Wireless Wide Area Network
CRUD	Create, Read, Update, Delete
SIM	Subscriber Identity Module
JSON	JavaScript Object Notation
PCI	Placa de Circuito Impresso

AS	Aprendizado Supervisionado
CdC	Comportamento da Colmeia
SDN	Software Defined Networks
UAV	Unmanned Aerial Vehicles
AM	Aprendizado de Máquinas
Wsn	Wireless Sensor Networks
IaaS	Infrastructure as a Service
MLP	Perceptron Multicamadas
SVM	Support Vector Machines
DDS	Detecção-Direto-Serviço
GSM	Groupe Special Mobile
RA	Reference Architecture
KNN	K-Nearest Neighbours
SaaS	Software as a Service
LTE	Long Term Evolution
PaaS	Platform as a Service
M2M	Machine-to-Machine
CPU	Central Process Unit
WAN	Wide Area Network
RRU	Remote Radio Unit
ML	Machine Learning
IoT	Internet of Things
BBU	Baseband Unit
DP	Desvio Padrão
GW	Gateway

LISTA DE SÍMBOLOS

°C Graus Celsius

SUMÁRIO

1	INTRODUÇÃO	25
1.1	Motivação	27
1.2	Questão de Pesquisa	28
1.3	Objetivos	29
1.4	Contribuição Científica	30
1.5	Organização do Trabalho	30
2	FUNDAMENTAÇÃO TEÓRICA	33
2.1	IoT	33
2.1.1	Camada de detecção	35
2.1.2	Camada de rede	35
2.1.3	Camada de serviço	38
2.1.4	Camada de interface	40
2.2	Aprendizado de Máquina	41
2.2.1	Aprendizado Supervisionado	41
2.2.2	Aprendizado Não-Supervisionado	44
2.2.3	Algoritmos	45
2.3	Comunicação de Dados Móveis	47
2.3.1	Passado: 1G, 2G, 3G, 4G	47
2.3.2	Presente: 5G	49
2.3.3	Futuro: 6G	50
3	TRABALHOS RELACIONADOS	53
3.1	Metodologia da Pesquisa	53
3.2	Resultado da Pesquisa	55
3.3	Discussão da Pesquisa	56
3.4	Análise dos trabalhos relacionados	58
3.5	Oportunidades de pesquisa	60
4	MODELO PROPOSTO	63
4.1	Visão geral do modelo	63
4.2	Camada de Detecção	65
4.2.1	Módulo de Entrada	66
4.2.2	Módulo de Processamento	68
4.2.3	Módulo de Saída	68
4.2.4	Módulo de Alimentação	69
4.3	Camada de Comunicação	69
4.3.1	Redes e Protocolos	69
4.3.2	Modos de Transmissão	71
4.3.3	Provisionamento	72
4.4	Camada de Serviço	72
4.4.1	SaaS - Multitenancy	73
4.4.2	Entrada de Dados	74
4.4.3	Componentes	75
4.4.4	Componente - Back-End / Front-End	75
4.4.5	Componente - Banco de Dados	75
4.4.6	Componente - Aprendizado de Máquina	76

4.5	Elaboração e treinamento do novo modelo de aprendizado de máquina	76
5	IMPLEMENTAÇÃO DO MODELO	81
5.1	Camada de Detecção	81
5.1.1	Desenvolvimento do Protótipo	81
5.1.2	Módulo Entrada	83
5.1.3	Módulo de Processamento	85
5.1.4	Módulo Saída	88
5.1.5	Módulo Alimentação	88
5.1.6	Pré-Validação do Protótipo	90
5.2	Camada de Serviço	92
5.2.1	Desenvolvimento Componente - Back-End	92
5.2.2	Desenvolvimento Componente - Banco de Dados	94
5.3	Elaboração e treinamento do novo modelo de aprendizado de máquina	94
5.4	Considerações	102
6	RESULTADOS E DISCUSSÕES	103
6.1	Resultados	103
6.1.1	Sustentabilidade e eficiência energética	103
6.1.2	Dados coletados na colmeia	105
6.1.3	Predição dos dados pelo novo modelo de aprendizado de máquina	109
6.2	Discussões	114
6.2.1	Sustentabilidade e eficiência energética	114
6.2.2	Dados coletados na colmeia	116
6.2.3	Predição dos dados pelo novo modelo de aprendizado de máquina	120
6.3	Considerações	122
7	CONSIDERAÇÕES FINAIS	123
7.1	Conclusão	123
7.2	Contribuições	124
7.3	Limitações	125
7.4	Trabalhos Futuros	125
	REFERÊNCIAS	127
	APÊNDICE A – PESQUISA FEITA COM O TERMO IOT NO GOOGLE	133
	APÊNDICE B – PESQUISA FEITA COM O TERMO ML NO GOOGLE	135
	APÊNDICE C – QUESTIONÁRIO DE INSPEÇÃO	137
	APÊNDICE D – RESULTADOS DO QUESTIONÁRIO DE INSPEÇÃO	139
	APÊNDICE E – CUSTOS DA CAMADA DE DETECÇÃO	141

1 INTRODUÇÃO

À apicultura no Brasil é uma atividade na qual iniciou oficialmente no ano de 1839, e desde então, tornou-se uma valiosa fonte de renda para diversas famílias brasileiras (ABELHAS, 2021a). Segundo dados da Confederação Brasileira de Apicultura (2019) os apicultores brasileiros são predominantemente de pequeno porte, em que a atividade complementa seu sustento e quase metade dos produtores no País possuem até 50 colmeias. A indústria da apicultura por meio de seus produtos, vêm conquistando cada vez mais consumidores no mundo, gerando perspectivas de crescimento em sua participação no agronegócio brasileiro, conforme Ahmad Saeed et al. (2014). Atualmente a China é o maior produtor mundial de mel; o Brasil figura em 11º lugar na produção global, apresentando evolução marcante entre os principais países produtores, de acordo com pesquisa do IBGE ¹.

As atividades da apicultura ocorrem em locais destinados para esse fim, denominado de apiário, onde devem estar dispostos por questão de segurança em localidades não frequentadas por pessoas (ABELHAS, 2021b). Dentre as tarefas que os apicultores desempenham nos apiários estão a exploração do mel, principal objetivo do segmento, além da extração de outros produtos produzidos pelas abelhas, como cera, pólen apícola, própolis, geleia real, enxames e crias (Ahmad Saeed et al., 2014).

Dessa maneira, para obter uma boa produção dos itens que as abelhas produzem, é importante o apicultor definir uma agenda para o manejo das colmeias, em que consiste em inspeções periódicas para avaliar a saúde das abelhas (ABELHAS, 2021c). Estas visitas aferem a presença de pragas (formiga e traça), possíveis doenças, como enxameação (fuga de abelhas), colmeia zanganeira (somente produção de zangão), presença ou não da abelha rainha, estado dos favos, produção de mel, o peso da colmeia na qual reflete na saúde e produtividade da colônia e dentre outros (TECNICO, 2007).

Os procedimentos nas quais o apicultor precisa realizar a fim de revisar os apiários constituem em vestir a indumentária para sua segurança, acender o fumigador, deslocar-se até o apiário, aplicar a fumaça no alvado, retirar a tampa da caixa e assim iniciar a inspeção (APRENDIZAGEM RURAL, 2010). Apesar das novas técnicas adotadas na preservação do enxame e colmeias (BERALDO; BARBOZA, 2011), este processo implica em consequências negativas como um todo dentro da apicultura. A primeira delas são os problemas na saúde e impacto no bem estar das abelhas, dado que o método causa grande excitação e esgotamento aos insetos, implicando na diminuição da produção (TECNICO, 2007). Para os apicultores brasileiros, constitui-se uma tarefa árdua, feita em sua íntegra de forma manual, sem nenhuma intervenção tecnológica (Mesquita et al., 2020). Tendo em vista que os apiários localizam-se em regiões distantes, onde em algumas não contam com conectividade de internet ou telefonia móvel, as quais, viabilizaria auxiliar e facilitar nas etapas da missão. Soma-se a um processo exaustivo e suscetíveis a falhas, em que há necessidade de repetição da operação em todas as colmeias do

¹Instituto Brasileiro de Geografia e Estatística website: <<http://www.ibge.gov.br>>.

apiário, e, finalmente, a insegurança da operação.

A partir deste cenário exibido, buscou-se na literatura o estado da arte, a fim de verificar perspectivas e soluções diante ao exposto. Constatou-se que os estudos expõem lacunas diante às questões relacionadas a predição de eventuais dificuldades vistas na caixa, em específico a enxameação e colmeia zanganeira, bem como, a carência de novos atributos na construção de novos modelos de aprendizado de máquina. Também fora encontrada vácuos nas áreas de sustentabilidade e eficiência energética, assim como, respostas alternativas diante às adversidades relativas às transmissões de dados.

Em cima desta situação na qual apresenta-se à área da apicultura e acrescido ao que existe atualmente de conhecimento no segmento tecnológico, percebe-se a oportunidade de buscar melhorias nas atividades deste setor, baseando-se no emprego computacional. Esses avanços objetivam contribuir no trabalho do apicultor, minimizando as intervenções junto às colmeias e garantindo assim o bem estar das abelhas. Neste sentido, destacam-se às tecnologias voltadas à monitoração inteligente de objetos que buscam trazer importantes noções ao usuário. Segundo Kridi et al. (2014) o monitoramento possibilita verificar o comportamento e saúde das abelhas, e a partir disto, definir formas de melhorar a produção e prevenir doenças na colônia.

Diante do exposto, este trabalho objetiva como propósito principal, elaborar um modelo de arquitetura computacional para a apicultura, na qual nesta dissertação é nominada e referenciada por *BeeBr*. Seu propósito focal é monitorar e gerenciar os apiários, partindo das seguintes premissas: (1) Sustentabilidade e eficiência energética; (2) Manipulação inteligente nos dados captados do apiário, por meio de IoT, na qual, a partir destes, seja capaz a predição de possíveis problemas que venham a ocorrer na colmeia, assegurando desta forma, o bem estar das abelhas e colaborando junto ao apicultor, mediante a diminuição do manejo ; (3) Flexibilidade no ponto de vista relacionada à transmissão dos elementos coletadas no apiário, visto que, conforme citado, a conectividade móvel é inexistente em algumas regiões. (4) Por fim, dado ao exposto cenário da apicultura Brasileira, formada por pequenas famílias, nas quais dispõem de insuficiente base tecnológica e de baixo poder em investir em seus apiários, a fim de agregá-los tecnologicamente, propõem-se um modelo de baixo custo operacional. Com a intenção de atingir tais objetivos expostos e diferenciar-se do estado da arte, esta dissertação mira procurar soluções em campos de estudos de interesse deste pesquisador, dentre elas, *IoT, comunicação de dados móveis e aprendizado de máquinas*, ntais quais são os pilares para a elaboração do *BeeBr*.

Em frente ao narrado, este estudo apresenta como contribuição científica, uma proposta de arquitetura computacional para a apicultura, denominada *BeeBr*; na qual oferece a elaboração e treinamento de um novo modelo de aprendizado de máquina, em que busca a predição de problemas encontradas na caixa, em particular à enxameação. Finalmente, expor soluções energeticamente sustentável e eficiente no consumo energético por parte dos equipamentos IoT.

1.1 Motivação

A motivação para iniciar esta pesquisa, moveu-se a partir de conversas com familiares que exercem a atividade da apicultura. Nestas discussões debateu-se informações gerais do segmento, desde aspectos básicos até avançados. Quais equipamentos são usados na atividade ? Como é o processo para captar as abelhas nas caixas ? Qual o fluxo completo do manejo e sua importância ? Essas foram algumas questões abordadas nos debates, bem como o procedimento de colheita, produção e venda do mel, além dos problemas e dificuldades encontradas na área, como por exemplo, as presenças de formiga, traças e vespas, que são umas das principais pragas na apicultura, e por fim, as doenças de enxameação, nas quais as abelhas deixam as colmeias (TECNICO, 2007). Tais ideias foram importantes para um entendimento amplo da área, visto que até então a apicultura era um negócio totalmente desconhecido por parte deste pesquisador.

A partir desta compreensão, percebeu-se a oportunidade em descobrir avanços nesta atividade, dado que é uma área de pouco desenvolvimento tecnológico para a grande parte do público que a pertence, formada em sua maioria de apicultores de pequeno porte (APICULTURA, 2019). Dentre as tarefas executadas na apicultura, entendeu-se que o manejo é a etapa a ser desenvolvida tecnologicamente, pois identificou-se por parte deste pesquisador, a fase mais exaustiva, perigosa, repetitiva em todo o processo da colheita do mel. Em cima do desejo de ajudar essas famílias, empregando a tecnologia em seu cotidiano, veio a motivação para desenvolver esta dissertação.

Com base no interesse em sanar as adversidades encontradas no manejo da apicultura, oportunizou-se o momento de buscar soluções em cima de áreas de estudos, tais como, *IoT*, *comunicação de dados móveis* e *aprendizado de máquina*, nas quais, entende-se por parte deste investigador, que seja as bases para a construção de uma proposta, que venha a sanar tal desejo.

Desta forma, procurou-se na literatura trabalhos que auxiliassem nesta atividade, fazendo uso dos segmentos citados, além de examinar de modo geral, as soluções comerciais, a fim de entender o que existe atualmente. Diante das pesquisas realizadas, foram verificados estudos que realizavam monitoramento junto a colmeia, apresentando resultados de algumas variáveis do apiário. Mas, conforme exposto em seção 3.4 constatou-se lacunas referente a análise destes dados, dificuldades na comunicação das informações entre apiário/nuvem e, a não preocupação do gasto energético da infraestrutura usada. Por fim, as soluções comerciais existentes dispõem de alto custo operacional, inviabilizando sua implantação para os padrões nacionais.

Portanto, este estudo tem a oportunidade de implementar uma prova de conceito denominada BeeBr, na qual contemple às áreas de *IoT*, redes de comunicação móveis e algoritmos de aprendizagem de máquinas, na qual ofereça tecnologia ao segmento, a fim de reduzir o manejo e consecutivamente, dispor melhorias no bem estar das abelhas. Por meio de um novo modelo de aprendizado de máquina, que busca a preempção de anomalias na caixa, em particular à enxameação, bem como, ofertar recursos computacionais, na qual possibilite ao BeeBr realizar sua atividade. Finalmente, transformar o BeeBr em um projeto de impacto social, tornando

um produto simples e de baixo custo, na qual contribua e auxilie o trabalho dos apicultores brasileiros.

1.2 Questão de Pesquisa

Este trabalho busca responder a seguinte questão de pesquisa:

Como seria um modelo de arquitetura de software na apicultura, que desfrute de sustentabilidade energética, flexibilidade na comunicação de dados, e, explore IoT e técnicas de predição, na qual objetiva minimizar o uso do manejo e assim, assegurar o bem estar das abelhas ?

A fim de responder ao questionamento, é importante o entendimento do contexto de alguns termos citados na pergunta:

- Sustentabilidade energética: oferecer uma solução de energia sustentável, em que disponha à camada de IoT do modelo, executar sem a interferência de uma distribuidora elétrica, bem como, atender de forma eficiente, no que tange o seu consumo.
- Flexibilidade na comunicação de dados: dispor mecanismos que viabilizem ao modelo cumprir sua função, mesmo não havendo redes de dados.
- IoT: área na qual refere-se a coleta das variáveis do apiário, por meio de dispositivos da internet das coisas, a fim de desfrutar das informações para o monitoramento e preempção de problemas na colmeia.
- Técnica de Predição: nesta pesquisa, atém-se ao estudo dos algoritmos de aprendizado de máquina; baseando-se nos elementos levantados na caixa, objetiva buscar a predição de anomalias encontradas no segmento, a fim de diminuir o emprego do manejo.
- Bem estar das abelhas: define-se este termo, englobando as características de redução do manejo manual; afastamento de doenças e pragas encontradas na caixa; precaução em assegurar que a colmeia tenha a presença da abelha rainha; ou qualquer problema/evento que prejudique a saúde das abelhas. Tais aspectos identificam, descrevem e caracterizam o bem estar das abelhas.

A partir dos termos expostos, entende-se que este trabalho busca oferecer um modelo computacional, designado de BeeBr; na qual permita trazer benefícios ao setor da apicultura brasileira, focando em especial na etapa do manejo, o que consecutivamente obtém-se ganhos para a saúde das abelhas. Baseado na ideia de ser um projeto altamente sustentável do ponto de vista elétrico, por parte dos dispositivos IoT e que disponha de eficiência energética. Versátil em relação a comunicação das informações, trazendo expectativas ao usuário que não dispõem

de uma infraestrutura de redes de dados. O BeeBr deve dispor da possibilidade de implementar algoritmos da área de aprendizado de máquina a fim de prever anomalias e problemas na colmeia, tais como, à enxameação (fuga de abelhas), colmeia zanganeira (somente produção de zangão) e presença ou não da abelha rainha. Usufruir das técnicas do segmento da internet das coisas, possibilitando desta forma, a monitoração junto aos apiários. Além de uma infraestrutura pensada em atingir uma maior gama de apicultores. Para o desenvolvimento do BeeBr, propõem-se a criação de um protótipo para coleta de dados no apiário, na qual, em cima destes valores, pretende-se trazer elementos que busquem minimizar o manejo, assegurando assim, o bem estar das abelhas e uma melhora no trabalho do apicultor.

1.3 Objetivos

Esta dissertação dispõe como principal objetivo elaborar um modelo de arquitetura computacional nominado de BeeBr, aplicado ao segmento da apicultura, na qual propõem-se a trazer benefícios para a área. Essas melhorias têm como foco levar ao apicultor uma solução tecnológica de custo baixo, em que busca auxiliar em suas rotinas, em específico ao do manejo. Além de ganhos para às abelhas, dado que, evitando-se as inspeções e dispondo de sistemas inteligentes que sejam capazes de monitorar e prever situações desagradáveis aos insetos, tende a trazer impactos positivos na saúde e bem estar destes animais. Na concepção do BeeBr, pensou-se em uma solução completa, com o intuito de atender ao ponto central, na qual deve-se dispor da possibilidade de implementar algoritmos de aprendizado de máquina, flexibilidade na transferência de dados à nuvem, recursos de hardware e software a fim de coletar e processar as informações dos apiários, alternativas no desenvolvimento sustentável, bem como, uma solução de predição com foco na enxameação. Seguem descrições dos objetivos específicos, com o propósito de responder ao ponto fundamental.

- Idealizar um modelo computacional que disponha em suas características, funcionalidades que atendam ao propósito principal.
- Desenvolver um protótipo de hardware, que seja capaz de monitorar e coletar às variáveis do apiário, na qual é formado por elementos IoT.
- Planejar e desenvolver soluções que atendam a área de sustentabilidade e eficiência energética, no que tange as necessidades dos dispositivos de IoT.
- Desenvolver componentes de softwares que processam as informações levantadas no apiário.
- A partir dos dados coletados, elaborar e treinar um novo modelo de aprendizado de máquina, constituído por classes oriundas dos elementos coletados da colmeia por meio de modernos sensores IoT, bem como, de pesquisas efetuadas em campo, a partir de especia-

listas do segmento, a fim de trazer predição de eventuais eventos na colmeia, em especial à enxameação (abandono das abelhas da colmeia).

- Validar o modelo através das implementações realizadas (protótipo de HW e componentes de SW desenvolvidos) dentro de um apiário real.
- Dispor mecanismos que ofereçam ao apicultor usufruir os recursos do BeeBr, mesmo não dispondo de comunicação.

1.4 Contribuição Científica

Este trabalho pretende colaborar ao segmento da apicultura a partir de uma proposta de modelo de arquitetura de software denominada BeeBr; em que oferta aspectos tecnológicos, preditivos e sustentáveis energeticamente, baseado nas áreas de estudos de IoT, comunicação de dados móveis e aprendizado de máquina. A partir do BeeBr, pretende-se atingir os objetivos propostos e assim auxiliar a comunidade científica.

Neste sentido, esta dissertação apresenta as seguintes contribuições científicas; propõe-se primeiramente a elaboração e treinamento de um novo modelo de aprendizado de máquina, que tem o propósito na predição de eventos referentes à problemas na caixa, em particular a enxameação. Tanto a finalidade quanto na concepção das classes dispostas neste modelo, não são encontrados no estado da arte. Atributos estes, originados por meio de novos sensores IoT, bem como, por coletas praticadas em campo (apiário).

Finalmente, expor soluções energeticamente sustentável e eficiente no que tange ao consumo energético por parte dos equipamentos IoT, nas quais efetuam as leituras dos elementos da colmeia; tais resultados não foram encontrados nos trabalhos relacionados.

1.5 Organização do Trabalho

Este estudo está organizado em 7 capítulos, nas quais estão estruturados da seguinte maneira. O primeiro exhibe a definição do problema, motivação, questão de pesquisa, objetivos e a contribuição científica. O capítulo 2 descreve a fundamentação teórica, conceitos estes, necessários para a compreensão da proposta; são expostos os temas IoT, aprendizagem de máquina e comunicação de dados móveis, sendo os pilares para a elaboração do BeeBr. O capítulo 3 apresenta os trabalhos relacionados encontrados na literatura, mostrando comparações, discussões e análises destes; por fim as oportunidades de pesquisas. O capítulo 4 expõem o modelo proposto por esta dissertação nominado de BeeBr; no capítulo seguinte é visualizada a implementação do BeeBr, expondo a metodologia de desenvolvimento da camada de detecção, serviço e a elaboração e treinamento do novo modelo de aprendizado de máquina. O capítulo 6 mostra os resultados e discussões obtidas a partir da avaliação dos temas: (1) Sustentabilidade e eficiência energética; (2) Dados coletados na colmeia; (3) Predição do novo modelo de aprendizado de

máquina, nas quais englobam as conclusões referentes aos objetivos desta dissertação. Por fim, o último capítulo exhibe as conclusões, contribuições, limitações, trabalhos futuros e trabalhos publicados por este pesquisador.

2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo serão apresentados os conceitos teóricos necessários a fim de entender o modelo computacional BeeBr, bem como, os fundamentos para alcançar os objetivos propostos por essa dissertação. Baseado no entendimento das áreas expostas, torna-se possível a percepção das tecnologias disponíveis que podem ser empregadas neste trabalho. Este capítulo organiza-se a partir de três campos de estudos que são os pilares na concepção do BeeBr; primeiro expõem os princípios da internet das coisas, o segundo aborda o tema aprendizado de máquina e finalmente é exibido ideias sobre comunicação de dados móveis.

2.1 IoT

A primeira área de estudo que objetiva na elaboração do BeeBr, surgiu em uma apresentação para executivos da empresa P&G, em meados de 1999, o então pesquisador britânico Kevin Ashton usou pela primeira vez o termo IoT (Internet of Things) (Ashton, 2008). Após alguns anos, o acrônimo começou a levantar o interesse da comunidade, principalmente a partir de 2013, quando as buscas pelo termo cresceram de forma exponencial, conforme apresentado no Apêndice A. Desde então, o IoT tornou-se uma tecnologia em crescimento e evolução, e hoje temos um ecossistema composto por diversos elementos, como sensores, microcomputadores, software embarcados e sistemas de alta complexidade, que possibilitam a Internet das Coisas tornar-se hoje uma realidade.

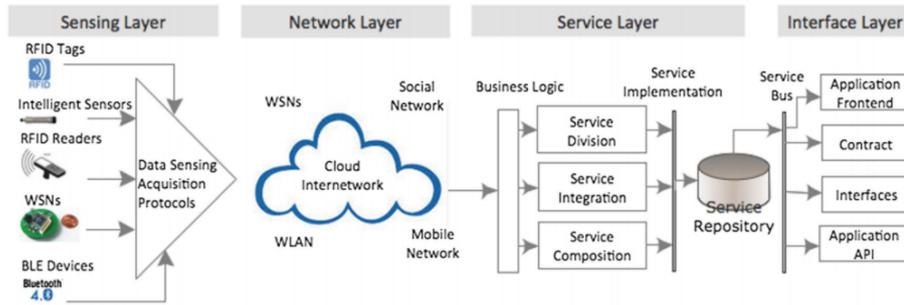
A IoT atualmente é presente em diversos segmentos da sociedade, como em linhas de produção de fábrica, que a partir do monitoramento do processo de fabricação são possíveis realizar a manutenção proativa de equipamentos; câmeras de segurança alertando quando um setor for violado, relógios que verificam os batimentos cardíacos de um paciente e aponta um possível problema; sistemas para a agricultura que monitoram o campo, verificando a umidade do solo e temperatura; em automações residenciais, nas quais os eletrodomésticos auxiliam na gestão e controle do ambiente e por fim nas cidades inteligentes (Islam et al., 2015).

Dentro da literatura, não há consenso definido sobre IoT, visto que existem visões distintas de diferentes pesquisadores (Perera et al., 2014), mas pode-se definir como um novo paradigma de conectividade de equipamentos, na qual é constituída por rede(s) de dado(s), composta por objetos físicos que enviam pacotes de conhecimento à uma estação base de comunicação; através desta, encaminha os dados para aplicações que dispõem da responsabilidade de processar e armazenar, sem a intervenção humana em todo o processo.

Dentre os pensamentos, destaca-se a arquitetura proposto pelos autores Roman e Lopes (2009) e encontrada em demais artigos (Li; Xu; Zhao, 2015; Miladinovic; Schefer-Wenzl, 2018; Guth et al., 2016), na qual busca dispor camadas de abstrações e fluxos de comunicação diferenciada (Miladinovic; Schefer-Wenzl, 2018). Baseia-se no conceito SoA (Service Oriented Architecture), em que propõe a segmentação da IoT em quatro camadas. Nominalmente é ex-

posto a *Camada de detecção*: hardware e sensores; *Camada de rede*: infraestrutura; *Camada de serviço*: serviço e aplicações; *Camada de interface*: interligação entre usuário e aplicações. Na Figura 1 é exibido em detalhes a arquitetura sugerida, em que, a partir deste ponto é referenciada como RA (Reference Architecture).

Figura 1 – RA (Reference Architecture)



Fonte: Li, Xu e Zhao (2015)

Na *Camada de detecção* encontra-se os objetos físicos que devem oferecer a capacidade de monitorar e analisar elementos do ambiente em que depara-se, executar ações e transmitir os dados pela rede. A efetivação deste processo, faz necessário a composição de sensores, atuadores e microcontroladores. Os sensores são responsáveis pelo monitoramento das variáveis do espaço que se situa. Os atuadores têm o trabalho de executar tarefas no mundo externo, como movimentar um objeto de local. Finalmente os microprocessadores detêm do processamento e oferecem a conectividade para todo o processo envolvido (Roman; Lopes, 2009).

Dentro da *Camada de redes* apresenta-se uma infraestrutura que têm o objetivo de levar as informações da camada de detecção até a de serviço. Para atingir este fim, compõem-se uma estrutura de redes de dados, elaborada e implementada para suportar as adversidades e desafios encontrados no segmento da IoT. Uma das soluções propostas e desenvolvidas é a WSN (Wireless Sensor Networks), que nasceu com o intuito principal de obter material do ambiente onde se encontra (Roman; Lopes, 2009), sendo responsável pela comunicação, interligação dos objetos da camada de detecção. A WSN é caracterizada por ser uma rede de baixo custo, com restrições de energia, processamento limitado e alcance reduzido. Outra rede que faz parte desta camada é a Internet, na qual é denominada de "rede de disseminação", em que é encarregado por transmitir os dados para a camada seguinte.

A camada definida como *Camada de serviço* goza da incumbência de processar os elementos recebidos pela camada de detecção e, a partir dos materiais coletadas e recebidos, são executadas as rotinas de processamento e armazenamento, a fim de que os dados estejam disponíveis para os usuários e aplicações (Li; Xu; Zhao, 2015). Consiste em diversos sub-processos processados por aplicações, na qual uma destas é a etapa analítica, que a partir de algoritmos de análises de dados, consegue-se extrair informações relevantes e de valor para o usuário. As aplicações disponibilizam APIs (Application Programming Interface) como uma interface para envio/recebimento de dados e na sua grande maioria estão hospedados na nuvem.

Por fim, a *Camada de interface* apresenta um barramento heterogêneo, disposto para as aplicações e serviços que venham a ter acesso ao conhecimento armazenado na camada de serviço (Li; Xu; Zhao, 2015). Tendo em vista que os dispositivos da camada de detecção são desenvolvidos por fabricantes diferentes, é de suma importância haver uma interface simples e fácil para que sistemas ou qualquer meio, consiga acessar os dados. Nas próximas subseções são detalhados e discutidos o comportamento de cada camada.

2.1.1 Camada de detecção

Os objetos desta camada contam com a capacidade de ler os elementos do ambiente em que encontram-se, executar atividade físicas ou lógicas e transmitir pacotes para as camadas superiores. Qualquer elemento que seja capaz de realizar tais atividades, são definidos dentro do contexto IoT como "Things"; sendo constituída por 1 ou mais sensores e atuadores inteligentes que estão interligados a um microcontrolador (Kocakulak; Butun, 2017).

Os sensores convertem as variáveis do ambiente como luz, fumaça, temperatura e som em sinais elétricos, sendo estes transmitidos ao microcontrolador. Do lado oposto, existem os atuadores, que exercem o processo inverso ao de um sensor, transformando um sinal elétrico em uma grandeza física, como movimentos, magnetismo e calor. O microcontrolador, dispositivo composto por CPU (Central Process Unit), memória RAM, conexões para entradas de módulos externos para expansão, como placa de rede e as conectividades para os sensores e atuadores, na qual Caracterizam-se no segmento IoT por limitada capacidade de processamento, baixo custo e eficiência energética.

Os dados coletados podem ser armazenados localmente a partir de banco de dados e/ou pendrive ou por meio de módulos de expansão, na qual possibilita-se transmitir via Ethernet, WI-FI, rádio e redes móveis. Na seção seguinte (2.1.2) é discutido a transmissão dos dados da camada de detecção para as superiores.

2.1.2 Camada de rede

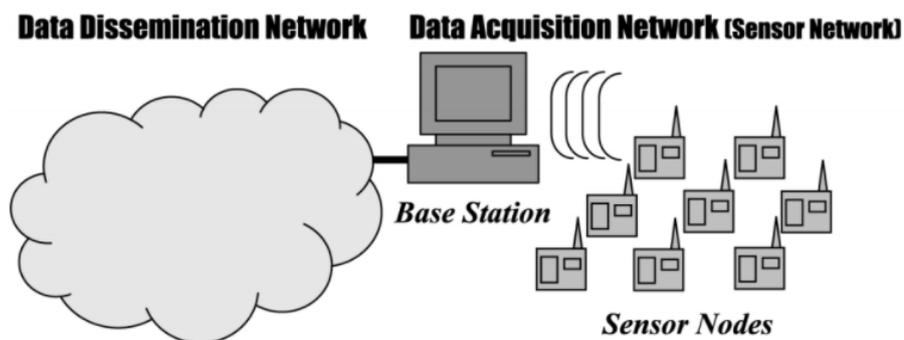
Os dados provenientes da camada de detecção são transmitidas para as superiores, mediante a uma agregação heterogênea de redes de comunicação, denominadas WSN (Wireless Sensor Network) e Internet (Li; Xu; Zhao, 2015). A WSN foi desenvolvida para atender o seguinte propósito: atuar em redes que dispõem de sensores que monitoram, obtém e alertam sobre as variáveis e condições do ambiente. Caracteriza-se por ter recursos de auto configurabilidade, autonomia, facilidade de implantação (Roman; Lopes, 2009), além de tamanho, preço, capacidade de processamento, armazenamento e transmissão de dados reduzidos e por fim, restrições de energia.

A Wireless Sensor Network é composta por dois elementos centrais, denominados "nós sensores" e "estação base"; o primeiro elemento tem a responsabilidade de capturar as informações

do ambiente e transmitir para a estação base. No olhar da camada de rede, uma "Things" é nominada como "nós sensores" (Roman; Lopes, 2009). A estação base por sua vez é um objeto titulado de GW (Gateway), em que atua no recebimento dos pacotes oriundos dos nós sensores, e os encaminha para as redes de disseminação. Segundo Roman e Lopes (2009) podem haver 3 tipos de GW, sendo eles, "front-end proxy solution", "gateway solution" e "TCP/IP Overlay Solution", na qual as diferenças estão principalmente, na questão de acesso e segurança dos nós sensores em relação a rede de dispersão (Internet).

Existem cenários de implantação de IoT, em que os nós sensores são capazes de transferir seus pacotes diretamente à Internet, sem o intermédio de um GW e assim, seus dados serem recebidos pela camada de serviço. Na Figura 2 é exposta uma visão geral da rede WSN.

Figura 2 – WSN - Wireless Sensor Network



Fonte: Roman e Lopes (2009)

O Gateway é um elemento que encontra-se na borda da WSN, fazendo a ponte entre os nós sensores e a rede de disseminação de dados (Internet), na qual essa possui a ligação junto a camada de serviço.

Protocolos de comunicação de todos os níveis da camada TCP/IP (enlace, rede, transporte e aplicação) devem ser suportados pelo Gateway, visto que a WSN é uma rede heterogênea, contendo variados tipos de nós sensores que transmitem por diferentes protocolos. Os nós sensores podem trabalhar com qualquer protocolo, sendo necessário por parte do fabricante, fazer a melhor escolha, baseando-se na especificação do dispositivo (memória, CPU, armazenamento e tempo de vida da bateria) (DASTJERDI et al., 2016). Na sequência são expostos os principais protocolos de comunicação utilizados nas redes WSN e Internet

No topo da pilha TCP/IP encontra-se a *camada de aplicação*, na qual é composta por protocolos simples e leves; características importantes quando trata-se de redes que trabalham com dispositivos com pouco poder de processamento e recursos (DASTJERDI et al., 2016). Cita-se o CoAP (Constrained Application Protocol), protocolo muito similar ao HTTP, porém projetado para ser mais fácil, demandando menos memória, processamento e consequentemente energia. Sua diferença em relação ao HTTP, está no fato que o primeiro faz uso de UDP na camada de transporte, enquanto o segundo, vale-se do TCP. Ao usar UDP, o CoAP favorece os dispositivos

IoT, pois dispõem de leveza, sem a complexidade de processamento do TCP; contribuindo na redução de consumo energético (DASTJERDI et al., 2016). Por fim, o MQTT (Message Queue Telemetry Transport) é um protocolo M2M, baseado em um modelo publish-subscribe em cima de TCP, na qual cada máquina publica suas informações em um tópico de mensagens. Cada dispositivo que estiver interessado em receber mensagens de outro equipamento, necessita fazer a inscrição no tópico e assim que for postado uma nova, o mesmo recebe a postagem de forma assíncrona. Por estar em cima do protocolo TCP, o MQTT é mais confiável nas entregas das mensagens em relação ao CoAP, sendo indicada para redes não-confiáveis, mas, esta vantagem traz um custo maior de processamento e consumo de energia por parte dos dispositivos IoT (DASTJERDI et al., 2016).

Dentro da *camada de enlace* estão os protocolos que definem os aspectos relacionados a taxa e faixa de transmissão dos dados, banda de frequência, latência de rede, além de detecção e correções de erros de comunicação. Dentre estes protocolos destacam-se o RFID, que segundo Li, Xu e Zhao (2015) são sistemas que podem fornecer informações em tempo real, bastante úteis para os fabricantes, distribuidores e varejistas; alguns exemplos destas aplicações estão no gerenciamento da cadeia de suprimentos melhorando a administração de estoques.

Surgindo como uma versão experimental em meados dos anos 1990, o padrão IEEE 802.11 (Tanenbaum, 2011) é enquadrado como um protocolo para redes locais sem fio WLAN (Wireless Local Area Network). Desde então vem havendo evoluções em termos de segurança, latência de rede e maiores taxas de transmissões, sendo hoje uma tecnologia consagrada, difundida e acessível. Algumas destas evoluções históricas, estão no padrão IEEE 802.11a, na qual se atinge taxas de 54Mbit/s a uma distância de 20 metros. Após a IEEE 802.11b conseguiu chegar a longitudes máximas de 100 metros. No decorrer do tempo surgirão as versões IEEE 802.11e, IEEE 802.11i, IEEE 802.11g, IEEE 802.11n e IEEE 802.11ac, nas quais as mudanças estão nas funcionalidades de segurança, qualidade de serviço e aumento na taxa de transmissão e frequência.

Desenvolvida por uma aliança de empresas de diferentes segmentos do mercado, denominada ZigBee Alliance, o protocolo ZigBee é um padrão global aberto para comunicação sem fio, destinado a hardware que necessita economizar ao máximo o seu consumo energético, devido ao ambiente que encontra-se. Para a realização deste feito, os módulos ZigBee quando não operados, entram em um estado de "sleep" e assim seu consumo declina drasticamente. Opera no padrão IEEE 802.15.4, podendo transmitir 256 kbps em distâncias de 100 metros e suportando até 65.000 nós por rede (DASTJERDI et al., 2016).

Vindo da expressão em inglês "Long Range", o LoRa é uma tecnologia emergente que foi implementada para atender as redes de dados constituídas por nodes com poucos recursos energéticos e que estão em localidades afastadas. Dessa forma, este protocolo dispõe as características de baixo consumo energético e transferência de dados a longas distâncias. Alguns estudos apresentados demonstram transmissões de pacotes na faixa de 15km (Lee; Ke, 2018).

Os protocolos da camada de enlace expostos são os que figuram a serem utilizados na trans-

ferência de dados entre as camadas da arquitetura de referência, visualizada na Figura 1; a escolha do protocolo depende de fatores a serem debatidos pelos administradores da rede IoT, tais como, quantidade de CPU, memória, armazenamento e tempo de vida das baterias dos nós sensores; a localização e ambiente dispostos dos equipamentos e por fim, o valor a ser investido no projeto IoT (DASTJERDI et al., 2016).

A Tabela 1 mostra as características referentes a taxa/faixa de transmissão e suas respectivas bandas de frequências relacionadas aos protocolos da camada de enlace.

Tabela 1 – Características dos Protocolos da camada de enlace

Protocolo	Taxa de Transmissão	Frequência	Faixa de Transmissão
RFID	424 kbps	2.4 GHz	> 1.5 m
IEEE 802.11	50–320 Mbps	2.4-5.8 GHz	100 m
ZigBee	256-20 kbps	2.4 GHz-900 MHz	10 m
LoRa	50 kbps	915 MHz	-+ 15 km
Bluetooth	1 Mbps	2.4 GHz	10 m
UMTS/CDMA	2 Mbps	896 MHz	-
EDGE/MBWA	2 Mbps	896 MHz	-

Fonte: Elaborado pelo autor. (2021)

Em cima da infraestrutura provida pela camada de rede, os dados originados dos nós sensores são recebidos na camada de serviço.

2.1.3 Camada de serviço

Segundo Li, Xu e Zhao (2015) a camada de serviço contempla todas as atividades orientadas a serviços, como informações, intercâmbio, armazenamento, gerenciamento de dados, ontologias de base de dados, motores de busca e comunicação. Para o feito destas tarefas, a camada oferece uma infraestrutura de hardware e software na Internet, a fim de aproveitar as características de escalabilidade, custo acessível, virtualização, segurança, flexibilidade e alta performance que a computação em nuvem dispõe. Empresas como ThingSpeak ¹, Amazon ², Google ³ e IBM ⁴ são algumas que oferecem persistências, ferramentas para construção de modelos de aprendizado de máquina; gerenciamento de nós sensores, comunicação multi-protocolos, segurança em nível de rede e aplicação e fácil escalabilidade (GUBBI et al., 2013).

Tais companhias dividem seus serviços em três categorias: IaaS (Infrastructure as a Service), PaaS (Platform as a Service) e (SaaS — Software as a Service). O modelo IaaS oferece uma infraestrutura referente à capacidade do hardware, que corresponde a virtualização da memória, armazenamento em disco, processamento e rede. O segundo disponibiliza todos os recursos ofertados pelo primeiro, além de ferramentas para o desenvolvimento de software, tais como,

¹<https://thingspeak.com/>

²<https://aws.amazon.com/pt/iot/>

³<https://cloud.google.com/iot-core>

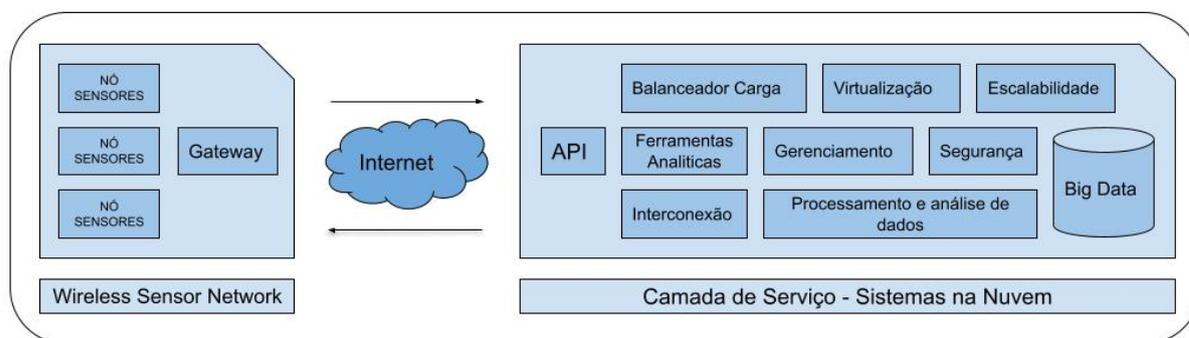
⁴<https://www.ibm.com/cloud/internet-of-things>

sistemas operacionais, compiladores, máquinas virtuais e bibliotecas. O SaaS, contempla as soluções anteriores, somado aos serviços em nível de software, em que representa aplicações com objetivo de atender e resolver a determinado tipo de problema (GUBBI et al., 2013).

Dentro da abordagem SaaS, apresenta-se o conceito Multitenancy, na qual constitui-se uma arquitetura de software que provê uma única instância de aplicação, compartilhada por vários clientes (tenant). Segundo Bezemer et al. (2010) é um princípio de arquitetura de software relativamente novo no modelo de negócios SaaS (Software as a Service). As vantagens em adotar este paradigma de arquitetura são diversas, sendo elas a possibilidade de compartilhamento de recursos de hardware e software, o que traz significativa redução de custos ao modelo. Proporciona fácil escalabilidade, por exemplo, caso um apicultor tenha necessidade de mais processamento ou espaço em disco, a mudança é facilmente resolvida. Provisionamento junto aos clientes de forma automática, necessitando simplesmente informar suas credenciais e por fim, facilidades em configurações e atualizações de softwares.

Por fim, as categorias são agrupadas em dois tipos, públicos e privados; na primeira os recursos são compartilhados entre vários usuários, e na seguinte, de exclusividade do cliente. A Figura 3, mostra uma visão da infraestrutura ofertada pelos serviços na nuvem.

Figura 3 – Camada de Serviço - Serviços na Nuvem



Fonte: Elaborado pelo autor. (2021)

Application Programming Interface é o ponto inicial da camada de serviço, na qual consiste em uma interface de comunicação com o intuito de ser a porta de entrada e saída de dados dos nós sensores. As API's são disponibilizadas pelos provedores de serviços por meio de ferramentas gráficas; em cima de interfaces visuais, o usuário configura o protocolo de aplicação usado em sua rede WSN (HTTP, CoAP ou MQTT) (Li; Xu; Zhao, 2015). Realizada a configuração, os pacotes provenientes das camadas inferiores estão aptos a serem enviados ou recebidos junto à camada de serviço.

Um fato importante dentro da IoT são o grande volume de pacotes geradas por partes dos nós sensores. Estes pacotes podem surgir em intervalos de tempo muito curto, o que provoca a necessidade de uma imensa quantidade de espaço para armazenamento. E por se tratar de uma rede heterogênea, os dados oriundos da WSN não são estruturados, havendo uma combinação de tipos de dados distintos, sem nenhuma padronização. Dentro deste contexto, entra o

conceito de Big Data, na qual consiste em uma ferramenta responsável por armazenar, analisar e interpretar em tempo real, grandes volumes de conhecimento, a partir de dados estruturados (números, texto) e não-estruturados (Figura, mídias sociais e etc.).

Baseado nas informações armazenadas na Big Data, são praticadas rotinas de análise e monitoramento, a fim de extrair elementos de valor ao negócio e juntamente com o gerenciamento destes. As rotinas fazem uso de algoritmos de inteligência artificial que aplicando técnicas e modelos de aprendizado de máquinas, são capazes de prever algum tipo de situação e assim agir de forma reativa ou proativa, sem a interferência humana. Segundo Dastjerdi et al. (2016) as abordagens atuais de análise de dados se concentram principalmente em lidar com Big Data, no entanto, processar dados gerados a partir de milhões de sensores e dispositivos em tempo real é mais desafiador.

Um das características oferecidas por partes das empresas são a fácil configuração da escalabilidade dos serviços na nuvem, na qual traz grandes vantagens para o usuário; a depender de sua necessidade, pode-se aumentar ou diminuir recursos computacionais, tais como, memória, CPU e armazenamento, além de controle de balanceamento de carga, incrementando assim a elasticidade do sistema. Por meio dos recursos oferecidos, a camada seguinte viabiliza interfaces para os usuários contemplarem as noções dos objetos de uma rede IoT.

2.1.4 Camada de interface

A derradeira camada da arquitetura SoA proposta por Roman e Lopes (2009) define-se como um barramento heterogêneo de dados, que disponibiliza os elementos providos pelos nós sensores de forma fácil e transparente. Conforme (Li; Xu; Zhao, 2015) há uma forte necessidade de um mecanismo de interface eficaz para simplificar a gestão e interconexão das coisas.

Pelo fato de haver uma grande quantidade de nós sensores de diferentes fabricantes em uma rede WSN é importante encontrar meios que ofereçam de maneira padronizada o acesso às informações. Uma das soluções para este desafio são as API, um modo simples de disponibilizar tais conhecimentos. Outro recurso são softwares que possibilitam ao usuário visualizar e gerenciar os recursos. Toda essa infraestrutura é oferecida pelas mesmas empresas prestadoras de serviços em nuvem. Por fim, possui níveis de acesso às bases, em que um cliente ao fazer uso de uma API ou um recurso visual, faz necessário a sua autenticação, e, a depender do nível dos privilégios da sua conta, será capaz de acesso a determinados recursos ou não.

Apresentou-se nesta seção a fundamentação teórica relacionada a área de IoT, na qual oferece ao modelo BeeBr, a capacidade de leituras dos dados no apiário. Baseado nestes elementos coletados, possibilita-se a partir do tema exposto na sequência, a elaboração e treinamento do novo modelo de aprendizado de máquina, bem como, demais rotina planejadas pelo BeeBr.

2.2 Aprendizado de Máquina

O segundo campo de pesquisa que serve de pilar na concepção do BeeBr, denomina-se AM (Aprendizado de máquinas) também conhecido como ML (Machine Learning); na qual é umas das técnicas dentro da área da Inteligência Artificial, em que visa aplicar algoritmos computacionais em cima de conjuntos de dados, a fim de reconhecer padrões ou prever resultados de acordo com os parâmetros informados. Implementado através do conhecimento de diversos ramos de estudo, os algoritmos dispõem em suas saídas modelos matemáticos, que possibilitam fazer inferências ou previsões a partir de novos valores. Conforme Mitchell (1997) o aprendizado de máquina baseia-se em ideias de um conjunto diversificado de disciplinas, incluindo inteligência artificial, probabilidade e estatística, complexidade computacional, teoria da informação, psicologia e neurobiologia, teoria do controle e filosofia.

No campo histórico, o AM não é um segmento novo dentro da computação, na qual os primeiros estudos relatados datam por volta dos anos 1950; mas apenas nas últimas décadas começou a popularizar-se, muito devido ao aumento da capacidade de processamento computacional e o grande volume de dados gerados a serem processados e analisados. Baseado nestes fatos, possibilitou a elaboração de modelos matemáticos que contribuem em vastos campos da sociedade, como a previsão do tempo, reconhecimento de imagem, análises precisas dentro do ambiente de negócio, previsões de faturamento e lucros; checagem se um e-mail é spam e dentre outras. As possibilidades dentro do AM são grandes, o que fez a sua procura aumentar na última década, como é visto nesta pesquisa feita no Google, exposta no Apêndice B.

No aprendizado de máquina os algoritmos são alimentados por um conjunto de dados, denominado dentro do contexto de ML de "dataset". Apoiado nestes, os modelos iniciam a execução com o foco em se adaptar, modificar e melhorar a sua experiência a fim de encontrar um padrão e assim conseguir um melhor desempenho. Segundo Mitchell (1997) um programa de computador aprende pela experiência E , em relação à algum tipo de tarefa T e alguma medida de desempenho P se, o seu desempenho P na execução da tarefa T , melhora com a experiência E . Por fim, existem duas categorias no AM, nas quais são nominadas de supervisionado e não-supervisionado; em que as diferenças estão pelo modo de reconhecimento dos padrões encontrados nos conjuntos de dados". Por fim, demais características do AM serão discutidas nas subseções seguintes.

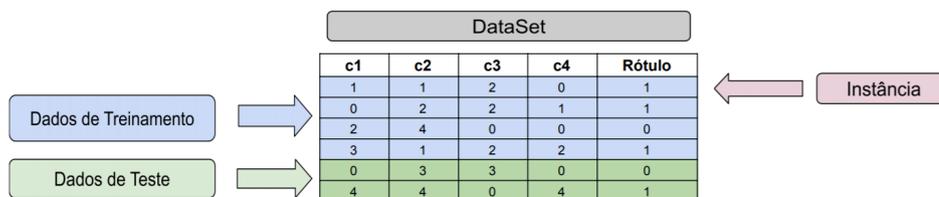
2.2.1 Aprendizado Supervisionado

O AS (aprendizado supervisionado) ocorre quando o algoritmo aprende a partir de valores pré-definidos e seu produto são a previsão de um ou mais resultados, em cima de classes de algoritmos nominados classificação e regressão (DIETTERICH, 2003). Essas rotinas são executadas em cima do dataset, que estão estruturados em forma de tabela, na qual cada linha é referenciada como uma instância a ser processada, sendo ela dividida em duas partes. A pri-

meira são as entradas, também chamadas de características e, por fim, o valor final, nominada de rótulos em que consiste na saída desejada. Conforme Escovedo e Koshiyama (2020) no AS, o modelo (ou algoritmo) é construído baseado no dataset, que são expostos em forma de pares ordenados (entrada — saída desejada).

Entradas do tipo características são variáveis que identificam atributos, aspectos, traços, particularidades e propriedades do contexto em que estamos trabalhando. Em um cenário de verificação a spam de e-mails, pode-se definir o endereço do remetente, hora de envio, país de origem como entradas deste tipo. Os resultados que busca-se prever são definidos como valor rotulado; sendo a saída esperada para o conjunto de entradas. Sendo assim, o nome aprendizado supervisionado vem do fato que os algoritmos dispõem de uma "supervisão" dos valores rotulados, como referencial daquilo que é verdadeiro ou falso. Na Figura 4 é visualizado a estrutura de um dataset.

Figura 4 – DataSet do Aprendizado Supervisionado



Fonte: Elaborado pelo autor. (2021)

Conforme verifica-se na Figura 4, o dataset é organizado em forma de matriz, em que cada linha é uma instância, formada por 5 variáveis, sendo as 4 primeiras do tipo características e a última por rótulo. Os conhecimentos no dataset dividem-se em duas partes, na qual a primeira denomina-se dados de treinamento e a segunda dados de teste ou validação. Esta segmentação se faz necessária a fim de que o algoritmo seja capaz, no primeiro conjunto a compreender, localizar padrões e desta forma, construir o modelo. Na etapa final, o algoritmo recebe a parte restante dos dados, aplicando neste momento o modelo feito na primeira fase, e deste modo, verifica-se o quanto aprendeu com as informações nas quais não a conhece. Na aprendizagem de máquina existem fases que constituem o processo global de aprendizado. Estes estágios são importantes dentro do aprendizado, pois auxiliam na criação e escolha do modelo (Escovedo; Koshiyama, 2020). Na Figura 5 podem ser visualizadas as etapas.

Os dados armazenados no conjunto de dados devem passar por uma etapa de pré-processamento que visa eliminar e corrigir conhecimentos incompletos ou inconsistentes. Tal fase é importante para o processo de aprendizado, pois elimina inconsistências nos resultados finais. Padrões de avaliação eficazes e os critérios são tão importantes para fornecer aos usuários um grau de confiança para os resultados derivados dos algoritmos usados, segundo Xu e Wunsch (2005). Dentre as técnicas existentes nesta fase, encontra-se o tratamento de outlier, em que consiste na busca de valores fora da curva do contexto e por fim a remoção dos dados incoerentes.

No estágio seguinte, escolhe-se o algoritmo para a elaboração do modelo. No aprendizado

Figura 5 – Fluxograma do Aprendizado Supervisionado



Fonte: Elaborado pelo autor. (2021)

supervisionado são definidas duas classes, nominadas classificação e regressão. A primeira busca um ou mais valores dentro de um conjunto de informações em sua saída, e a última, objetiva retornar um único valor. Conforme Escovedo e Koshiyama (2020) quando se deseja fazer a predição de um valor, o problema pode ser de Classificação (quando a variável a ser predita é categórica, geralmente uma classe) ou de Regressão (quando se deseja fazer a predição de um valor numérico, contínuo ou discreto). O detalhamento dos algoritmos são discutidos na seção 2.2.3.

O ponto de partida dos algoritmos de AS são os dados fornecidos pelo dataset, nas quais estão divididos em dois grupos, treino e teste. Ao realizar a segmentação das entradas são geradas as camadas de características e rótulos, na qual essa divisão é conhecida como x e y do treino e x e y do teste, sendo que x representa um vetor com as características e y o rótulo. A expressão seguinte representa esta definição:

$$(x_i, y_i), i = 1, \dots, N$$

na qual N é a cardinalidade finita da coleção de dados representativo disponível no dataset (XU; WUNSCH, 2005). Existem técnicas para a separação dos dados de treinamento e teste. A primeira delas é Treino e Teste simples (Hold-out), na qual divide-se 80% para treinamento e o restante para os testes. Uma das mais conhecidas é a validação cruzada (k-fold Cross-Validation) em que é definido um grupo para testar o modelo durante o estágio de treinamento. Envolve-se a partição do conjunto de dados de treinamento em subconjuntos, em que um subconjunto é mantido para testar o desempenho do modelo.

Ao iniciar a sua execução, o algoritmo é alimentado por x e y do treino, a fim de buscar e construir o modelo. Finalizada, ocorre a predição das classes, inserindo os valores de x e y do teste ao algoritmo, com o objetivo de verificar o quão ele aprendeu. Dentro da predição dos valores de teste, o algoritmo não se vale de y do teste, sendo estes valores utilizados futuramente para medição da capacidade de aprendizado.

Na etapa de avaliação dos resultados do modelo, apura-se a qualidade do aprendizado; o quão acertado foi a generalização das informações das quais o algoritmo não detinha conhecimento. Dentro do ML, existem técnicas, ferramentas e métricas para medir a capacidade do aprendizado, e umas das primeiras a serem utilizadas são a de investigação de ocorrência Over-

fitting e Underfitting. Um modelo sofre de Overfitting quando adapta bem aos dados com que foi treinado, mas que não consegue executar corretamente as inferências ou predições dos dados de teste (DIETTERICH, 2003). Por outro lado, um modelo que não tenha se adaptado de forma positiva tanto com dados de treino quanto de teste, sofre de Underfitting.

Outra maneira de averiguar a qualidade são por meio das métricas de acurácia e matriz de confusão, usadas nas classes de algoritmo de classificação. A acurácia representa o percentual de acertos e a matriz de confusão disponibiliza um detalhamento do desempenho do modelo. Por fim, nas classes de regressão a métrica mais empregada é RMSE (Root Mean Square Error), na qual quanto menor o seu valor, melhor o modelo de regressão analisado (Escovedo; Koshiyama, 2020).

Após as análises dos resultados, pode-se realizar alterações no comportamento do modelo, a fim de obter um melhor ganho. Obtidas as melhorias desejadas, o modelo está pronto, caso contrário, deve-se escolher um novo algoritmo, e assim recomeçar o processo, conforme apresentado no fluxograma da Figura 5.

2.2.2 Aprendizado Não-Supervisionado

O ANS (aprendizado não-supervisionado) tem a característica de não dispor dos dados rotulados no dataset. Desta forma, o algoritmo precisa aprender por si só, sem nenhum supervisionamento para a realização dessa tarefa. Conforme Escovedo e Koshiyama (2020) no ANS, não existe a informação dos rótulos históricos, ou seja, não temos as saídas desejadas a serem estimadas e, por este motivo, dizemos que nossos dados não são rotulados. O objetivo é buscar identificar padrões entre as informações a fim de agrupá-los ou organizá-los em função das similaridades que apresentam entre si. O ANS divide-se em duas classes de algoritmos, sendo a primeira denominada clusterização e/ou agrupamento e por último a associação. As etapas expostas no fluxograma da Figura 5 também são válidas para o ANS, diferenciando-se na fase de 'Execução Algoritmo' dado que neste, nunca há divisão do conjunto de dados de treino e teste, pois não dispõem dos rótulos.

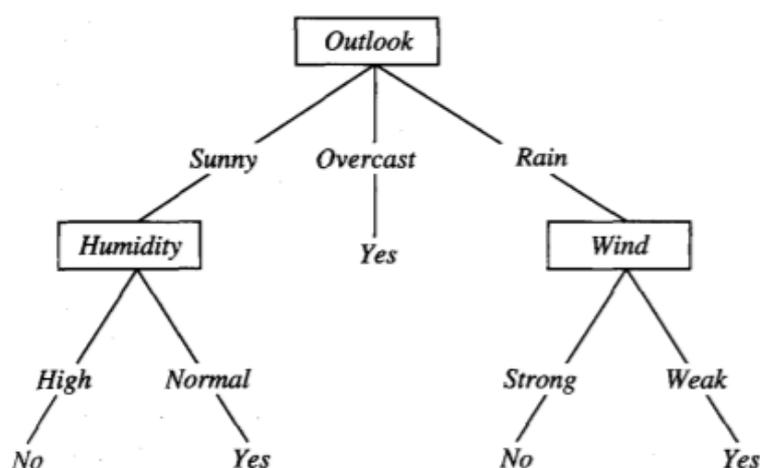
O foco na clusterização são o agrupamento das ideias em grupos, na qual cada elemento que faça parte de um "cluster", compartilhe propriedades e características em comum. Segundo Xu e Wunsch (2005) o objetivo do agrupamento é separar um conjunto de dados em um conjunto finito e discreto de elementos "naturais". Partitivos e hierárquicos são as duas categorias que constituem a clusterização. Pode-se usar o coeficiente de 'silhouette' nas medições de qualidade, na qual calcula-se a distância média de um ponto para todos os outros em um cluster. Por fim, nas classes de associação, volta-se em identificar associações entre registros que de alguma maneira, estão ou devem estar relacionados.

2.2.3 Algoritmos

Nesta seção expõem-se características de um conjunto de algoritmos usados no aprendizado supervisionado *Árvore de Decisão*, *KNN (K-Nearest Neighbours)* e *SVM (Support Vector Machines)* e no aprendizado não-supervisionado como K-means e Apriori.

Segundo Mitchell (1997) o aprendizado da árvore de decisão é um método para aproximar funções alvo de valor discreto, no qual a função aprendida é representada por uma árvore de decisão. As instâncias são classificadas da raiz até algum nó folha, fornecendo a classificação da mesma. Cada nó especifica um teste de um atributo da instância, e cada ramo descendente desse nó corresponde a um dos valores possíveis para a propriedade. As implementações destes algoritmos são realizadas através de recursividade, particionando os conjuntos de dados que os tem e estruturando essas informações conforme uma árvore, favorecendo o entendimento das pessoas.

Figura 6 – Algoritmo Árvore de decisão

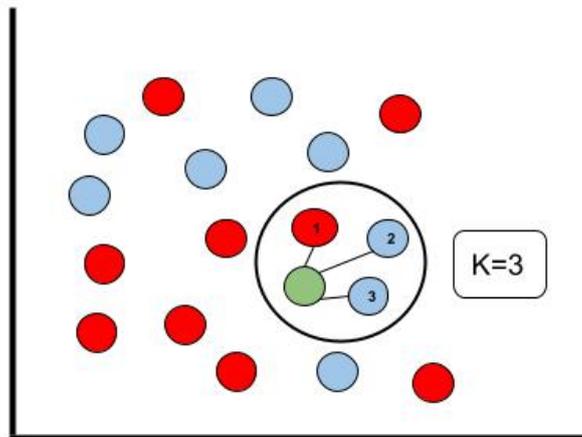


Fonte: Mitchell (1997)

Na Figura 6 ilustra a classificação em um cenário de previsão de tempo, sendo os atributos ensolarado (Sunny), chuvoso (Rain) e nublado (Overcast).

O algoritmo exposto na sequência é conhecido pelo acrônimo em inglês KNN (K-Nearest Neighbors), na qual usa-se em problemas de classificação e regressão. O nome vem da sua intenção de buscar a menor distância entre um dado não classificado em relação a cada um dos outros vizinhos que já estão considerados. O total de vizinhos a ser comparado pelo novo dado, é feito pelo valor de entrada da variável K. Para Escovedo e Koshiyama (2020) no KNN, o conjunto de dados de treinamento são armazenados e, quando um novo exemplo entra, é comparado a todos os demais, a fim de identificar os k (que é um parâmetro de admissão do algoritmo) vizinhos mais próximos (mais semelhantes) de acordo com alguma métrica de espaço (por exemplo, distância euclidiana). Na Figura 7 é mostrado o funcionamento do KNN.

Figura 7 – Algoritmo KNN K-Nearest Neighbors



Fonte: Elaborado pelo autor. (2021)

A partir de um novo valor (bola verde) é definida $K=3$; começa-se a procura por três informações não classificadas ao redor do novo dado, a fim de calcular sua distância. Percebe-se as bolas azuis formadas com menores valores números de afastamento em relação a esfera verde; desta forma, o KNN conclui que novo valor faz parte do grupo das azuis.

O algoritmo SVM (Support Vector Machines) têm como função a determinação de limites de decisão que produzam uma divisão adequada entre classes por meio da minimização dos erros (VAPNIK, 1995). É efetuado um mapeamento não linear para alterar os dados de treino originais em uma dimensão maior, rastreando nesta nova dimensão um hiperplano que separe os dados linearmente de forma ótima. Conforme Escovedo e Koshiyama (2020) o SVM é implementado usando funções kernel, objetos matemáticos que permitem que trabalhem um espaço de dimensão maior. Os tipos de kernel mais utilizados são linear, polinomial e radial. Cabe mencionar que existem outros algoritmos no campo do AS como o Regressão Linear, em que é exclusivo para problemas de regressão e Naïve Bayes para os de classificação.

No ANS, o K-means consta na classe de algoritmos de clusterização. Seu funcionamento baseia-se na distância em que é informado inicialmente o valor do parâmetro K , que define o número de clusters e atribui cada observação a um dos k clusters. O objetivo é encontrar o centroide mais próximo através de uma métrica de espaço e atribuir o ponto de localização a esse cluster. Após esta etapa, os centroides são atualizados pegando-se o valor médio de todos os pontos daquele cluster. Havendo dados numéricos para o cálculo do intervalo, os valores nominais então podem ser mapeados em binários para o mesmo cálculo. Com sucesso na operação, os elementos são separadas e rotuladas e os centroides viram referência para novas classificações. Para Escovedo e Koshiyama (2020) o K-means apresenta bom desempenho quando os grupos são densos, compactos e bem separados uns dos outros, além de ser computacionalmente rápido e de fácil entendimento e implementação. Entretanto, há a necessidade de especificar previamente o parâmetro k , sendo necessário realizar experimentos para determinar

o valor adequado.

Por fim, o algoritmo de mineração de dados Apriori tem como foco extrair um conjunto de regras a partir da conjunção de itens mais frequentes na base de dados, até que critérios de parada sejam atendidos, como tamanho máximo da regra, suporte mínimo da premissa e confiança mínima da regra, segundo Escovedo e Koshiyama (2020). Inicia-se pesquisando os itens frequentes, determinado um limite mínimo para o apoio; após, busca-se todos os conjuntos de elementos que mostram-se com um apoio maior a essa barreira. O passo seguinte é a montagem de regras baseadas nos conjuntos de itens localizados na primeira fase. A confiança de cada regra é calculada, conservando aqueles em que a confiança é maior que o limite imposto pelo usuário.

Exibiu-se nesta seção a teoria referente ao campo de aprendizado de máquina, na qual oferece ao modelo BeeBr, conhecimento necessário para a elaboração e treinamento do novo modelo de aprendizado de máquina. Na sequência expõem-se os fundamentos nas quais são responsáveis por transmitir as informações coletadas no apiário por meio de IoT até a base de processamento da arquitetura BeeBr.

2.3 Comunicação de Dados Móveis

Conforme estudo realizado pela Cisco (CISCO, 2019a) os smartphones ultrapassaram 90% do tráfego de dados móveis até 2022; a previsão que exista 12.3 bilhões de dispositivos móveis conectados em 2022; as velocidades de conexão em rede móvel aumentaram mais de três vezes até 2022 e 79% do tráfego de dados móveis do mundo será de vídeo. A pesquisa traz uma visão do emprego e necessidade que a sociedade têm em relação às redes móveis e conforme Lu e Zheng (2020) o desenvolvimento contínuo da tecnologia da informação e comunicação têm desempenhado um papel vital na atualização contínua do conhecimento e a prosperidade da sociedade.

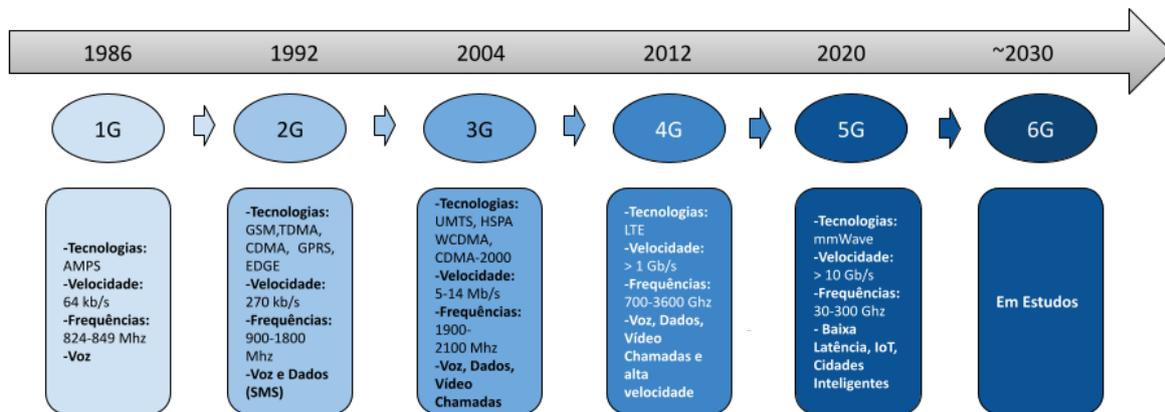
A evolução tecnológica começou em meados dos anos 80 quando surgiu a primeira versão de redes móveis de comunicação, nominada 1ª Geração (1G). Na Figura 8 é exposto o progresso das redes, apresentando o ano de lançamento, velocidades e principais tecnologias utilizadas por cada geração. Nas décadas seguintes são lançadas a segunda, terceira, quarta e quinta, sendo que a sexta está prevista para ser lançada em 2030 (Nawaz et al., 2019).

2.3.1 Passado: 1G, 2G, 3G, 4G

A primeira geração foi um marco mundial para as telecomunicações e trouxe um grande impacto para a sociedade, até então, acostumada a comunicação por telefonia fixa. Esta fase utilizou a tecnologia analógica AMPS (Advanced Mobile Phone System) e trabalhava na faixa de frequência de 800 Mhz e velocidades para transmissões de dados em torno de 60Kb/s.

Em meados dos anos 90 surge a segunda geração, na qual o analógico é substituído pela

Figura 8 – Evolução das Redes de Comunicação Móveis



Fonte: Elaborado pelo autor. (2021)

comunicação digital e acrescido maiores velocidades. Dentre as tecnologias envolvidas estão a TDMA (Time Division Multiple Access), CDMA (Code Division Multiple Access) e GSM (Groupe Special Mobile). O GSM acabou se tornando o padrão de comunicação por ter atraído mais usuários e investimentos. Operava na faixa de 900 a 1800 MHz e conseguia oferecer 270Kb/s em transmissões de dados (Zhang et al., 2020). Outro ponto importante é a inclusão do SIM (Subscriber Identity Module), também conhecido como 'chip', em que se tratava de um dispositivo de segurança, na qual armazenava informações do usuário e sua agenda telefônica.

O envio de mensagens de texto (SMS) foi outra característica incorporada nesta geração (Ferreira; Correia, 2018). A fim de melhorar o GSM desenvolveu-se duas novas tecnologias, sendo a primeira chamada de GPRS (General Packet Radio Services) e após sua extensão conhecida como EDGE (Enhanced Data Rates for GSM Evolution), nas quais trouxeram ganhos substanciais de velocidade para o usuário, tornando o uso da internet mais agradável.

No final dos anos 90, a ITU (International Telecommunications Union) começou a solicitar propostas para o sistema 3G a fim de padronizar a geração. Após esta fase surgiram novos padrões tecnológicos das quais as redes 3G se basearam para a transmissão dos pacotes, sendo elas UMTS (Universal Mobile Telecommunications System) e CDMA-2000 (SHETH et al., 2020).

A 3GPP (Third Generation Partnership Project) desenvolveu a UMTS, concebida para ser a sucessora do GSM. Utiliza W-CDMA (Wideband Code Multiple Access) como interface de rádio e estava baseada em Internet Protocol. Nos anos seguintes implantou-se o HSPA (High Speed Packet Access) como uma evolução do UMTS, na qual trouxe ganhos na velocidade e transmissão de dados, como também aumento da capacidade da rede. Conseguiu-se tais melhorias por meio da união de duas novas tecnologias, denominadas HSDPA (High Speed Downlink Packet Access) e HSUPA (High Speed Uplink Packet Access). Com este sistema possibilitou o uso eficiente de serviços, como web browsing, streaming de vídeo e áudio, email e transferência de ficheiros, conforme Ferreira e Correia (2018).

Implementada pela 3GPP2 (Third Generation Partnership Project 2) o padrão CDMA-2000 compõem-se pelas tecnologias CDMA 1xEV-DO e CDMA 1xEV-DV, na qual suas principais diferenças estão no fato que a primeira transmitia somente dados e a segunda dados e voz.

No ano de 2007 deu início no Japão a quarta geração e teve como principal objetivo atender as seguintes características; comunicação bilateral de voz, vídeo e dados, altas taxas de transferência, baixa latência e uma melhor qualidade nos serviços ao usuário final. Com uma arquitetura de rede mais simples e permitindo a transmissão de dados a velocidades mais elevadas, tendo sido desenhada para comunicações multimídia, segundo Ferreira e Correia (2018).

A geração empregou a tecnologia LTE (Long Term Evolution) em que oferecia taxas de download em torno de 100 Mb/s por canal, explorando uma faixa de frequência de 20 MHz. Conforme Oliveira, Alencar e Lopes (2018) a maior parte do ganho de velocidade e capacidade era obtida por meio de melhorias na eficiência do uso do espectro (como em técnicas de modulação e esquemas de codificação). Um passo seguinte ao LTE foi o desenvolvimento das tecnologias LTE-Advanced e LTE Advanced Pro, na qual introduziu mais velocidades nas transmissões, alcançando valores na casa dos Gigabyte.

2.3.2 Presente: 5G

A introdução ao 5G no ano de 2020 trouxe maiores velocidades e uma série de benefícios para a comunicação móvel, tais como: alta capacidade, diminuição da latência, grande densidade na conexão, melhor rendimento espectral, conectividade perfeita, ampla cobertura e maior eficiência energética da rede (TANG et al., 2021). Em cima destas vantagens espera-se um crescimento em seu uso, como indica o estudo realizado pela Cisco (CISCO, 2019a) que aponta que em 2022, as redes 5G representaram 3,4% das conexões, 11,8% do tráfego total e conexões 5G geraram 2,6 vezes mais tráfego do que a 4G.

Busca-se atingir uma melhor qualidade de serviço em cenários que a anterior não conseguia alcançar de forma satisfatória. A Internet das coisas, realidade aumentada, e-healthcare, comunicação dispositivo e tecnologias emergentes como as cidades inteligentes são algumas que usufruíram destes benefícios (CISCO, 2019a).

Com o aumento do poder de processamento/armazenamento, virtualização de sistemas e automações de processos de negócios, tornou possível as configurações e gerenciamento através de software e não mais por hardware (CISCO, 2019b). A partir da gerência por programa, três novas funcionalidades agregaram a nova rede: (1) NFV (Network Function Virtualization): processo de virtualizar objetos da rede, antes executado por um hardware dedicado/proprietário. Isto reduz os custos operacionais e diminui o tempo para ativação de novos serviços; (2) SDN (Software Defined Networks): camada de abstração que possibilita especificar a arquitetura da rede com uma maior flexibilidade, sem se ater com detalhes, como por exemplo, tecnologias proprietárias; (3) Fatiamento da rede: criação de instâncias de redes na qual são fatiadas de forma virtual a fim de atender diversos nichos de negócios. Assim, a operadora tem a flexibili-

dade através de uma única rede física, repartir e disponibilizar para necessidades e segmentos específicos.

Outra inovação dentro da arquitetura 5G são a C-RAN (Cloud RAN), uma evolução da RAN em redes móveis. Na RAN as estações base são compostas por RRH (Remote Radio Unit) e BBU (Baseband Unit), em que as duas se comunicam entre si, havendo um mapeamento de dispositivos de 1:1. Na nova C-RAN existe a RRH que atende uma região em conjunto com BBU, na qual responde a diversas RRH. Neste moderno cenário o mapeamento de dispositivos é 1:N. As vantagens encontradas em implementar a C-RAN são a diminuição de custos, aumento de desempenho, facilidade de manutenção e compartilhamento de BBU's entre operadoras (Oliveira; Alencar; Lopes, 2018).

2.3.3 Futuro: 6G

O início das atividades comerciais das redes 5G ao redor do mundo no ano de 2020, fez com que a comunidade voltasse seu foco para a sexta geração de redes móveis, e assim, abrir caminho para as pesquisas de novas tecnologias para a evolução da atual (Zong et al., 2019). Neste começo, os trabalhos desenvolvidos por pesquisadores, buscam trazer visões, desafios, tendências, oportunidades e tecnologias para o futuro na nova geração, sendo esse esforço a ser concluído até 2030, definido com o marco de lançamento mundial do 6G (Nawaz et al., 2019).

Busca-se enfatizar as novas tecnologias e aplicativos com altas taxas de dados, confiabilidade elevada, baixa latência e transmissão eficiente e segura (Elmeadawy; Shubair, 2019). As inovações passam por áreas de arquitetura, infraestrutura, meios de comunicação, equipamentos, segmentos como IoT, IA, BlockChain, UAV (Unmanned Aerial Vehicles), segurança da informação e se estendendo a novos negócios.

Um papel importante dentro do 6G são as comunicações TeraHertz "THz", que forneceram mais largura de banda e capacidade, taxas de dados ultra-altas e transmissões seguras. As ondas eletromagnéticas estarão na faixa de 0,1 e 10 THz o que traz a possibilidade de maiores taxas de transferências. A banda THz apoiará o desenvolvimento de minúsculas células em dimensões nanométricas a micrométricas fornecendo comunicações de alta velocidade, segundo o autor (Elmeadawy; Shubair, 2019).

A tecnologia MIMO Massivo, empregada nas antenas da quinta geração, deverá estar presente no 6G. Pesquisas devem continuar no segmento a fim de otimizar os algoritmos empregados atualmente, melhorando questões de eficácia energética, escalabilidade, interoperabilidade, padronização e segurança (Bi, 2019).

Com elevado custo energético dos objetos que compõem o ecossistema da rede 6G, os problemas de eficiência energética serão mais sérios, sendo necessário tomar atitudes a fim de reduzir o consumo (LU; ZHENG, 2020). O aumento da eficácia energética em redes móveis auxilia a reduzir os gastos operacionais e diminuir as emissões de carbono para o ambiente, segundo o autor Nawaz et al. (2019).

As redes de integração espaço terrestre, conhecida como (ISTN), trazem uma série de desafios tecnológicos que precisam ser superados, como suporte a comunicações sem fio verdadeiramente globais em qualquer lugar e hora. Para o autor Huang et al. (2019) espera-se que o ISTN esteja no núcleo dos sistemas 6G, e que seja de alta capacidade e baixo custo, visto que possui um papel de extrema importância no desenvolvimento de sistemas globais 6G.

A comunicação holográfica na qual é possível transmitir todas as informações dos sentidos humanos (olfato, paladar, visão, audição e tato) está prevista para a nova geração. Os aplicativos com comunicação holográfica devem fornecer alta precisão, o que exige um grande volume de informação (SHETH et al., 2020).

Por fim, as redes 6G têm que se valer do Blockchain para o gerenciamento, organização de big data e gestão da grande conectividade gerada pela nova geração. Problemas relacionados a segurança dos dados quando há integrações de sistemas IoT e blockchain são os desafios relacionadas neste segmento (Sekaran et al., 2020).

Desta maneira, apresentou-se neste tópico os fundamentos relacionados aos campos de estudos que são as bases na concepção e desenvolvimento do modelo BeeBr.

3 TRABALHOS RELACIONADOS

Neste capítulo é exposto os trabalhos relacionados que direcionam esta pesquisa, nas quais tratam do assunto principal (*apicultura*) e, contemplam os tópicos aprendizado de máquina, IoT e comunicação de dados móveis; sendo os pilares para atingir os objetivos desta dissertação. O capítulo divide-se nas seções: Metodologia da Pesquisa 3.1, os Resultado da Pesquisa 3.2, as Discussões da Pesquisa, 3.3, a Análise dos trabalhos relacionados 3.4 e as Oportunidades de pesquisa 3.5.

3.1 Metodologia da Pesquisa

A metodologia de pesquisa proposta por Kitchenham (Kitchenham, 2004) empregou-se com objetivo de buscar na literatura existente os trabalhos sobre o tema apicultura, na qual contemplem igualmente os assuntos aprendizado de máquina e IoT, bem como, expor os estudos existentes no campo da comunicação de dados móveis.

Em cima desta orientação, o pesquisador investiga os trabalhos relacionados e a partir de critérios estabelecidos, chega-se no final, com os que mais se caracterizam e relacionam com o objeto da pesquisa; podendo desta forma analisar as lacunas deixadas por estes estudos. As etapas que constituem a proposta são: (1) Definição da Pergunta de Pesquisa; (2) Processo de Busca; (3) Processo de Seleção e (4) Avaliação Qualitativa.

A *definição da pergunta de pesquisa* é a primeira fase para a busca dos trabalhos relacionados, na qual seu propósito é trazer uma indagação específica que o pesquisador almeja responder para abordar o assunto investigado, tornando ela o condutor aos estudos a serem buscados. Sendo o tópico principal a área da apicultura, aventou analisar estudos deste segmento e, que houvesse algum relacionamento com também áreas de interesse de pesquisa, sendo elas, IoT, aprendizado de máquinas, redes de comunicação móveis, bem como sustentabilidade e eficiência energética no que tange as necessidades dos dispositivos de IoT; contemplado essa última, por meio da definição e busca do termo IoT. Desta forma, a pergunta elaborada para este fim, descreve-se no quadro abaixo:

Quais trabalhos existentes sobre a apicultura, estão relacionados com aprendizado de máquina, IoT e comunicação de dados móveis ?

Na etapa de *processo de busca* foram acessadas bases de dados armazenadas na internet, a fim de realizar as pesquisas dos trabalhos relacionados. Os sites usados para a averiguação dos estudos são: ScienceDirect ¹, IEEE Xplore², ACM Digital Library ³, Google Acadêmico ⁴. Para o andamento da pesquisa deve-se definir uma string de busca, que tem a finalidade de

¹<https://www.sciencedirect.com>

²<https://ieeexplore.ieee.org>

³<https://dl.acm.org>

⁴<https://scholar.google.com.br>

trazer resultados de estudos que de alguma forma respondam a pergunta de pesquisa. Definiu-se duas strings de busca, uma em língua portuguesa e outra na inglesa, com objetivo de investigar o máximo de registros dos trabalhos relacionados na literatura mundial.

String em Português:

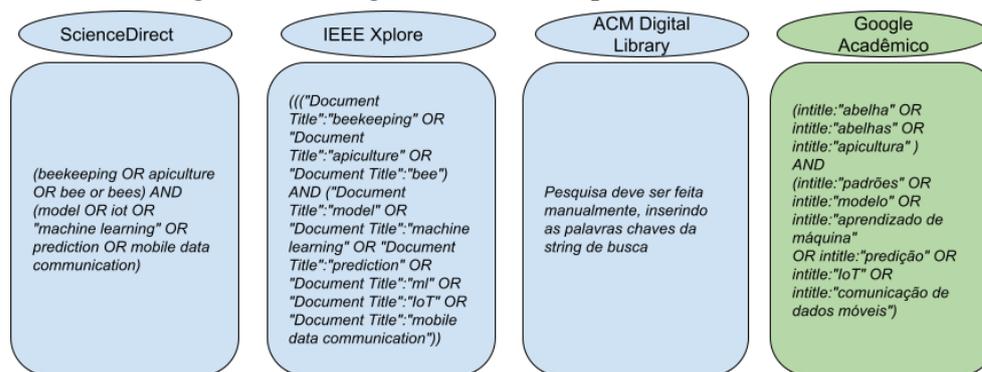
**apicultura OU abelha OU abelhas E (padrões OU modelo OU aprendizado de máquina
OU predição OU IoT OU comunicação de dados móveis)**

String em Inglês:

**beekeeping OR apiculture OR bee OR bees AND (model OR machine learning OR
prediction OR ml OR IoT OR mobile data communication)**

Por se tratar de aplicações distintas e sem nenhuma integração, o método de pesquisas nestas bases de dados são diferentes uma da outra. Desta forma, deve-se adequar a string de busca para cada sistema. Na Figura 9 é visualizada as string para cada site.

Figura 9 – String de busca usada por cada base de dados



Fonte: Elaborado pelo autor. (2021)

Cabe ressaltar que no site (*ACM Digital Library*) deve-se preencher campos para a realização da pesquisa, não sendo possível criar uma string de busca. A procura focou-se em buscar os resultados nos títulos dos estudos. Pesquisas realizadas em língua inglesa destacam-se na cor azul claro (*ScienceDirect*, *IEEE Xplore*, *ACM Digital Library*) na Figura 9, na qual estas bases concentram grande quantidades de pesquisas desenvolvidas mundialmente. Com o objetivo de conhecer o que a comunidade científica Brasileira desenvolveu de estudos sobre o tema apicultura, definiu-se a busca em português no *Google Acadêmico*.

Baseado nos resultados obtidos pelo processo anterior é iniciado a fase *processo de seleção*, em que visa aplicar filtros definidos pelo pesquisador a fim de remover estudos inconsistentes e indesejados. Sendo assim, aplicou-se três critérios de exclusão (CE):

- CE1: Estudos publicados a partir do ano de 2014.
- CE2: Estudos duplicados, em que foram publicados em mais de uma base de dados.
- CE3: Estudos da área de computação ou área correlata com ênfase na apicultura.

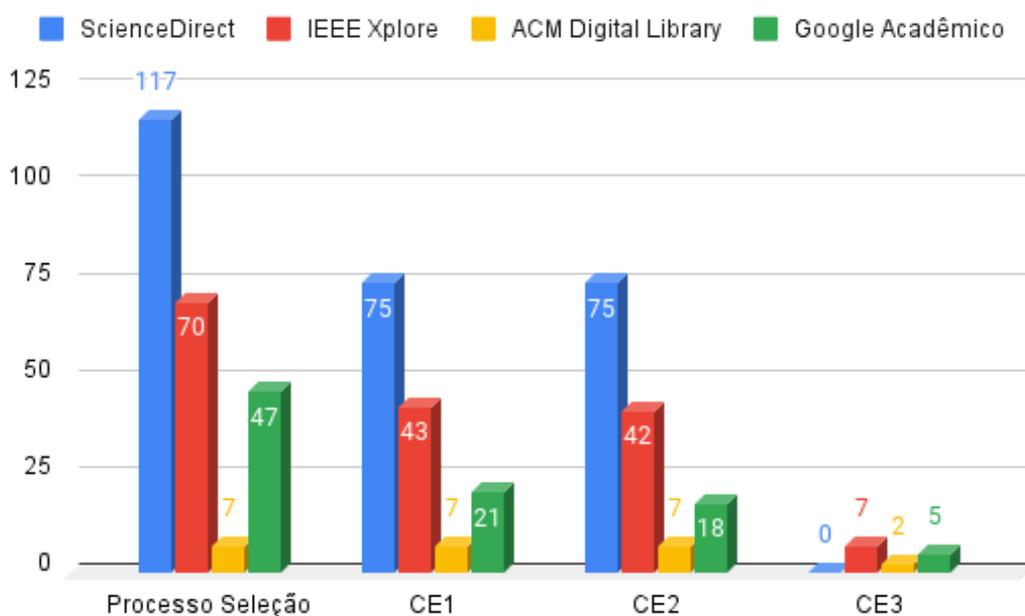
Áreas de objetos desta pesquisa, como IoT e aprendizado de máquinas, são temas recentes e modernos em se tratando das suas aplicabilidades no uso comercial ou científico; sendo assim, adotou-se o critério CE1. Estudos duplicados foram outro filtro de exclusão, definido em CE2. Por se tratar de um segmento discutido nas áreas de Biologia e Agronomia ou mesmo encontrado como nomes de algoritmos da inteligência artificial (*Bee Colony*), fez com que definisse o CE3; na qual objetivasse somente estudos da área de computação ou área correlata com ênfase na apicultura.

Por fim, não foi adotado a última etapa da proposta definida por (Kitchenham, 2004), denominada *avaliação qualitativa*, pois concluiu-se que ao aplicá-la a partir de critérios qualitativos mínimos, o resultado final da pesquisa dos trabalhos relacionados seria muita baixa.

3.2 Resultado da Pesquisa

Nesta seção é apresentado os resultados obtidos pela etapa 3.1, na qual a Figura 10 demonstra os estudos alcançados inicialmente pelas bases de dados e, após aplicado os critérios de exclusão (CE1,CE2,CE3).

Figura 10 – Resultado da Pesquisa



Fonte: Elaborado pelo autor. (2021)

O resultado adquirido pelo processo de seleção, retornou um total de 241 estudos. Após aplicar o CE1, este número caiu para 146. Quatro pesquisas foram excluídas pelo CE2, visto que estavam publicadas em mais de uma base de dados. Percebeu-se ao aplicar o último critério de exclusão (EC3), que muitos destes trabalhos excluídos estavam relacionados diretamente com o algoritmo de IA (*Bee Colony*) citado anteriormente. Já os demais, eram estudos de

pesquisas sobre apicultura ou abelhas e contava com os termos inserido na string de busca (*padrões OU modelo OU aprendizado de máquina OU predição OU IoT OU comunicação de dados móveis*), mas que não havia um relacionamento direto com o objeto desta pesquisa. Como mencionado na seção anterior, não foi adotada a fase de avaliação qualitativa, na qual caso houvesse implementada, haveria terminado em 7 estudos. Desta forma, o resultado final conta com um total de 14 trabalhos relacionados, sendo 7 destes da base de dados IEEE Xplore, 5 do Google Acadêmico e 2 ACM Digital Library. Todas as pesquisas de ScienceDirect foram excluídas pelo CE1 e CE3. Os estudos relacionados são discutidos na seção seguinte.

3.3 Discussão da Pesquisa

Nesta seção são discutidos os 14 trabalhos selecionados, de acordo com os critérios de seleção da seção 3.1. Os próximos dois estudos são do site ACM Digital Library. Os autores Ammar et al. (2019) propõem um projeto denominado *Makers 'Beehives* que visa desenvolver um sistema para a apicultura capaz de avaliar automaticamente a saúde e o status de ameaça de colônias de abelhas por meio da Internet das coisas (IoT) e Inteligência Artificial (IA). Na sua arquitetura são usados Raspberry Pi 3 como microcontrolador e sensores de temperatura, umidade, ruído, iluminação, monóxido de carbono, dióxido de nitrogênio e Raspberry Pi Camera para monitoramento da colmeia. Este sistema é ligado em uma bateria de 5V com capacidade de 33000 mAh na qual é carregada por um painel solar. A cada hora os elementos são coletadas e enviadas para a Internet a fim de aplicar um modelo de aprendizado de máquina.

Em seu artigo os autores Kridi et al. (2014) propõem um algoritmo de ML para colônias de abelhas, capaz de detectar temperaturas adversas na colmeia e alertar para possíveis dificuldades na colmeia. O algoritmo k-means é usado para prever situações contrárias para as abelhas, em cima das leituras dos sensores de temperatura interna e externa. Através de um microprocessador Arduino Uno e placa XBee, os dados são enviados via rádio para serem processados no algoritmo de ML.

Os próximos cinco trabalhos são da base de dados do Google Acadêmico. Um sistema de baixo custo cujo objetivo é realizar o monitoramento da temperatura interna e externa das caixas das abelhas foi desenvolvido no trabalho de Mesquita et al. (2020). Sensores de temperatura DS18B20 e microcontroladora ESP8266 são responsáveis por coletar e enviar as informações para a nuvem, na qual posteriormente são armazenados no banco de dados PostgreSQL. Em cima destes valores é possível a geração de relatórios de acompanhamento da colmeia.

No estudo de Rodrigues et al. (2021) é proposto um modelo de aprendizagem de máquina capaz de detectar anomalias em padrões acústicos de abelhas melíferas africanizadas *Apis mellifera L* e em padrões de temperatura e de umidade sazonais internas em colmeias com abelhas de raça europeia. Três modelos preditivos foram implementados, Modelo de Mistura de Gaussianas (GMM), Máquina de Aprendizado Extremo (ELM) e Máquina de Vetor de Suporte para uma classe (OC-SVM).

O autor Maciel (2018) propôs em sua dissertação a identificação de padrões comportamentais sazonais de colônias de abelhas *Apis mellifera* a fim de auxiliar o apicultor no manejo e na manutenção de suas colmeias. A metodologia consistiu na aplicação da técnica de clusterização em dois datasets reais de colmeias com dados de temperatura, umidade e massa. A partir da aplicação do índice Calinski-Harabasz e do algoritmo K-means, foram identificados padrões coerentes e associados às transições entre as estações do ano. Além disso, foi possível concluir que a colônia mais forte é mais eficiente ao tentar manter o microclima da colmeia durante o inverno.

É apresentado por Herreira e Ferreira (2018) um modelo matemático para estudar o crescimento populacional de uma colônia de abelhas. Para tal, desenvolveu-se um sistema baseado em regras Fuzzy (SBRF) capaz de estimar o parâmetro de crescimento da população de abelhas, o qual foi incorporado ao modelo matemático.

Para verificar o nível de bem estar das colônias de abelhas, a tese de Braga (2020) aplica três métodos de análise de dados que calibram algoritmos de classificação e regressão com base em abordagens de aprendizado de máquina supervisionado e não supervisionado. Para validar o primeiro método aplicado, foi utilizado um conjunto de dados reais de duas colmeias obtidas no portal HiveTool com temperatura da cria, umidade relativa e peso de colmeias de abelhas *Apis mellifera*. A partir do índice Calinski-Harabasz e do algoritmo k-means, foram encontrados 6 padrões sazonais de colônia relacionados às transições entre as estações do ano. Os resultados obtidos com a aplicação dos métodos sugerem que os algoritmos de classificação e regressão são eficientes para a obtenção de modelos de alta precisão para predição de níveis de bem estar das colônias de abelhas *Apis mellifera*.

Os sete estudos finais são da base de dados IEEE Xplore. O autor Mahamud et al. (2019) propôs um sistema de baixo custo para monitoramento das colmeias, nas quais são possíveis a visualização dos elementos de temperatura, umidade e ruído. A arquitetura proposta baseou-se no microcontrolador ESP8266 e sensores de captação das variáveis do ambiente. Através do ESP8266 que executa o papel de gateway da rede WSN, os dados são recebidos e publicados no servidor baseado no protocolo MQTT que posteriormente disponibiliza as informações para ser consumida por uma aplicação mobile.

Um protótipo de sistema autônomo, aplicando IoT é proposto por Zabasta et al. (2019). O objetivo do protótipo é fornecer dados sobre o estado da colmeia e do apiário (interno e ambiental como temperatura, umidade, peso das colmeias e visual) para os usuários, de modo que eles possam avaliar o status da colmeia e tomar providências. O sistema é composto por microcontrolador ESP8266 e diversos sensores para o monitoramento da colmeia. Uma câmera de vídeo é usada para a monitoração do apiário. Os autores desenvolveram uma rede WSN na qual a comunicação se dá via rádio a partir do ESP8266 até o gateway, que em seguida, encaminha para a nuvem os dados através de conexão 4G. Em cima destes conhecimentos o protótipo disponibiliza ferramentas para acompanhar em tempo real o estado do apiário.

O objetivo do trabalho de Arruda et al. (2018) é classificar as espécies de abelhas através do

emprego de algoritmos de aprendizagem. Os dados são coletados com etiquetas RFID coladas ao tórax das abelhas. Sempre que uma abelha entra ou sai da colmeia, o carimbo de data e hora é armazenado. Os dados foram analisados em formato de série temporal, na qual as atividades das abelhas são agrupadas em diferentes categorias. Três métodos de aprendizado de máquina foram escolhidos para classificar as espécies das abelhas. Os modelos usados são RF (Random Forest), MLP (Perceptron multicamadas) e SVM (Support Vector Machines).

Um sistema capaz de prever e mapear o comportamento de polinização de insetos voadores, e que permite aos produtores de frutas detectar, quantificar e otimizar a atividade das abelhas, foi proposto por Van Goethem et al. (2019). A partir de modelos SVM (Support Vector Machines), o sistema consegue identificar e inspecionar as espécies de insetos nos campos.

No trabalho de Khairul Anuar et al. (2019) é proposto um sistema IoT que monitora o peso, as temperaturas interna/externa e a umidade relativa da colmeia. A configuração do hardware são o microcontrolador NodeMCU e os sensores de leitura. Os dados são coletados e armazenados na nuvem, na qual é usado o serviço da plataforma Google.

A depender do local que um equipamento gateway de rede WSN está localizado, pode haver problemas para alimentá-lo energeticamente, devido ao fato de estar distante de fontes de alimentação. A partir desta situação os autores Vidrascu e Svasta (2017a) propõem um projeto livre de manutenção, energeticamente independente usando supercapacitores como energia de armazenamento. A apicultura foi usada para validar a proposta, na qual são instalados sensores em caixa de abelhas e, as informações coletadas, transmitidas para o gateway alimentado pelos supercapacitores.

Por fim, os autores Vidrascu e Svasta (2017b) destacam em apresentar os componentes de hardware que fazem parte da área de IoT. Em cima do segmento da apicultura, os pesquisadores montaram uma configuração na qual utilizava o microcontrolador ESP8266 e alguns sensores a fim de demonstrar o uso destes dispositivos dentro de uma colmeia.

3.4 Análise dos trabalhos relacionados

Nesta etapa são analisadas as lacunas deixadas pelos estudos da seção 3.3 em relação ao objeto de pesquisa. Os estudos examinados tratam sobre o tema principal apicultura, na qual contemplem de alguma maneira os assuntos aprendizado de máquinas, IoT e comunicação de dados móveis, nas quais são de interesse deste pesquisador. A partir destas definições é apresentado uma síntese de temas abordados por cada trabalho na Tabela 2. Os acrônimos significam: $(A + IoT)$ = Apicultura + IoT; $(A + AM)$ = Apicultura + Aprendizado de Máquina; $(A + CDM)$ = Apicultura + Comunicação de Dados Móveis.

Dos 14 trabalhos relacionados, 11 (78,50%) tratam de $(A + IoT)$, 7 (50,00%) de $(A + AM)$ e 1 (7,14%) de $(A + CDM)$, bem como, estudos com a interseção de todos os temas interessados somam-se 0 (0,00%).

Percebe-se que os estudos que tratam do assunto $(A + IoT)$ se restringem ao uso de deter-

Tabela 2 – Comparação dos trabalhos relacionados

	(A + IoT)	(A + AM)	(A + CDM)
Ammar et al. (2019)	x	x	-
Kridi et al. (2014)	x	x	-
Mesquita et al. (2020)	x	-	-
Rodrigues et al. (2021)	-	x	-
Maciel (2018)	x	x	-
Herreira e Ferreira (2018)	-	-	-
Braga (2020)	-	x	-
Mahamud et al. (2019)	x	-	-
Zabasta et al. (2019)	x	-	x
Arruda et al. (2018)	x	x	-
Van Goethem et al. (2019)	x	x	-
Khairul Anuar et al. (2019)	x	-	-
Vidrascu e Svasta (2017b)	x	-	-
Vidrascu e Svasta (2017a)	x	-	-

Fonte: Elaborado pelo autor. (2021)

minado conjunto de sensores para a leitura das variáveis do apiário, tais como, temperatura, umidade, peso, ruído e gás, sendo que somente o autor Zabasta et al. (2019), acrescentou o emprego do vídeo. Os trabalhos de Vidrascu e Svasta (2017b) e Vidrascu e Svasta (2017a) que contemplam o tema (*A + IoT*), enfatizam em expor questões relacionadas ao armazenamento energético e dispositivos IoT, sendo a apicultura apenas uma área de experimentos destas pesquisas.

Observa-se nos trabalhos de Mahamud et al. (2019), Zabasta et al. (2019), Khairul Anuar et al. (2019) e Mesquita et al. (2020) que os elementos levantados pelos sensores instalados no apiário, servem de subsídios para o desenvolvimento de ferramentas no monitoramento da apicultura. Estes sistemas trazem interfaces gráficas para o apicultor verificar o estado das abelhas, mas não se verificou uma análise profunda em cima dos dados, de tal modo pudesse trazer predições de eventos e problemas na colmeia.

Dentro do campo energético e sustentável, no que tange as necessidades dos dispositivos de IoT, apenas o estudo Ammar et al. (2019) tratou do ponto de sustentabilidade, na qual o autor adicionou um painel solar em seu modelo, mas, não obteve-se resultados em relação a eficiência energética.

Esperou-se nas pesquisas de abordagem do tema (*A + AM*) que trouxessem de alguma forma, predições nas quais oferecesse o progresso de uma determinada etapa na apicultura, contemplando neste sentido preempções de problemas na colmeia. No estudo de Ammar et al. (2019), é proposto um modelo de aprendizado de máquina, predizendo a saúde das colônias, mas conforme os resultados e conclusão avaliados, percebeu-se que o trabalho não atingiu com o objeto da pesquisa.

Convém mencionar os estudos de Rodrigues et al. (2021), Maciel (2018) e Braga (2020) em que trouxeram resultados nas quais limitam-se as previsões de detecções de temperaturas,

irregularidades em padrões acústicos e reconhecimento sazonal de colônias. Salienta-se que seus dataset aplicados com a finalidade de alimentar os algoritmos de AM, foram baixados por plataformas oriundas da Internet, que oferecem conjuntos de dados para análise, sendo os sites Hobos ⁵ e HiveTool ⁶, a partir de um conjunto limitado de características (temperatura, umidade e som). Segundo o autor Maciel (2018), se fez necessário esta abordagem, devido às localizações remotas dos apiários, o que ocasionou dificuldades técnicas relacionadas às transmissões dos dados, visto que não havia nenhuma conectividade de Internet e todas as outras soluções também falharam. Diante deste fato, reforça-se o quão é difícil a disponibilidade de Internet no interior do Brasil, na qual dificulta os avanços e progressos em diferentes áreas.

Diante do exposto, não obteve-se pesquisas e/ou resultados satisfatórios em estudos com a temática AM na preempção em específico nos problemas nas colmeias, tais como a enxameação (fuga de abelhas) e colmeia zanganeira, que são algumas das dificuldades que o apicultor precisa lidar, sendo que tais problemas atrapalham a saúde e bem estar das abelhas. Os estudos, conforme exibido abordaram em seus modelos questões de detecção de temperaturas adversas na colmeia, anomalias em padrões acústicos, reconhecimento de padrões sazonais de colônia e classificação de insetos por meio de RFID.

Por fim, obteve-se como resultado na temática (*A + CDM*) o estudo Zabasta et al. (2019) em que trouxe uma arquitetura voltada a apicultura na qual possibilitava-se a transmissão das informações coletadas no apiário para a Internet por meio de comunicação de dados móveis, baseado na rede 4G (LTE). Demais pesquisas apontavam conectividade a internet, mas com tecnologias diversas; o estudo de Kridi et al. (2014) transmitiu a rádio até sua base de processamento; o trabalho de Mesquita et al. (2020) não foi identificado a conexão usada no envio do apiário até o gateway, mas segundo o autor, a transmissão dos dados para a nuvem é feita por ADSL. Demais estudos não apontavam meio, forma de conectividade junto a rede mundial de computadores.

Diante das dificuldades de encontrar e não dispor dos estudos relacionados (*A + CDM*) referencia-se o trabalho de Yaacoub e Alouini (2020), na qual fora pesquisado por este autor, no ano de 2020/2, a partir de uma revisão sistemática de literatura sobre redes 6G. Dentre 112 estudos revisados, este foi o único a tratar e apresentar uma visão da comunicação 6G diante à agricultura. Mesmo sendo atividades distintas, a apicultura e agricultura relacionam-se por serem cultivadas e desenvolvidas no campo.

3.5 Oportunidades de pesquisa

Nesta seção é exposto as oportunidades de pesquisa, mediante as lacunas deixadas pelos trabalhos relacionados. Diante da inexistência de trabalhos que envolvem a predição de problemas na colmeia, em específico as dificuldades de enxameação e colmeia zanganeira, pode-se

⁵<http://hobos.de>

⁶hivetool.net

destacar nesta direção, a fim de trazer noções que auxiliem o trabalho do apicultor e de alguma forma, venha diminuir o manejo. Em face deste cenário, oportuniza-se a elaboração e treinamento de um moderno modelo de AM, que objetiva a preempção no tocante às adversidades relacionadas a enxameação e colmeia zanganeira.

Percebe-se que os sensores instalados nos apiários, restringem-se à temperatura, umidade, gaz, peso e ruído. Desta forma, existe a oportunidade de inserir novos sensores IoT, que venham dispor de atuais valores captados no ambiente, bem como, à dados coletados por inspeções realizadas diretamente no apiário por partes dos apicultores, a fim de constituir os atributos para a elaboração de um novo modelo de aprendizado de máquina.

Dentro ainda da área de captação das informações do apiário, apresenta-se uma situação de estudo na qual objetiva a expor soluções energeticamente sustentável e eficiente, no que refere-se as necessidades energéticas dos dispositivos de IoT.

Tendo em vista a escassa existência de estudos que reportam a comunicação de dados móveis diante a apicultura, observa-se uma grande chance de expor novas tecnologias para a transmissão de redes móveis. Neste sentido, oportuniza-se pesquisas relacionadas respectivamente a recente e futura redes 5/6G.

Diante dos trabalhos apresentados, percebe-se que no Brasil não existe um projeto, solução ou ferramenta que ostente características de colaboração, facilitação e evolução ao trabalho dos apicultores, em conjunto ao mundo tecnológico. Projetos semelhantes com tais aspectos, podem ser encontrados em desenvolvimento em Israel, pela startup Beewise ⁷. O sistema engloba avanços tecnológicos na qual a empresa disponibiliza a instalação, suporte e software com um custo mensal de \$400.00, ou pela empresa SolutionBee ⁸ em que uma solução para uma caixa de abelha, gira em torno de \$1,099. Apesar de ser uma atividade lucrativa, grande parte do mel produzido no país vem de pequenos apicultores sem capacitação técnica e dedicação plena aos apiários. Desta forma, os lucros que um apicultor brasileiro dispõe, torna inviável a manutenção do pagamento de sistemas informatizados de alto custo.

Portanto, diante do exposto, este trabalho tem a oportunidade inicial de expor uma arquitetura para o segmento da apicultura denominada BeeBr, na qual oferta um novo modelo de AM, constituído por novos atributos até o momento não empregados, que resulta a buscar respostas no que refere-se à problemas na caixa, em particular a enxameação e colmeia zanganeira, bem como, expor soluções no tocante a sustentabilidade energética por parte dos dispositivos IoT; e em etapas futuras, seguir avançando, a fim de transformar este estudo em um projeto de impacto social, na qual acrescente, colabore e contribua no trabalho dos apicultores brasileiros.

⁷<https://www.beewise.ag/>

⁸<https://solutionbee.com/store/>

4 MODELO PROPOSTO

Este capítulo objetiva expor uma visão sistemática do modelo computacional proposto por esta dissertação, denominada de BeeBr. Sua concepção originou-se inicialmente a partir de conversas com os apicultores responsáveis pelos apiários. As discussões objetivaram entender melhor a área da apicultura, dispor conhecimento das etapas de colheita dos produtos; compreender as peças que moldam todo o ecossistema do segmento; dominar os jargões usados pelos apicultores e descobrir as dificuldades encontradas. Com base nestes debates, conseguiu-se apossar de uma base de conhecimento. Baseado nas compreensões citadas e somadas as lacunas apontadas pelos trabalhos relacionados em capítulo 3, definiu-se os pontos na formação do modelo BeeBr.

A primeira seção 4.1 descreve uma visão geral e premissas básicas para o sua elaboração. Nas seções seguintes são reveladas as três camadas que compõem o modelo: (1) Camada de Detecção 4.2; (2) Comunicação 4.3; (3) Serviço 4.4 e por fim, é exibido a elaboração e treinamento do novo modelo de aprendizado de máquina com foco na predição em enxameação.

4.1 Visão geral do modelo

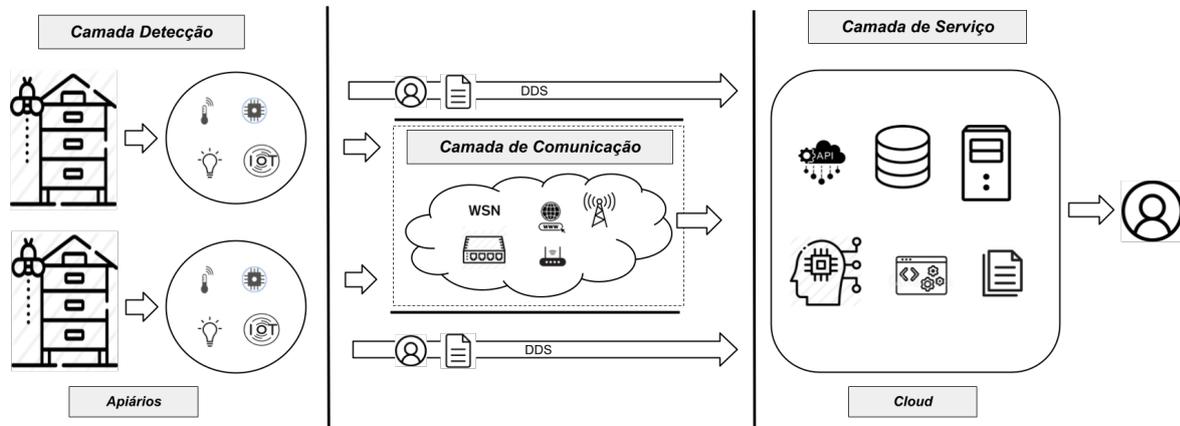
Buscou-se na modelagem do BeeBr, aspectos que buscassem a diminuição do manejo; predição de possíveis problemas nas colmeias e melhoras no bem estar das abelhas, a partir do emprego tecnológico na apicultura. O BeeBr constitui-se por três camadas (*Detecção, Comunicação e Serviço*), que destinam-se a realçar e atuar em determinados estágios dentro deste modelo, englobando campos de estudos de interesse desta pesquisa, dentre elas, *IoT, comunicação de dados móveis e aprendizado de máquinas*.

O BeeBr têm como foco levar ao apicultor uma solução tecnológica de custo baixo, na qual traga benefícios ao apicultor, em específico na etapa do manejo. Além de vantagens para às abelhas, dado que, evitando-se as inspeções e dispondo de sistemas inteligentes que sejam capazes de monitorar e prever situações desagradáveis aos insetos, tende a trazer impactos positivos na saúde e bem estar destes animais. Na concepção do BeeBr, pensou-se em uma solução completa, na qual ofereça a chance de implementar algoritmos de aprendizado de máquina, facilidade e flexibilidade na transferência de dados à rede, hardware e software que colem e processem as informações dos apiários, bem como, alternativas no desenvolvimento sustentável.

Tendo em vista a complexidade que propõem-se nesta arquitetura, pretende-se segmentá-la a partir da construção de camadas, de modo que, facilite o entendimento, planejamento, desenvolvimento, execução e, principalmente, tornar o projeto flexível. Desta forma, diminuiu-se o acoplamento, possibilitando plugá-los de forma transparente e, favorecendo a evolução das futuras fases do modelo. Mediante a interfaces coesas, possibilita-se um canal de entrada e saída de dados, que propicie as integrações entre os segmentos. Na Figura 11 apresenta uma

visão geral do BeeBr.

Figura 11 – Visão Geral do Modelo BeeBr



Fonte: Elaborado pelo autor. (2021)

Observa-se na Figura 11 as três camadas que compõem o modelo. A camada de detecção, na qual localiza-se nos apiários, responsabiliza-se em ler, armazenar e/ou transmitir os elementos da colmeia, por meio da internet das coisas. Com base nas variáveis coletadas, surge a comunicação, transmitindo os pacotes com destino ao serviço, através de redes de dados (WSN/Internet). Situada na cloud, o último nível objetiva-se a processar, persistir e visualizar as informações ao usuário. Este é o fluxo normal de conversação entre as camadas são: (1) Detecção -> (2) Comunicação -> (3) Serviço; mas o BeeBr expõem uma solução que objetiva a depender da necessidade do apicultor, encaminhar os dados coletados pela primeira camada até à última por intermédio de arquivo, sem a intervenção da camada do meio, na qual é chamada de DDS (Detecção-Direto-Serviço).

Nas bordas do modelo estão as colmeias e os apicultores, sendo a primeira o agente que dispõem as bases, para o início do fluxo do processo. Na extremidade oposta, encontram-se os apicultores, que com base nos dados coletados, transmitidos e processados pelas camadas, usufrui um ambiente de aplicações gráficas, possibilitando-os a visualizar e gerenciar os recursos.

Por fim, como objetivo exposto nesta dissertação, apresenta-se dentro desta visão geral do BeeBr, uma alternativa na elaboração e treinamento de um novo modelo de aprendizado de máquina, na qual objetiva na predição da enxameação (fuga das abelhas), a fim de diminuição das intervenções junto ao apiário (manejo), garantindo assim o bem estar das abelhas. Na sequência são expostos características específicas na qual o BeeBr, necessita apresentar.

Primeiramente, o BeeBr deve gozar de recursos computacionais, em que viabilize ofertar e fazer uso da área de aprendizado de máquina. A partir deste conceito, entende-se que o BeeBr disponha de componentes de softwares que possibilite-o, a inclusão, gerenciamento e execução de modelos de AM, ou seja, na concepção desta entidade, compreende-se que sua estrutura deve ser flexível no ponto de vista da inclusão de novos algoritmos de AM; como exemplo, a opção proposta do moderno modelo de aprendizado de máquina focado na predição, citado

anteriormente.

No olhar da camada de detecção, considera-se a condição de ser sustentável e eficiente energeticamente, pontos importantes na concepção deste modelo, visto que, as localidades onde encontram-se os apiários são remotas, sem nenhuma infraestrutura, por questões de segurança, e devem observar a distância mínima de 400m em relação a casas e outros locais frequentados por pessoas (ABELHAS, 2021b).

Deve-se possibilitar ao apicultor solicitar os dados originados pela camada de detecção e encaminhar diretamente para a de serviço, de forma manual, sem a interferência com a de comunicação, na qual este recurso conforme citado, denomina-se por DDS. A partir deste artifício, propicia aos usuários que estão em regiões do Brasil, sem conectividade a redes móveis, fazer uso do modelo BeeBr.

Conforme Ahmad Saeed et al. (2014) o apicultor possui baixo nível de profissionalização, existe dificuldade de acesso a tecnologias e assistência técnica. Entende-se que o perfil deste usuário é de pouco conhecimento tecnológico; sendo assim, o BeeBr deve disponibilizar sistemas fáceis de serem manipulados e gerenciados, interfaces simples, noções com jargão da área e descomplicada instalação no apiário.

Por fim, o modelo BeeBr precisa dispor de uma infra-estrutura que ofereça os recursos necessários para seu trabalho, mas com um baixo custo. Constitui-se por elementos de hardware, software, infraestrutura e mão de obra. Dentre os equipamentos, cita-se sensores, microcontroladores, baterias e módulos de comunicação. Na parte da infra, é necessário a contratação de empresas prestadoras de serviços na nuvem, que dispõe de recursos para processar e armazenar as informações. A mão de obra humana é indispensável em uma arquitetura deste porte, devido às etapas de implementação das funcionalidades de aplicações.

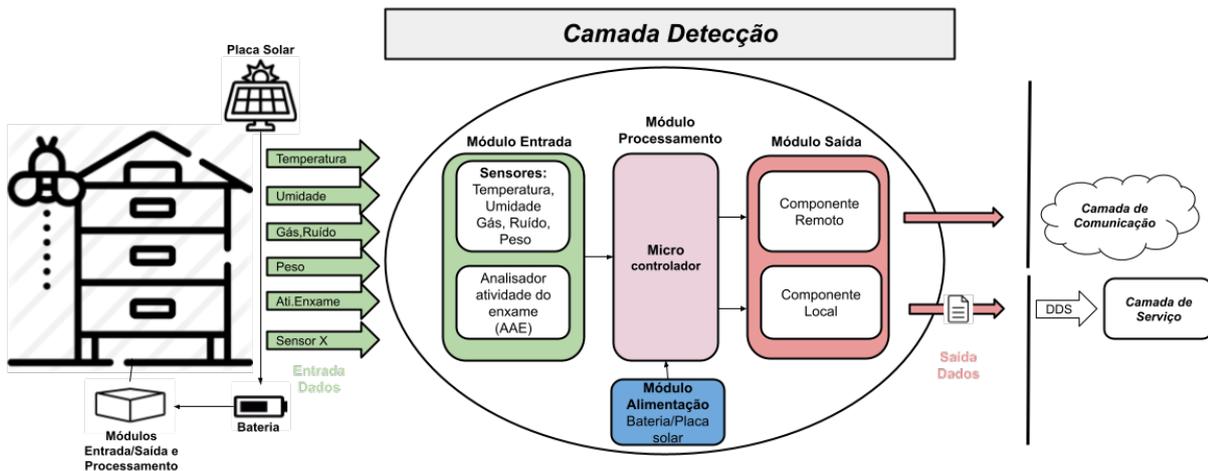
Desta forma, apresentou-se nesta seção uma visão geral do comportamento e funcionalidades do modelo BeeBr; nas subseções seguintes serão tratadas os detalhes do funcionamento das camadas reportadas, bem como, a especificação na elaboração e treinamento do novo modelo de aprendizado de máquina.

4.2 Camada de Detecção

A camada de detecção encontra-se os dispositivos que devem dispor da capacidade de coletar, processar, salvar e/ou transmitir os pacotes da colmeia, implementando conceitos apresentados e discutidos na seção 2.1.1. Este conjunto de equipamentos são formado por sensores, microcontrolador, módulos de transmissão/armazenamento dos dados, bateria e painel solar. Constitui-se a área de interesse de estudos sobre IoT dentro do modelo BeeBr, na qual é composto por elementos da internet das coisas. Na Figura 12 é exposto em detalhes a camada.

Os dispositivos estão agrupados em quatro módulos, nas quais dispõem da responsabilidade de executar determinada tarefa dentro da camada. Segue uma breve descrição de suas atribuições:

Figura 12 – Camada de Detecção



Fonte: Elaborado pelo autor. (2021)

- Módulo Entrada: Ler/Receber os elementos da colmeia.
- Módulo Processamento: Processar as informações dos módulos de entrada/saída e gerenciar dispositivos.
- Módulo Saída: Encaminhar os pacotes coletados aos demais segmentos.
- Módulo Alimentação: Fonte energética para os hardwares.

4.2.1 Módulo de Entrada

O módulo de entrada contempla sensores que monitoram o ambiente da colmeia. Cada sensor realiza a leitura de uma determinada informação e em seguida encaminha ao microcontrolador. O BeeBr propõe a inclusão de seis sensores para o acompanhamento destes dados.

Segundo Lorenzon (2004) o desenvolvimento e o comportamento de colônias de abelhas recebem influência de fatores ambientais como temperatura, umidade relativa do ar e radiação solar. A fim de investigar as reações das abelhas, devido às influências do clima, sensores de temperatura e umidade compõem o módulo de entrada.

Dentro do manejo, o apicultor precisa estar atento a possíveis problemas na colmeia, garantindo o bem estar das abelhas. A enxameação é um processo natural de reprodução do enxame, onde grande parte das operárias e a rainha velha, ou uma rainha nova, saem do ninho para formar outra colônia e as colmeias zanganeiras caracterizam pela grande presença de zangões na colmeia, devido a falta da abelha rainha (TECNICO, 2007). Estas duas adversidades da natureza trazem para a apicultura grandes dificuldades e impactam diretamente na produção de mel. Uma das características destas situações na colmeia é a presença de ruídos maiores do que o normal. Em colmeias zanganeiras se tem um zumbido forte e no processo final de enxameação, a rainha emite um som especial. Para o autor Bencsik et al. (2011) o monitoramento de áudio

pode ser aplicado como estimativa no comportamento das abelhas e utilizados para identificar padrões de zumbido do enxame que indique um possível abandono em massa da colônia. Por este motivo, o modelo BeeBr traz um sensor de ruído, com o intuito de investigar estes aspectos sonoros.

Um sensor de gás foi inserido no modelo a fim de verificar a qualidade do ar ao redor da colmeia. Através deste detector é possível analisar amônia, óxido nítrico, álcool, benzeno, dióxido de carbono e fumaça. Todos estes prejudiciais para a saúde e bem estar das abelhas.

Um apicultor que segue boas práticas de manejo e com boa localização do apiário, próxima a flora, consegue colher mel o ano inteiro. Mas existem meses em que a safra é maior que outras por questões do ciclo de produção da abelha. Desta forma, saber a eficiência do apiário é uma informação valiosa na apicultura. Para o autor Fitzgerald et al. (2015) peso da colmeia é um indicador de saúde e produtividade da colônia. Por meio deste valor possibilita-se diagnosticar problemas em colmeias que estejam produzindo menos que o esperado. Deste modo, um sensor de peso é uma peça importante para a construção do BeeBr.

Com o intuito de verificar o fluxo de entrada e saída das abelhas, pensou-se em um primeiro momento, o desenvolvimento de um recurso que pudesse realizar esta tarefa. Mas por questões de custo e tempo, propõem-se inicialmente uma solução formada por software e hardware, na qual objetiva efetuar tal trabalho, sendo neste estudo, denominado o conjunto citado de AAE (Analisador Atividade Enxame). Por intermédio de um sensor de obstáculo infravermelho, pretende-se ter uma amostra do quão as abelhas estão se movimentando na entrada da colmeia (alvado). Mediante esta informação, pretende-se averiguar situações de abandono em massa da colônia, bem como, trazer subsídios para os algoritmos de aprendizado de máquina, a fim de prever esses acontecimentos (enxameação).

Na Tabela 3 é apresentado uma síntese dos itens e seu propósito no modelo BeeBr.

Tabela 3 – Composição Módulo de Entrada

Item	Propósito
Sensor Temperatura	Investigar as reações das abelhas, devido a influência da temperatura.
Sensor de Umidade	Examinar as ações das abelhas, por causa da atuação da umidade.
Sensor de Gás	Verificar a qualidade do ar, a fim de garantir a saúde das abelhas.
Sensor Ruído	Observar o zumbido do enxame, o que pode apontar abandono da colmeia.
Sensor Peso	Analisar a saúde e produtividade da colônia.
Sensor infravermelho (AAE)	Averiguar situações de abandono em massa da colônia. (Enxameação)

Fonte: Elaborado pelo autor. (2021)

Por fim, cabe mencionar que o módulo de entrada permite adicionar outros novos sensores, havendo apenas a necessidade de implementação em nível de software.

4.2.2 Módulo de Processamento

Sendo o núcleo central da camada de detecção, este módulo tem como foco, processar todas as informações originadas pelos módulos de entrada/saída e gerenciar os hardware nele conectado. Através de um software desenvolvido dentro do escopo deste estudo, é realizado o processamento em que contempla a recepção aos dados oriundos dos sensores e transferência destes aos dispositivos do módulo de saída. Com a finalidade de realizar tais atividades torna-se necessário um microprocessador, na qual, possibilite programá-lo, a fim de implementar algoritmos que realizem as tarefas citadas e, conectável a qualquer novo periférico, de forma que expanda suas funcionalidades. Sensores, placas de transmissão e armazenamento dos dados são exemplos destes equipamentos.

Tendo em vista que o fator energético é um ponto extremamente relevante para o sucesso do modelo BeeBr, fez com que a escolha do microprocessador se torne uma etapa importante, dado que sua principal característica a ser observada é seu baixo consumo de energia e, em seguida, programável e modulável. O fornecimento energético ao microprocessador por parte do módulo de alimentação, assegura energia aos demais dispositivos conectados.

Propõe-se o emprego de microprocessador de baixo poder de processamento, tendo em vista que as atividades executadas são de pequeno custo computacional, somado ao foco de baixo custo do projeto.

4.2.3 Módulo de Saída

Após processada as informações, apresenta-se o módulo de saída, que tem como objetivo, encaminhar os elementos coletados as demais camadas, na qual este modelo planeja implementar duas formas de encaminhamento dos dados. Os dois tipos são definidos como componentes *Remoto e Local*, em que o primeiro envia por meio de redes de dados e o segundo, persistindo às variáveis localmente. A escolha dá-se na primeira vez que o sistema é executado.

O componente remoto transfere os dados para o segmento de comunicação através de uma placa expansora de transmissão WI-FI, conectada ao microprocessador. Pretende-se trabalhar com a tecnologia IEEE 802.11b, tendo em vista que é uma via de comunicação bastante difundida, conhecida e barata, sendo a escolhida para transmissão entre a colmeia e o gateway da rede WSN (mais detalhes sobre comunicação na seção 4.3).

O componente local persiste os dados coletados em arquivo texto (dataset), por intermédio de leitor/gravador SD Card plugado ao microprocessador. Propõe-se através desta funcionalidade incorporada ao BeeBr, uma alternativa aos apicultores que encontram-se em localidades distantes, sem conectividade de redes móveis. Desta forma, a camada de comunicação não é acionada, sendo os dados repassados diretamente ao segmento de serviço, por intermédio de um processo manual. Este recurso é denominado no BeeBr pelo nome de DDS (Detecção-Direto-Serviço). De posse do arquivo, armazenado em cartão de memória, na qual contém todas as

leituras realizadas na colmeia até o momento, o apicultor disponibiliza-o como input para a seção de serviço.

4.2.4 Módulo de Alimentação

Responsável por fornecer energia para os componentes que fazem parte da camada de detecção, este módulo é composto por bateria e placa solar. Sendo uma peça fundamental ao modelo, dado que os ambientes nas quais encontram-se os apiários são distantes a qualquer tipo de infraestrutura, devido as questões legais de segurança.

Na primeira etapa do BeeBr, propõem-se a instalação dos componentes (bateria/placa solar) por colmeia. Em estágio futuro, pretende-se colocar baterias e placas solares de grande porte, de modo que alimenta todo o apiário, a fim de diminuir os gastos, tendo em vista que estes equipamentos são de alto valor, o que vai na contramão de baixo custo na concepção do BeeBr. Sendo assim, sugere-se o uso de baterias usadas em motos.

Planeja-se o uso de energia solar, a partir de placas solares que ofereça a carga consumida pelo sistema ao dia e recarregue a bateria durante este turno, deixando os componentes sempre operantes.

Por fim, espera-se implementar um algoritmo que otimize o uso dos recursos computacionais, de modo que faça o módulo de alimentação manter e disponibilizar a energia armazenada o máximo de tempo possível.

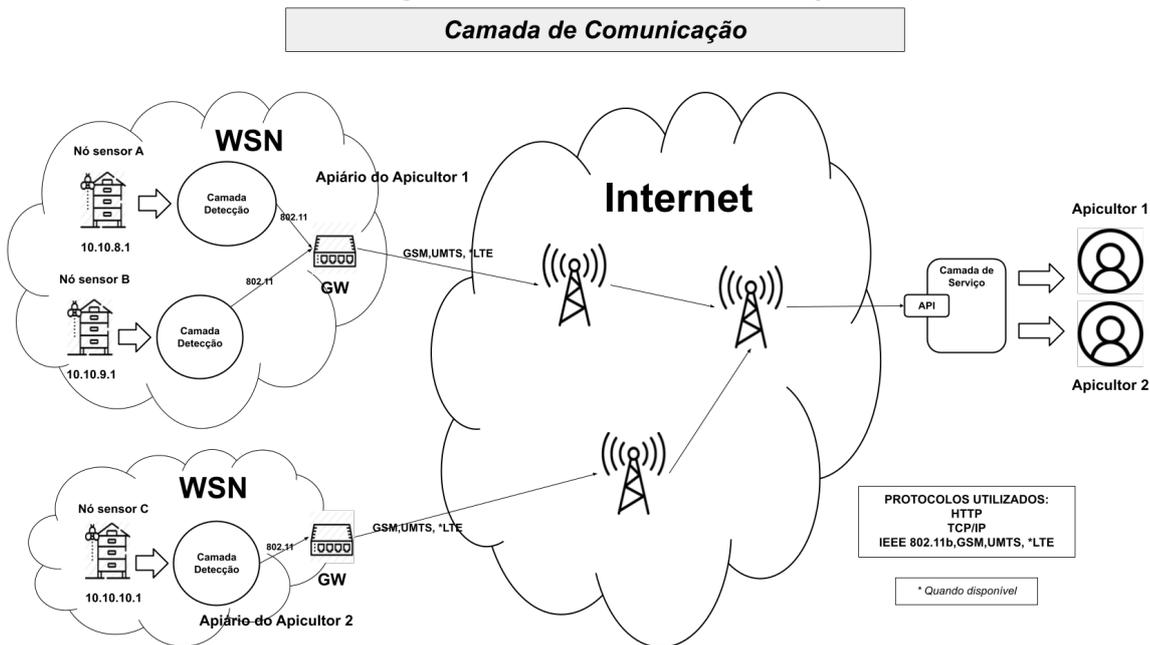
4.3 Camada de Comunicação

A camada seguinte a ser exposta, denomina-se Comunicação, na qual é responsável por transmitir as informações coletadas no apiário até ao segmento de serviço, por meio de redes locais e móveis, desenvolvendo as atividades apresentadas pela fundamentação teórica, disposta na seção 2.1.2. Dentro do modelo BeeBr, constitui-se o campo de interesse de estudos sobre comunicação de dados móveis. Propõem-se uma camada usual, eficiente, econômica e extremamente objetiva, visando uma topologia que faça uso de uma infraestrutura mínima de recursos, com baixos custos operacionais e que atenda a finalidade do projeto. Valendo-se de redes e protocolos de comunicação padronizado no meio da internet das coisas, modo de transmissão simplex e por fim, provisionamento por colmeia. A Figura 13 expõem em detalhes a comunicação.

4.3.1 Redes e Protocolos

Pretende-se usar duas redes heterogêneas, na quais comunicam-se entre si, a fim de realizar o objetivo de transmitir os dados provenientes do segmento de detecção até a camada de serviço, por meio da agregação das redes de comunicação WSN (Wireless Sensor Network) e Internet.

Figura 13 – Camada de Comunicação



Fonte: Elaborado pelo autor. (2021)

A opção por esta infraestrutura, se dá por ser um padrão estabelecido, testado e consolidado dentro da comunicação de objetos IoT (Li; Xu; Zhao, 2015).

O BeeBr apresenta uma Wireless Sensor Network implementada e estruturada via topologia estrela, responsável por trafegar os dados originados pelos nós sensores (colmeia/camada de detecção) até o concentrador da rede (Gateway), este por sua vez, estando na borda da rede WSN, encaminha os pacotes para a Internet.

A comunicação entre nó sensor e GW deve-se realizar através do padrão IEEE 802.11, na qual constituem-se protocolos de comunicação em nível de enlace. Para sua escolha, leva-se em conta dois pontos; o primeiro pelo fato de ser um meio difundido, conhecido e com custo acessível de implementar. Outro fator são as características de velocidades e raio de alcance, tendo em vista que as colmeias estão uma perto da outra e agrupadas em perímetros onde o 802.11 possibilita acesso. Caso haja nós sensores com distância superior a especificada pelo protocolo, é possível inserir repetidores de sinal.

Pretende-se usar microcontrolador de baixo custo para realizar o papel do GW, pois introduz valor reduzido em comparação a GW's proprietários e dedicados e por suas especificações atenderem ao modelo. Por ser um equipamento eletrônico, faz necessário uma fonte energética. Como mencionado na subseção 4.2.4, em versão futura do BeeBr, baterias e placas solares serão usadas para alimentar todos os dispositivos no apiário, mas em estágios iniciais é concentrado em uma bateria e placa solar. Deve-se dispor no GW a solução DDS, mecanismo na qual possibilita armazenar os dados das colmeias de forma local, caso não seja possível transmitir estes valores para a Internet e assim viabilizar ao apicultor, a recuperação das informações de todo apiário diretamente no dispositivo; sendo mais uma alternativa aos usuários que encontram-se

em localidades distantes e sem conectividade de redes móveis. Para o nó sensor, este recurso necessita ser transparente, visto que em sua configuração do módulo de saída, deve estar disposto (camada de detecção -> módulo de saída -> componente remoto).

Do ponto de vista da comunicação com a Internet, opta pelo uso das redes móveis 2G (GSM) e 3G (UMTS), por serem as alternativas mais viáveis nos aspectos técnico e financeiro. Em algumas regiões onde encontram-se os apiários, o acesso à comunicação é bastante limitada ou inexistente. Os elementos a serem transmitidas são pequenos pacotes de dados, não necessitando de grande largura de banda. Desta forma, através de um plano simples de internet 2/3G, o sistema está apto a comunicar-se. Após encaminhada a Internet, esta por sua vez, ofertada pelas operadoras de telecomunicação, encarrega-se de transmitir os dados até a camada de serviço. Caso o local tenha disponibilidade, o modelo suportaria redes 4G (LTE), havendo a necessidade de adicionar placas expansoras para este tipo de tecnologia.

A Tabela 4 expõem uma visão geral dos protocolos propostos no uso pelas redes, divididos por camadas do modelo TCP/IP.

Tabela 4 – Protocolos propostos empregados na camada de comunicação

Camadas TCP/IP	Rede WSN	Rede Internet
Aplicação	HTTP	HTTP
Transporte	TCP	TCP
Rede	IP	IP
Enlace	IEEE 802.11b/g/n	2G = GSM / 3G = UMTS / *4G = LTE

Fonte: Elaborado pelo autor. (2021)

Em nível de aplicação, pretende-se o usar o protocolo HTTP pelas suas características, como simplicidade, leveza, facilidade de implementação e integração junto aos serviços em nuvem que oferecem em sua grande parte, comunicação por meio de uma API, na qual usa-se a arquitetura REST, que se vale por sua vez do protocolo HTTP. Nas camadas de transporte e rede são empregado os tradicionais TCP/IP respectivamente. Por fim, na camada de enlace, é utilizado na rede WSN as especificações IEEE 802.11b/g/n e Internet (2G = GSM / 3G = UMTS / *4G = LTE), este último, quando disponível.

4.3.2 Modos de Transmissão

É definido o modo de transmissão simplex ou sentido único. Entende-se que o modelo necessita de trocas de pacote, somente na direção: camada de detecção -> camada de comunicação -> camada de serviço. Percebe-se a não obrigação da colmeia (camada de detecção), ser um receptor de dados, a fim de executar uma possível atividade na mesma.

4.3.3 Provisionamento

Deseja-se viabilizar no BeeBr o conceito de aplicação multitenancy, exposto na fundamentação teórica em seção 2.1.3, na qual permitirá aproveitar ao máximo a infraestrutura compartilhada entre os clientes (*locatários*).

A reprodução do cenário multitenancy é mostrada na Figura 13, nas quais existem dois apiários de apicultores distintos (*Apiário do Apicultor 1 / Apiário do Apicultor 2*), comunicando-se com o segmento de serviço compartilhada, que por sua vez, disponibiliza interfaces gráficas separadas por usuário. Mais detalhes na seção 4.4.

A partir deste recurso, uma colmeia que visa ser adicionada ao sistema pelo componente remoto 4.2.3 (camada detecção -> módulo de saída -> componente remoto) , faz necessário o seu provisionamento, na qual o apicultor digita as informações referente ao seu cadastro, dados relacionados ao apiário e/ou colmeia (localização, identificação) e por fim, o endereço IP do servidor que expõem a API.

4.4 Camada de Serviço

A última camada proposta por esta pesquisa, propõem-se a processar, analisar, prever e armazenar os elementos coletados no apiário e assim, ofertá-los aos usuários, a partir de uma plataforma contratada e hospedada na nuvem. Sendo implementada e viabilizada em cima da arquitetura SaaS, na qual oferece uma infraestrutura completa de hardware e software. Dispõem dos princípios expostos pela fundamentação teórica, posta na seção 2.1.3 .

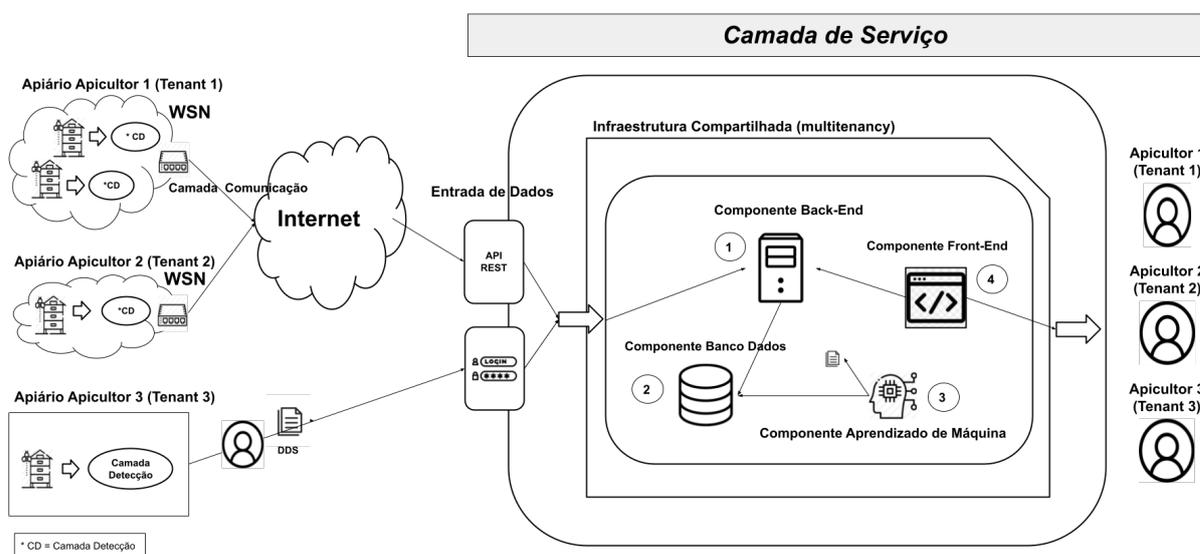
Entende-se que o modelo de negócio SaaS, obtém vantagens para o projeto, tanto do ponto de vista técnico quanto financeiro, visto que, a plataforma traz diversas facilidades técnicas, como ferramentas disponíveis para configurar, desenvolver, implantar e suporte ao usuário. E por trazer o conceito multitenancy, em que é possível compartilhar os recursos de infraestrutura entre os clientes, trazendo deste modo ganhos financeiros para a proposta.

Propõe-se um ecossistema de componentes de software independentes, que tem como objetivo executar as rotinas planejadas pela camada de serviço, nas quais usufruem dos recursos compartilhados pela plataforma. Na listagem a seguir, são expostos com seus respectivos e breves descrições:

- Back-End: Processar os elementos do apiário.
- Front-End: Disponibiliza uma interface web ao apicultor.
- Banco de dados: Manipulação do conhecimento.
- Aprendizado de Máquina: Gerenciamento de modelos de aprendizado de máquina.

Na Figura 14 é possível visualizar em detalhes a camada de serviço.

Figura 14 – Camada de Serviço



Fonte: Elaborado pelo autor. (2021)

Percebe-se que o fluxo é capaz de ser iniciado por meio de duas vias, sendo a primeira uma API REST e a seguinte, uma interface web, em que o apicultor disponibiliza as leituras diretamente ao sistema. Em cima destes dados injetados ao segmento, o componente Back-End (1) inicia seu trabalho de processamento das informações; inclui-se nesta tarefa operações junto ao banco de dados (2).

Por ser tratar de uma etapa importante dentro do BeeBr, optou-se na elaboração de um software específico na administração de modelos de aprendizado de máquina, denominado de componente de Aprendizado de Máquina (3). Por fim, o usuário tem acesso ao sistema, através de uma interface gráfica disponibilizada pelo componente Front-End (4). Na sequência são expostos os detalhes do comportamento das partes do segmento de serviço.

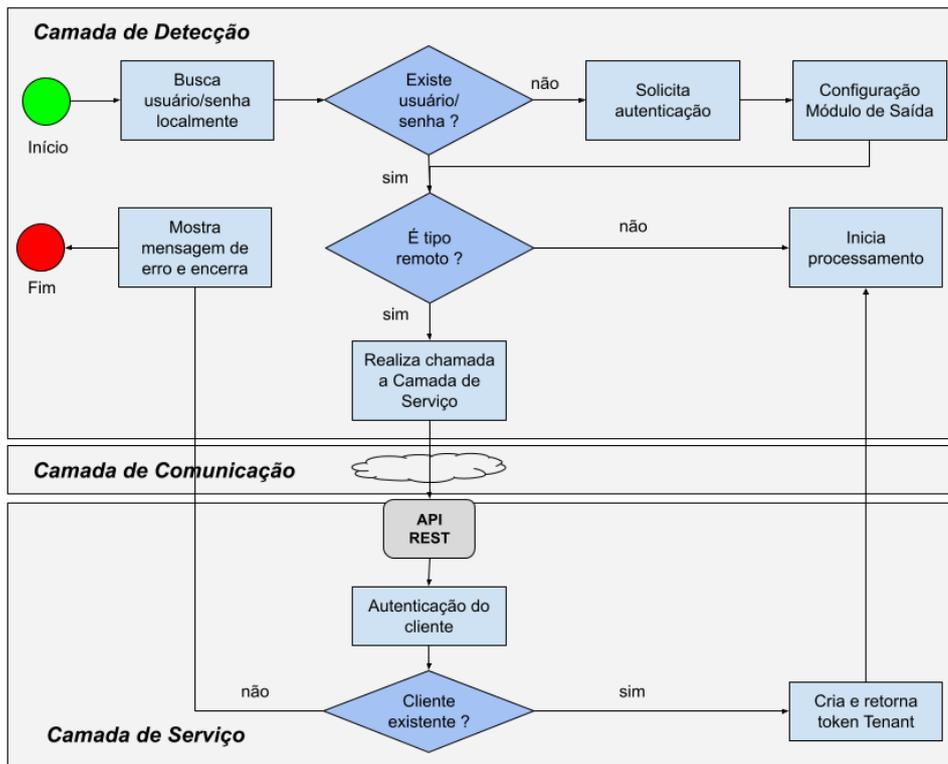
4.4.1 SaaS - Multitenancy

Percebe-se na Figura 14, três apicultores (Apicultor 1 - Tenant 1), (Apicultor 2 - Tenant 2) e (Apicultor 3 - Tenant 3) que usufruem de todo o potencial da camada. Internamente, é compartilhado os recursos (hardware, ferramentas e software desenvolvidos) entres os usuários, ofertados pelo conceito multitenancy. Todo o processo é transparente perante ao apicultor, sendo disponibilizado uma interface, na qual é apresentado somente dados relacionados aos seus apiários.

Para fazer uso destas capacidades faz necessário efetuar o registro do usuário. Com posse de sua credencial, o apicultor realiza o provisionamento da colmeia por meio do software embarcado instalado na camada de detecção. Na Figura 15 visualiza-se este fluxo.

Ao iniciar o sistema, deve-se verificar a existência no banco de dados local as informações

Figura 15 – Fluxograma no provisionamento das Colmeias



Fonte: Elaborado pelo autor. (2021)

de usuário e senha (credencial). Se não houver, é preciso solicitar ao operador a autenticação e modo de operação do módulo de saída, conforme discutido na subsecção 4.2.3, a fim de determinar qual meio de evasão da camada de detecção.

Ao selecionar o componente local, inicia-se o processamento, caso contrário, é feita a comunicação junto ao segmento de serviço, com as credenciais vinculada ao nó sensor. Na base de dados remoto é efetuada a autenticação. Estando válida, retorna-se um token, na qual é usada nas requisições subsequentes. Havendo problemas, apresenta-se um erro ao apicultor e encerra-se a execução.

4.4.2 Entrada de Dados

Entende-se como entrada de dados, um canal de comunicação entre o mundo externo e a camada de serviço. A partir desta porta, permite-se receber os elementos das colmeias e assim, iniciar o fluxo de processamento. Pretende-se contar com duas entradas de dados com propósitos específicos e características distintas.

O primeiro meio proposto é expor uma API REST em cima do protocolo HTTP. Por ser uma interface conhecida, de fácil implementação, disponível nos provedores e empregada atualmente como arquitetura na integração de sistemas, percebe-se neste método a forma mais adequada para disponibilizar o acesso ao mundo externo. Desta forma, as informações oriun-

das do apiário, transmitidas através das redes WSN e Internet, serão recebidas pela camada de serviço, por meio desta entrada de dados.

A fim de oferecer uma alternativa aos apicultores que não tem conectividade móvel na região onde encontram-se seus apiários ou para colmeias que estejam em posições na qual o sinal não é captado, é proposto a solução DDS, na qual faz uso deste canal de comunicação. Através de uma interface web, possibilita ao apicultor, realizar o upload do dataset da colmeia, diretamente na camada de serviço. Este cenário é visualizado na Figura 14 por meio do Apicultor 3. Em futuras atualizações do BeeBr, pretende-se viabilizar recursos de software em modo off-line, com o propósito do usuário ser capaz de efetuar tal atividade diretamente no apiário.

4.4.3 Componentes

Considera-se componente, um software que objetiva-se a processar, analisar, salvar, gerenciar modelos de AM, dispor informações ao apicultor; nas quais fazem parte do ecossistema de aplicações a serem modelados, implementados e/ou configuradas dentro do escopo deste estudo. É segmentado em dois tipos, na qual a primeira são aplicações desenvolvidas em linguagem de alto nível e por fim a última, ferramentas que auxiliam como um todo.

Os quatro componentes propostos possuem características de autonomia e independência em relação à plataforma SaaS. Isto oferece flexibilidades no processo de software, tais como a análise, desenvolvimento, testes, implantação e execução das aplicações. Por fim, dentro da arquitetura SaaS - Multitenancy, os componentes são instâncias únicas de programas, nas quais compartilhadas entre os apicultores.

4.4.4 Componente - Back-End / Front-End

Os componentes back-end e front-end focam-se nas etapas de processar e visualizar respectivamente os elementos coletadas pelo sistema. Contempla no processamento do back-end, a exposição de API's Rest, nas quais tem como objetivo, receber os dados provenientes dos apiários e disponibilizá-los ao front-end. Também as validações em nível de negócio fazem parte e por fim, a criação, remoção, atualização e busca do conhecimento junto ao banco.

As atividades do front são apresentar uma interface gráfica web amigável e viabilizá-la de forma rápida, fácil e customizada ao usuário, de modo que possa-o acessar os valores do apiário. O front deve conectar o back por meio das API's. Para acesso à aplicação faz necessário realizar autenticação.

4.4.5 Componente - Banco de Dados

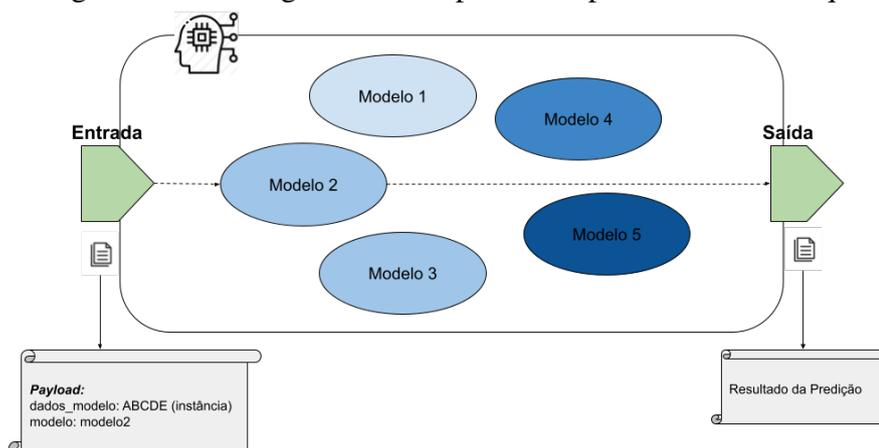
Dentre os quatro componentes, o banco de dados é o único a não ser desenvolvido, desta forma, ele encaixa-se no grupo de ferramenta, auxiliando no processo de manipulação das in-

formações - CRUD (Create, Read, Update, Delete). Pretende-se usar o modelo relacional como estrutura de armazenamento, devido às suas características de consistência, atomicidade e segurança.

4.4.6 Componente - Aprendizado de Máquina

O último componente a ser exposto, objetiva-se no gerenciamento de modelos de aprendizado de máquina. Desta forma, este software é formado em seu núcleo por modelos de aprendizado de máquina, que destinam-se na preempção de tais dificuldades encontradas na apicultura. Por isso, na concepção desta entidade, entende-se que sua estrutura deve ser plugável do ponto de vista da inclusão de novos algoritmos de AM; ou seja, ao criar-se um novo modelo, o objeto deve dispor de características que facilitem este processo. A Figura 16 apresenta uma visão geral do componente.

Figura 16 – Visão geral do Componente Aprendizado de Máquina



Fonte: Elaborado pelo autor. (2021)

Na Figura 16 expõe como exemplo, cinco modelos de AM, incluídos no componente, najs quais, estão treinados e prontos para utilização no BeeBr. Também percebe-se o elemento denominado *Entrada*, na qual é a porta de acesso das informações pelas quais serão processadas pelos modelos, bem como, o item *Saída* em que exibe os resultados da predição atingido pelo modelo processado. Deve-se viabilizar ao componente um payload que ofereça dados para o modelo (instância), assim como, qual algoritmo a ser invocado. Baseado nestas explicações, o objeto dispõe de elementos básicos para execução da predição. Na Figura 16 mostra um exemplo de execução do modelo 2.

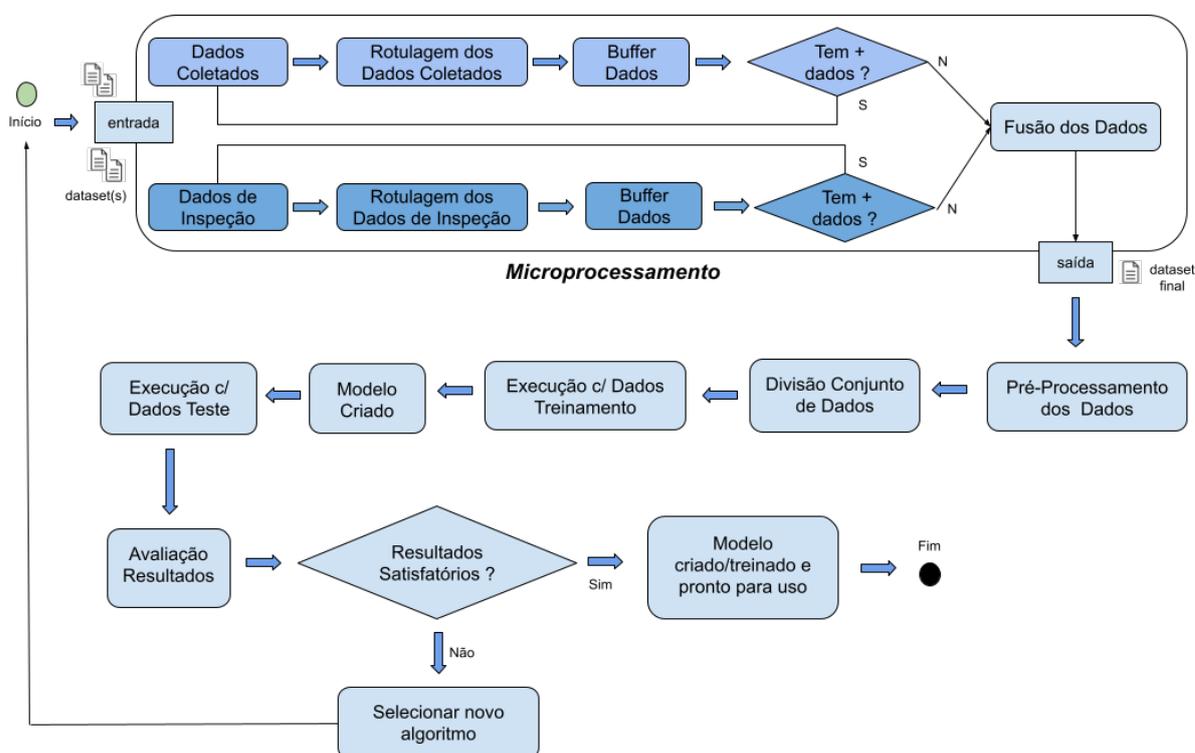
4.5 Elaboração e treinamento do novo modelo de aprendizado de máquina

Diante dos objetivos expostos nesta dissertação, das quais, apresenta-se a elaboração e treinamento de um novo modelo de aprendizado de máquina, expõem-se na sequência as etapas de

modelagem e treinamento do objeto citado. Neste sentido, é exibido um moderno modelo de aprendizado de máquina, na qual diferencia-se do estado da arte, a partir do momento que retrata o objetivo de predição da enxameação (fuga das abelhas); baseando-se na inclusão de classes oriundas de dados coletados da colmeia, em cima de modernos sensores IoT, bem como, de pesquisas efetuadas em campo, por meio de especialistas do segmento, a fim de trazer a preempção da enxameação. Ressalta-se que o mesmo dispõe de independente em seu funcionamento, ou seja, pode-se usá-lo sem a interferência do BeeBr, ou, acoplá-lo ao componente de aprendizado de máquina do modelo proposto BeeBr.

A fim de atingir o foco proposto, baseado nos princípios apresentados na seção 2.2, a Figura 17 expõem o fluxo a realizar, de finalidade a completar as etapas de criação e treinamento do novo modelo de aprendizado de máquina, apontado neste estudo.

Figura 17 – Fluxo de criação e treinamento do novo modelo de aprendizado de máquina



Fonte: Elaborado pelo autor. (2021)

Uma premissa básica na concepção do novo modelo de predição da enxameação, é o entendimento que este, encontra-se vinculado ao algoritmo de AM, visto que, baseado nas avaliações dos resultados do mesmo, e que por ventura seja negativa, todo o processo exposto na Figura 17 deve ser refeito. Perante ao exibido, a Figura 17 mostra a fase inicial denominada de *Microprocessamento*, na qual é formada por um conjunto de sub-rotinas, que tem como foco, receber, rotular, armazenar e unir todas as informações recebidas (dataset(s)). Todo o processo enlobado pelo microprocessamento é executado até não mais dispor de dados.

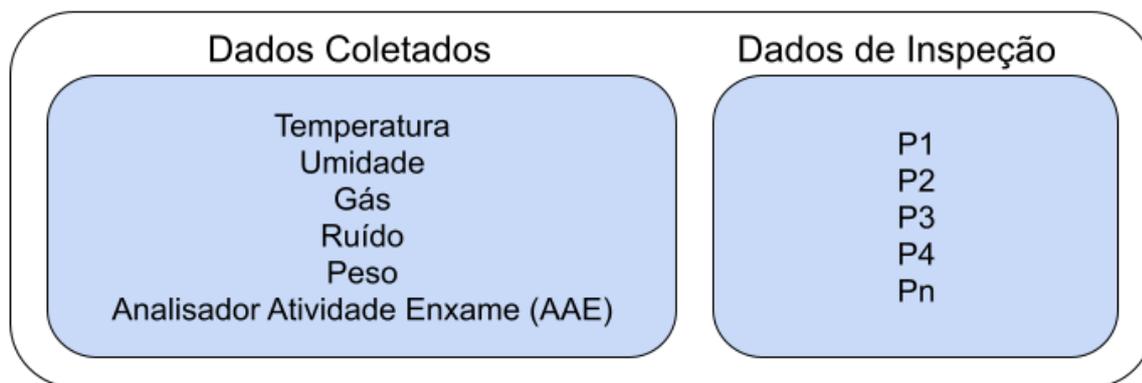
Inicialmente, propõem-se dois tipos de dataset(s) na entrada do microprocessamento, que

são ofertadas para a criação e treinamento do modelo. Baseado nesta união das informações recebidas, sugere-se a modelagem dos atributos/classes do modelo. A primeira coleção de informações, denomina-se *Dados Coletados*, nas quais são constituídas por dados coletados e recebidos diretamente da colmeia, por meio dos dispositivos de IoT. O modelo deve dispor em sua modelagem às variáveis de temperatura, umidade, gás, ruído, peso e analisador atividade enxame (AAE). Entende-se que AAE é uma peça fundamental na construção do modelo, dado que, mediante as leituras praticadas pelo sensor apresentado em seção 4.2.1 pretende-se averiguar situações de abandono em massa da colônia, bem como, trazer subsídios para os algoritmos de aprendizado de máquina, a fim de prever o acontecimento de enxameação. Tais elementos devem ser tipados com valores numéricos.

O segundo grupo de informações é nomeado de *Dados de Inspeção*, que tem como o foco oferecer conhecimento sobre o comportamento das abelhas, através de um olhar humano. Neste sentido, propõem-se a elaboração de perguntas pertinentes à apicultura, na qual busque trazer a compreensão destes insetos. Sugere-se que tais questionamentos sejam formulados, a partir de um pensamento científico, e, principalmente, da experiência dos apicultores.

Na Figura 18 é apresentado um resumo das classes propostas para o modelo, baseado no conjunto de dados expostos.

Figura 18 – Resumo das classes propostas para o modelo



Fonte: Elaborado pelo autor. (2021)

As questões a serem formuladas nos dados de inspeção são infinitas; pode-se então dispor de n perguntas ao conjunto. Ainda dentro da fase de microprocessamento, após o recebimento das informações, é disposto a etapa de rotulagem dos dados; que objetiva a rotular às amostras, a fim de identificar, descrever e caracterizar um aspecto e/ou traço da saúde e bem estar das abelhas, dentro do conjunto de dados coletados e/ou inspeção. Desta forma, as amostras levantadas devem dispor de um rótulo na qual indica o estado da colmeia, em específico características que denotam problemas de enxameação.

Após finalizada a rotulagem dos dados, os mesmos são inseridos em um buffer. Posteriormente, verifica-se a existência de demais dataset(s). Caso exista, o processo é executado

novamente, até que não disponha mais. A partir desta abordagem, entende-se que seja possível disponibilizar um dataset rico de instâncias de diferentes valores de classes e rótulos. Na sequência do fluxo de criação e treinamento do novo modelo de aprendizado de máquina, deve-se realizar a *Fusão dos Dados*, na qual consiste na união dos elementos dispostos pelo buffer, oriundos das coletas/inspeções e assim, resultar em um único arquivo. Por fim, o microprocessamento entrega em sua saída o dataset final gerada pela sub-rotina anterior.

Completada o estágio de microprocessamento, o fluxo apresenta as etapas habituais no treinamento de um modelo de AM, nas quais são: pré-processamento, divisão de conjunto de dados (treino/teste), execução do algoritmo selecionado através das informações de treinamento/teste e avaliação dos resultados; sendo essas fases detalhadas na seção 2.2. Finalmente, deve-se avaliar se as respostas geradas pelo algoritmo indicado são adequadas. Sugere-se a análise por meio das métricas: Accuracy, Precision, Recall, F1-score e Matrix de confusão. Caso a resposta seja apropriada, o modelo está criado, treinado e apto para ser incluído no componente de aprendizado de máquina. Na hipótese contrária, recomenda-se a escolha de um novo algoritmo e refazer todo o fluxo apresentado.

Desta forma, este capítulo expôs o modelo proposto denominado BeeBr, bem como a elaboração e treinamento de um novo modelo de aprendizado de máquina na qual visa a predição da enxameação. No capítulo seguinte é exposto as implementações e metodologias dos principais recursos do BeeBr.

5 IMPLEMENTAÇÃO DO MODELO

Conforme apresentado no capítulo 4, o modelo BeeBr é segmentado nas camadas de detecção 4.2, comunicação 4.3 e serviço 4.4. Desta maneira, este capítulo descreve os aspectos de seu desenvolvimento. Pretende-se enfatizar as camadas de detecção e serviço, implementando assim, protótipos com as principais características propostas nestes segmentos, bem como os objetivos definidos nesta pesquisa. Devido às dificuldades de conectividade encontradas onde localizam os apiários, entendeu-se que neste momento, a camada de comunicação não será implementada. A divisão deste capítulo encontra-se da seguinte maneira; a primeira parte é exposto uma prova de conceito da camada de detecção, em seguida a implementação dos componentes do segmento de serviço e por fim a elaboração e treinamento do novo modelo de aprendizado de máquina.

5.1 Camada de Detecção

Inicialmente, realizou-se uma prova de conceito da camada de detecção, na qual objetivasse abordar os principais aspectos discutidos no modelo BeeBr. Dentre estes, estão as implementações dos módulos de entrada, processamento, saída e alimentação, nas quais, são compostos por elementos de IoT. Numa primeira fase, efetuou-se uma pré-validação no protótipo, em que a partir dos resultados, proporcionou seu aperfeiçoamento.

5.1.1 Desenvolvimento do Protótipo

Inicialmente, percebeu-se que o manejo das abelhas são atividades muito perigosas, dado que, as picadas das abelhas em grande quantidade podem ocasionar a morte de um ser humano. Por este motivo e pelo fato do apiário estar a uma distância de 250 Km em relação ao domicílio deste pesquisador, solicitou-se aos donos do apiário, uma caixa de abelha, a fim de realizar testes em bancada com segurança e agilidade. Em cima da colmeia teste, possibilitou levantar as dificuldades e planejar soluções a ultrapassá-las. Sendo assim, foram encontrados no início do processo de desenvolvimento os seguintes desafios:

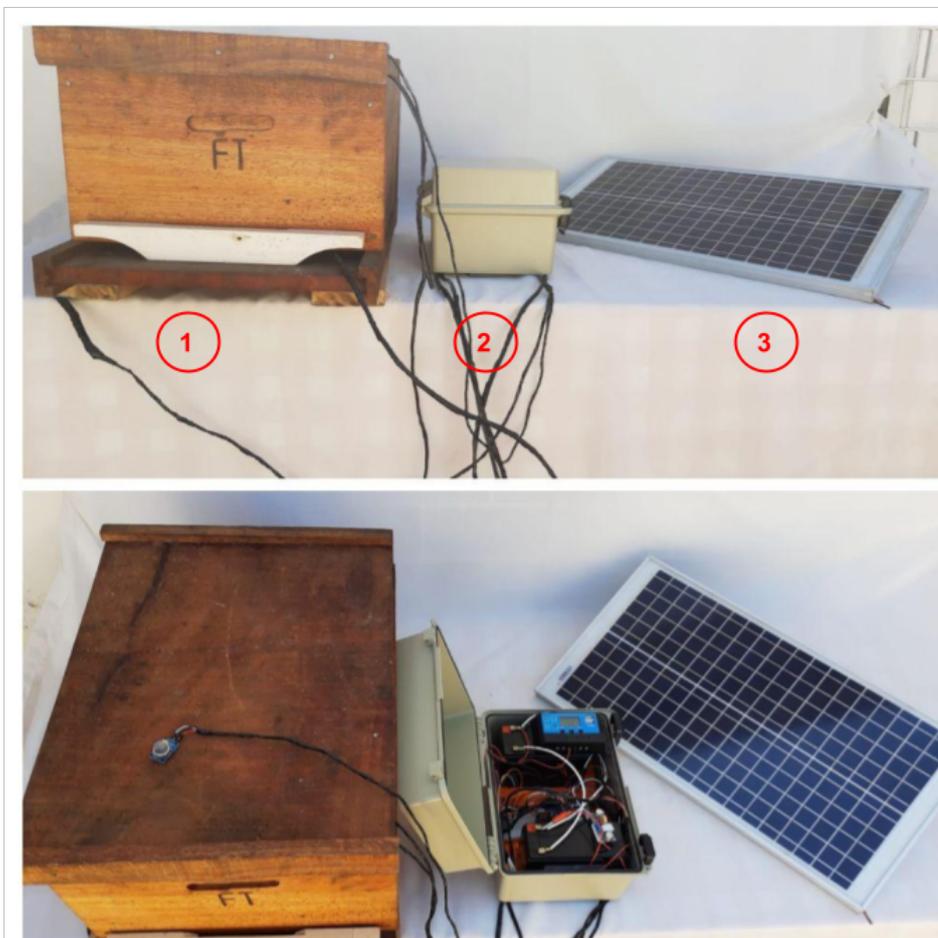
- A - Quais os melhores locais para colocar os sensores para que não atrapalhem o bem estar das abelhas ?
- B - Como proteger estes sensores das atividades das abelhas ?
- C - Como inserir um sensor de peso, debaixo da caixa ?
- D - Como proteger e qual local a ser colocado os equipamentos (microprocessador, baterias, cabos e controlador de carga)
- E - Como implementar o analisador de atividade do enxame ?

- F - Quais características importantes da placa solar a ser introduzida no projeto ?
- G - A internet móvel está disponível nos locais dos apiários ?

As indagações (A,B,C,D,E,F) são respondidas nas próximas subseções, mas, o item G (relacionada com a disponibilidade móvel) verificou-se que no local do apiário escolhido não constava de sinal de rede, visto que é um local muito íngreme. Desta forma, entendeu-se que não seria possível a implementação da camada de comunicação nesta pesquisa.

Após oito meses de planejamento, desenvolvimento e testes, surge a primeira versão do protótipo de hardware do segmento de detecção, na qual pode ser visualizada na Figura 19.

Figura 19 – Protótipo da Camada de Detecção



Fonte: Elaborado pelo autor. (2021)

A caixa de madeira, representada pelo número 1 na circunferência em vermelho, dispõe em sua entrada um item denominado no segmento da apicultura de **alvado**, na qual localiza-se um sensor. Em cima e internamente estão os demais sensores. Identificado ao meio da figura pelo valor 2, encontra-se o gabinete com resistência a água, que armazena os dispositivos eletrônicos que compõem a camada, e por fim, apontado pelo número 3, consta a placa solar.

Na Figura 20 é apresentado em detalhes o interior do gabinete com os periféricos.

Figura 20 – Detalhes do Gabinete e seu dispositivos



Fonte: Elaborado pelo autor. (2021)

A figura apresenta os itens identificados pela seguinte sequência numérica: (1) controlador de carga; (2) microprocessador Arduino; (3) regulador de tensão; (4) bateria A; (5) bateria B e (6) conectores e leitor/gravador SD Card. Nas subseções seguintes serão detalhadas as implementações dos módulos da camada de detecção e por consequência os equipamentos aqui relatados, do qual fazem parte.

5.1.2 Módulo Entrada

O módulo de entrada contempla os sensores que monitoram a colmeia. Cada sensor realiza a atividade de leitura de uma determinada informação e em seguida encaminha ao módulo de processamento. O modelo BeeBr propõe a inclusão de sete sensores e oportuniza a inserção de demais. Neste sentido, essa implementação atuará nos sensores de temperatura, umidade, ruído, gás, peso e analisador atividade enxame (AAE) conforme proposto.

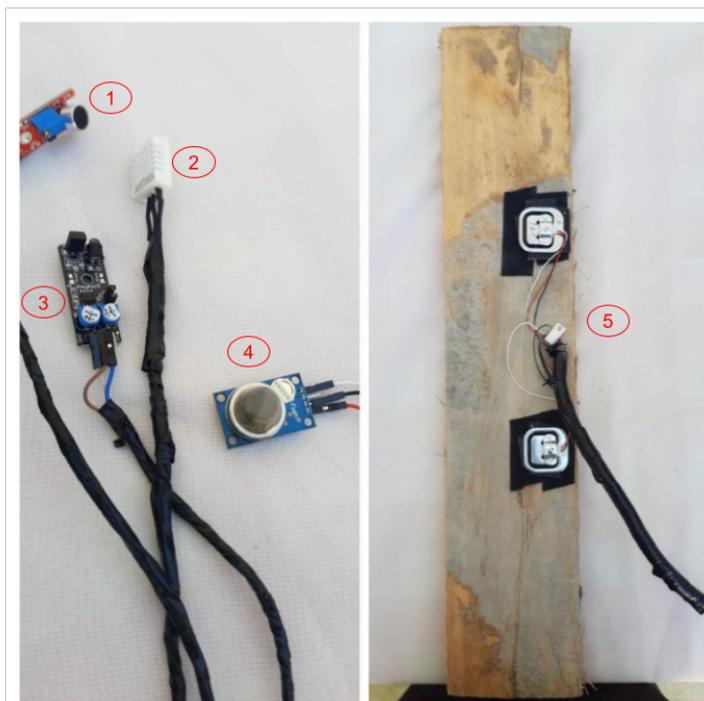
O mercado de sensores dispõe de uma grande variedade de opções e preços. Tendo em vista que uma das premissas definidas pelo modelo é o seu baixo custo, buscou-se então dispositivos de custo/benefício na qual atendessem com qualidade, somado ao valor acessível. No final, definiu-se os equipamentos baseado em depoimentos de usuários, além de testes em bancada a fim de determinar a melhor escolha. Os sensores escolhidos são discutidos na subseção Sensores 5.1.2.1.

Após realizada a escolha, levantou-se as indagações *A - Quais os melhores locais para colocar os sensores para que não atrapalhem o bem estar das abelhas ?*; *B - Como proteger estes sensores das atividades das abelhas ?* e *C - Como inserir um sensor de peso, debaixo da caixa ?* da listagem 5.1.1. Segundo os apicultores que auxiliaram nesta pesquisa, não seria interessante dispor os sensores em qualquer posição dentro da caixa, pois isto dificultaria o trabalho das abelhas. Outro ponto levantado por eles, seria a possibilidade das abelhas propolisarem os equipamentos. Para solucionar estes pontos, colocou-se sensores no piso da caixa, em posições que minimizem os prováveis problemas de trânsito dos insetos. Para proteção, inseriu-se pedaços de tecido em torno de cada sensor, aos quais estavam dispostos no interior da colmeia. Em relação ao questionamento C, observou-se possíveis dificuldades ao inserir uma balança debaixo da colmeia, haja vista que é uma operação delicada, pois é necessário movimentar o caixote para cima, o que faz com que as abelhas fiquem agitadas. Para contornar esta adversidade, optou-se em comum acordo (pesquisador e apicultores), uma simulação junto a caixa teste. Desta forma, todos praticam os movimentos de inserção do dispositivo, evitando assim, reverses na implantação real.

5.1.2.1 Sensores

Nesta seção é exposto os sensores selecionados e suas principais características. Exceto o sensor de temperatura/umidade, todos os demais precisaram efetuar calibrações, adequando as necessidades desta pesquisa. Na Figura 21 é possível visualizá-los.

Figura 21 – Sensores usados na camada de detecção



Utilizou-se o sensor de ruído modelo KY-038, identificado pelo número 1 na figura. Sua saída é digital e analógica, na qual selecionou-se a segunda, pois por meio desta configuração possibilita valores entre 0 à 1024 ajustável via potenciômetro. Sua instalação deu-se internamente, no assoalho da caixa.

Assinalado pelo valor 2 na figura, o sensor de temperatura e umidade modelo DHT22, dispõe das seguintes características. Faixa de medição de umidade entre 0 a 100% UR e temperatura -40° a $+80^{\circ}\text{C}$. Precisão na medição de umidade $\pm 2,0\%$ UR e temperatura $\pm 0,5^{\circ}\text{C}$. Medição da temperatura em Graus Celsius. Instalou-se junto ao sensor de ruído.

Item 4 da figura, o detector de gás modelo MQ-135, oferece a capacidade de identificar gases do tipo amônia, óxido nítrico, álcool, benzeno, dióxido de carbono e fumaça. Seus resultados são oferecidos por meio de uma porta analógica. Acomodou-se no piso do caixote.

Para medir o peso da colmeia, empregou-se duas células de carga modelo YZC-133, nas quais desfrutavam de capacidade total de 100 Kg. Em conjunto, usou-se o amplificador de sinal HX711 como o módulo conversor. Conforme visualizada na Figura 21, em destaque com o número 5, optou-se em instalá-los em um pedaço de madeira, com dimensões iguais às da caixa. Sua instalação se dá em uma extremidade debaixo do caixote, sendo necessário inserir na outra ponta, um mesmo fragmento de madeira, a fim de equilibrar o peso.

Por fim, identificado pelo valor 3, encontra-se o sensor de obstáculo infravermelho KY-032. Implementou-se um algoritmo que identifica e contabiliza objetos que passam em sua frente. Desta forma, obteve-se o que o modelo define como analisador atividade enxame (AAE), na qual seu objetivo é contar a quantidade de abelhas que passam ao seu redor. Definiu-se sua instalação em um dos orifícios do alvado, sendo que a entrada oposta, colocou-se obstáculos que impediam o acesso das abelhas. Baseado nestas soluções, possibilitou responder a indagação *E - Como implementar o analisador de atividade do enxame ?* da listagem 5.1.1.

5.1.3 Módulo de Processamento

Centro da camada de detecção, o módulo de processamento objetiva processar todas as informações originadas pelos módulos de entrada/saída, por meio de um microprocessador, na qual este é discutido em detalhes na seção seguinte 5.1.3.1. Somado a este dispositivo, a implementação dispõe de uma placa de ensaio (Protoboard), usada na conexão entre sensores e placas externas ao microprocessador através de jumpers. A estrutura completa é visualizada na Figura 20, em que a CPU (item 2) encontra-se conectada com a protoboard.

Levantou-se a indagação *D - Como proteger e qual local a ser colocado os equipamentos (microprocessador, baterias, cabos e controlador de carga)* da listagem 5.1.1, referente ao armazenamento e segurança destes dispositivos, visto que são objetos delicados que não suportam chuva, neve ou qualquer adversidade que os danifiquem. Por este motivo, comprou-se um gabinete que oferecesse total segurança a tais fatalidades. Um ponto levantado junto aos apicultores foi a escolha do local onde seria instalado o gabinete. Decidiu-se colocar em cima da colmeia,

na qual possui um telhado, oferecendo uma proteção a mais aos equipamentos.

5.1.3.1 Microprocessador

A escolha do microprocessador deu-se pelo fator custo e seu baixo poder de processamento, nas quais são os aspectos propostos pelo modelo. Baseado nestas características e as facilidades em prototipação de projetos, escolheu-se o microprocessador Arduino Uno ¹.

As principais especificações da CPU são clock de 16 MHz, SRAM de 2 KB, 14 portas digitais e programação em linguagem C. O fabricante disponibiliza um ambiente de desenvolvimento denominado Arduino IDE ² na qual simplifica a programação e implantação da aplicação no microprocessador.

5.1.3.2 Algoritmo

Nesta seção é descrito o desenvolvimento do algoritmo implantado no microprocessador Arduino, cujo objetivo é efetuar as leituras do módulo de entrada e encaminhar os resultados ao de saída, de forma eficiente no que tange a questão energética. A programação realizou-se em linguagem C através da ferramenta Arduino IDE. Planejou-se ao iniciar a implementação do algoritmo, otimizações que ofertasse ao máximo a diminuição do consumo energético, visando a eficiência. Para atingir o objetivo, fez necessários diversos testes a fim de verificar quais rotinas despendem mais energia. Verificou-se que as leituras feitas pelos sensores em períodos curtos, como por exemplo 1 segundo, decorreria grande processamento por parte da CPU, e por consequência acarretaria maior gasto. Desta forma, pensou-se em diminuir a frequência de consulta de alguns sensores, como método de reduzir o custo. Outro ponto percebido de alta consumação enérgica, surgiu por parte do módulo de saída, na qual nesta implementação realiza-se por meio de um leitor/gravador SD Card (camada de detecção -> módulo de saída -> componente local).

Baseado nestas apurações, tomou-se as seguintes abordagens para contornar o alto consumo energético; variáveis nas quais seus valores não tendem a sofrer alterações constantes, podem dispor de consultas efetuadas em períodos maiores que 1s. e, as gravações dos registros feito pelo SD Card, devem ser praticadas em longos intervalos. Assim, entendeu-se que os dados de gás, temperatura, umidade e peso enquadram-se nesta nova política, apoiado pelos argumentos descritos na sequência.

A carga das colmeias não pendem a sofrer mudanças substanciais em espaço curto de tempo, ou seja, leituras contínuas apresentaram resultados parecidos ou iguais. Percebe-se que as regiões onde encontram-se os apiários para esta pesquisa, constam com temperatura e umidade estáveis, não havendo mudanças abruptas, o que também acarreta mínimas variações em seus

¹<https://store.arduino.cc/products/arduino-uno-rev3>

²<https://www.arduino.cc/en/software>

valores. Em relação ao gás, optou pela abordagem, dado que o sensor usado precisa de um tempo para identificar alterações no ambiente que o encontra. Por fim, implementou-se uma lógica na qual os dados oriundos da caixa são armazenados em memória, até que a rotina de escrita do SD Card seja executada.

A fim de um amplo entendimento, é disposto um pseudocódigo do algoritmo, conforme visualizado na sequência. O código fonte completo da aplicação está armazenado no repositório do GitHub, através do link ³.

INICIO

INTEIRO tempo

SE tempo == eUmMinuto
 leiaSensorGas ();

FIM SE

SE tempo == eCincoMinuto
 leiaSensorTemperaturaUmidade ()
 leiaSensorPeso ()
 escreverModuloSaida ()

FIM SE

leiaSensorRuido ()
 leiaSensorAnalizadorAtividadeEnxame ()

FIM

A leitura do sensor de gás é realizada a cada minuto. Temperatura, umidade, peso e escrita dos dados são executadas em períodos de 5 minutos. As informações de ruído e analisador de atividade do enxame, exigem monitoramento em tempo real, o que faz com que as leituras sejam processadas em espaços de 100ms. Tendo em vista que as gravações são feitas em tempos de 5 minutos e, existem consultas com período inferior, implementou-se a lógica de substituir os valores antigos pelos novos, caso este seja maior que o anterior. Finalmente, ao escrever os dados no cartão, algumas variáveis são reinicializadas. O analisador de atividade do enxame é implementado como um contador, na qual é incrementado ao passar um objeto na frente do sensor de obstáculo infravermelho KY-032.

Por fim, seguindo e priorizando a eficiência energética, adotou-se a política de remoção de qualquer tipo de led luminoso, que esteja fixado na carcaça dos sensores. Não se percebeu benefícios nas indicações destes, e por menor que seja seu consumo, é um gasto que pode-

³https://github.com/mbsoaresinfor/mestrado/blob/master/codigo_fonte/modulo_IoT-leituraVariaveisColmeia/modulo_IoT-leituraVariaveisColmeia.ino

se evitar. Excluiu-se desta forma, um total de 4 leds. Apenas o microprocessador Arduino e regulador de tensão, não realizou-se tal procedimento, pois através destes emissores de luz, consegue-se visualizar seus estados (ligado/desligado).

5.1.4 Módulo Saída

Conforme discutido em seções anteriores, houve grandes dificuldades em obter sinais de redes móveis na região onde localizava o apiário, o que impossibilitou o planejamento e desenvolvimento do componente remoto do módulo de saída, na qual viabilizaria a transmissão dos pacotes para o segmento de comunicação por meio de uma placa expansora de transmissão WI-FI.

Tendo em vista que o modelo propõe uma alternativa para situações de não existência de conectividade, nominada de DDS (Detecção-Direto-Serviço), a implementação fez uso desta possibilidade, na qual denomina-se de componente local 4.2.3. Conforme proposta inicial, empregou-se um módulo leitor/gravador SD Card, com capacidade de 8GB para persistência dos elementos captados na colmeia. A conexão foi feita junto a placa protoboard, sendo exposta na Figura 20 item 6.

O módulo de saída ao receber os dados oriundos dos sensores, concatena-os com a data e hora em uma string, sendo separados por ponto e vírgula, conforme ilustra abaixo:

```
[dataHora;contadorAnalizadorAtividadeEnxame;gás;temperatura;umidade;som;peso]
```

Após formatado, é executada a rotina de persistência, inserindo-se a string em uma linha do arquivo. De acordo com a seção 5.1.3.2 o módulo de saída é processado a cada 5 minutos, o que produz 12 registros (linhas) por hora. Por fim, para efeitos de segurança, as coletas são inseridas no final do arquivo, não havendo nenhuma perda dos dados já coletados.

5.1.5 Módulo Alimentação

A fim de atender as premissas do modelo, no que diz respeito a ser sustentável e com eficiência energética, enfatizando neste sentido na camada de detecção, esta subseção apresenta os itens utilizados para auxiliar este fim. Segundo o autor Barros e Gedra (2015) a *eficiência energética* é o uso de modo eficiente da energia para se obter um determinado resultado e energia sustentável é toda energia elétrica originada a partir de fontes renováveis que não geram impactos ao meio ambiente. Desta forma, tentou-se diminuir ao máximo o consumo energético por parte da camada de detecção, pois ao alcançar esse objetivo, consegue-se reduzir a quantidade de equipamentos usados, diminui-se a despesa energética, e do ponto de vista da sustentabilidade, se tem uma fonte de energia renovável, limpa e contribuindo na diminuição da poluição e espalhamento dos gases de efeito estufa que afete a nossa atmosfera. Para a atender este pro-

pósito fez necessário a compra de baterias, placa solar e demais componentes, nas quais são detalhadas nas próximas subseções.

5.1.5.1 Baterias

As baterias são as responsáveis por energizar os equipamentos eletrônicos, que compõem a camada de detecção. Com esse propósito, inseriu-se duas baterias de moto, na qual cada uma dispõe de capacidade de 7(Ah), totalizando um total de 14 (Ah) ao sistema. As mesmas são visualizadas na Figura 20 itens 4 e 5. As características reportadas pelo fabricante são: bateria chumbo ácida regulada por válvula, carga em tensão constante (25 C), tensão nominal 12V, flutuação 13.5 a 13.8V, cíclico 14.1 a 14.4V e corrente máxima de recarga 1.02 (A). Mais detalhes na página⁴ do produto.

O Arduino trabalha numa faixa de tensão entre 5 a 12 V e as baterias podem dispor de tensões superiores ao suportado pelo microprocessador. Para evitar a queima do mesmo, inseriu-se um regulador de tensão, na qual foi configurado na faixa 10-12 V. Pode-se visualizá-lo na Figura 20 item 3.

5.1.5.2 Placa Solar

Conforme proposto pelo modelo, é usado energia solar, com sistema off-grid (quando não se tem acesso a uma distribuidora de energia elétrica). A partir de um painel fotovoltaico pretende-se recarregar as baterias de forma sustentável e possibilitar aos dispositivos carga suficiente para seu funcionamento. A placa selecionada dispõe de 30 W de potência, na qual oferece 1,62 (Ah), em sua total potência. Valor fornecido em tempo com sol radiante, sendo que em dias nublados, este número tende a cair. Baseado nas informações levantadas dos fabricantes, referente ao consumo do processador/sensores e somada as leituras preliminares do gasto energético de toda a aplicação, acredita-se que com esses parâmetros seja suficiente para atender o projeto por um determinado tempo de sua execução, mas somente após as execuções em campo, poderá ser possível averiguar o quão performático foi, visto que há outras variáveis nesta conta, tais como, quantidade de irradiação solar e temas abordados na subseção 5.1.3.2. Detalhes das características do painel é disposto no link link⁵ e sua apresentação dentro do protótipo na Figura 19 item 3.

Por fim, ao usar painéis solares é indispensável o uso do controlador de carga, na qual responsabiliza-se em controlar a energia fornecida pelas placas, enviando-as para as baterias. O modelo utilizado foi o controlador de carga solar fotovoltaico Pwm 30a Lcd 30A, visualizado na Figura 20 item 1. O equipamento dispõe de uma configuração de fábrica, que atende esta pesquisa. Baseada nesta estrutura elaborada, responde-se ao questionamento *F - Quais*

⁴<https://unipower.com.br/produto/bateria-12v-4ah-up12alarme/>

⁵<https://www.neosolar.com.br/loja/painel-solar-fotovoltaico-30w-resun-rsm030-p.html>

características importantes da placa solar a ser introduzida no projeto ? da listagem 5.1.1.

5.1.6 Pré-Validação do Protótipo

Com o protótipo finalizado, este estudo optou em realizar inicialmente uma pré-validação, a fim de verificar o respectivo comportamento em campo, de modo que trouxesse resultados que possibilitasse realizar melhorias em seu funcionamento. Desta forma, na data de 09/10/2021 no apiário localizado na cidade de Encruzilhada do Sul/RS - Brasil, pelas coordenadas (Latitude: -30.81065981325515 - Longitude: -52.49520522662678), realizou-se o primeiro teste.

Primeiramente inseriu-se os sensores de temperatura, umidade, gás e ruído dentro da caixa. O analisador de atividade do enxame colocou-se na entrada, a balança debaixo da colmeia, o gabinete na qual concentra-se as baterias e o microprocessador situou-se em cima da colmeia, protegida pelo telhado da mesma. As fiações foram postas de modo que ficassem seguras e não atrapalhasse as abelhas. Por fim, a placa solar instalou-se há uma distância de 2 metros em relação ao sistema, em uma posição que pudesse captar o máximo de irradiação solar. Na Figura 22 é possível visualizar este cenário.

A aplicação deu-se início no dia 09/10/2021 às 14:10 e encerrou-se na tarde seguinte às 14:00. Após remover sensores, caixa de proteção e placa solar, verificou-se alguns pontos, conforme abaixo:

- As abelhas começaram a "roer" e propolizar as proteções dos sensores.
- As baterias estavam em 9V, o que impactava o funcionamento do sistema.
- Excesso de abelhas na entrada da colmeia.

A partir destas observações, necessitou-se entender os motivos e implantar e executar um plano de ação, com objetivo de sanar estes empecilhos. Primeiramente, a questão das abelhas "roerem" as proteções dos sensores e em seguida iniciar a propolisação (informação levantada pelos apicultores), vem do fato que esses insetos, tendem a propolizar em situações ao qual encontram qualquer corpo estranho dentro da caixa, frestas abertas impedindo a entrada de frio ou de inimigos naturais (ABELHAS, 2021d). Isto acarretou a perda do sensor de temperatura e umidade. Desta forma, pensou-se para resolver este problema, inserir proteções mais resistentes, a fim de não danificarem o equipamento.

Em relação às baterias que estavam em baixa voltagem, diagnosticou-se que na data do teste, o tempo estava nublado e pouca irradiação solar, e desta forma, a placa solar não oferecia energia suficiente para alimentar o consumo total do sistema e carregar as baterias simultaneamente. Outro ponto encontrado foi a configuração do regulador de tensão, na qual estava configurado na faixa de 10-12V, devido ao microprocessador Arduino poder trabalhar no meio de 5-12V, sendo assim, como as baterias estavam com 9V, o regulador parou de enviar energia e consecutivamente o sistema parou de funcionar. Para solucionar este cenário, primeiramente

Figura 22 – Pré-Validação do Protótipo



Fonte: Elaborado pelo autor. (2021)

diminuiu-se a tensão de funcionamento do regulador, variando entre 7.5-12V, o que faz com que a aplicação funcione normalmente neste range. Isto aumenta o tempo de execução para mais horas, mas não impede do sistema parar completamente, caso persista mais dias nublados. Para resolver este ponto, seria necessário uma bateria de maior amperagem e placa solar com também maior potencial, a fim de energizar períodos com pouca luz solar. Esta solução foi descartada por questão de custo. Para mais detalhes, é apresentado na subseção 6.1.1 os resultados de consumo energético, tempo de carregamento e funcionamento.

Referente ao excesso das abelhas na entrada do alvado, verificou-se que é pelo fato de ter havido a obstrução de uma das duas entradas da colmeia, conforme discutido na subseção 5.1.2.1. Pensa-se em alterar futuramente o projeto de entrada, colocando dois sensores em posições que não afetem as duas entradas da caixa, mas isto impactaria no consumo energético.

Por fim, por meio deste primeiro teste, conseguiu-se ter alguns resultados das leituras, nas quais foram armazenados em arquivo através do SD Card (camada de detecção -> módulo de saída -> componente local) - (4.2.3). Percebeu-se o quão é o fluxo de abelhas, a temperatura, o erro na leitura da umidade, devido as abelhas roerem o sensor e por fim, a falha operacional por

parte deste pesquisador, que não posicionou corretamente a balança, ocasionando a não leitura desta variável. Neste link ⁶ é possível visualizar o dataset gerado.

5.2 Camada de Serviço

A camada de serviço é constituída por quatro componentes que objetiva executar, analisar, persistir e disponibilizar as informações ao usuário. Dispõem de autonomia e independência em relação à plataforma SaaS. Baseado nestas características, propõem-se a desenvolver os componentes Back-End, Banco de Dados e Aprendizagem de Máquinas com as funcionalidades propostas pelo modelo BeeBr, tais como, o suporte a arquitetura multitenancy.

Por fim, dado ao cenário de prototipação, os testes e simulações das aplicações serão executadas na máquina deste pesquisador, não fazendo uso de serviços em nuvem, autenticação das credenciais e geração de token para as chamadas API's.

5.2.1 Desenvolvimento Componente - Back-End

Realizou-se a implementação do back-end por meio do framework Spring Boot⁷, na qual é desenvolvido em linguagem de programação Java. Optou-se por sua escolha por trazer facilidades no desenvolvimento de aplicações que necessitam dispor de API's Web.

Segundo o modelo BeeBr, o back-end disponibiliza interfaces Rest, a fim de constituir um canal de entrada e saída de dados, que proporciona a transferência das variáveis coletadas pela camada de detecção, somada a comunicação junto ao front-end. Neste sentido, esta implementação introduz o desenvolvimento de oito endpoint Rest que objetiva a salvar as informações dos dados monitorados nas colmeias, cadastros de apicultor, apiário e colmeia, além das buscas destes respectivos valores. Na Figura 23 é possível visualizá-los por meio da ferramenta Swagger ⁸, que dispõe de uma interface gráfica amigável, que proporciona a apresentação dos endpoint implementados nas aplicações.

As quatro operações de persistência dos dados foram realizadas em cima do método POST do protocolo HTTP. As demais, realizou-se por meio do método GET. No corpo das mensagens experimentou-se o formato JSON (JavaScript Object Notation), nas quais são enviados os campos para cadastro ou busca das entidades.

Todas as requisições devem enviar a propriedade *idTenant*, na qual identifica o cliente que pertence o pedido. Por meio desta variável o componente oferece suporte ao multitenancy (mais detalhes na subseção 5.2.2). A Figura 24 ilustra o payload do endpoint responsável por salvar os dados monitorados na colmeia.

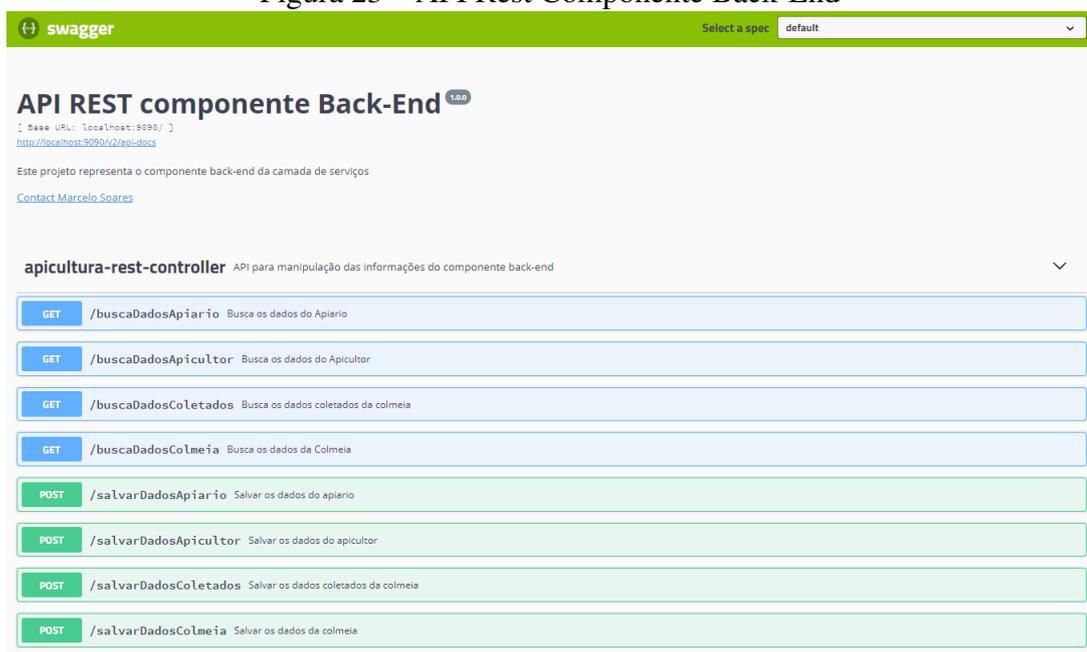
Percebe-se que no payload em formato JSON, são transmitidos as variáveis coletadas do apiário (analisador atividade enxame, gás, temperatura, umidade, som e peso), somado ao iden-

⁶https://github.com/mbsoaresinfor/mestrado/blob/master/dataset/abelhas/10_10_2021-1/DATASET.TXT

⁷<https://spring.io/projects/spring-boot>

⁸<https://swagger.io/>

Figura 23 – API Rest Componente Back-End



Fonte: Elaborado pelo autor. (2021)

Figura 24 – Payload - Salvar os dados coletados da colmeia

```
{
  "contadorAnaliseAtividadeEnxame": 0,
  "gas": 0,
  "idColmeia": 0,
  "idTenant": 0,
  "peso": 0,
  "som": 0,
  "temperatura": 0,
  "umidade": 0
}
```

Fonte: Elaborado pelo autor. (2021)

tificar obrigatório *idTenant* e *idColmeia*, responsável este por reconhecer a colmeia.

A fim de possibilitar maior coesão e menor acoplamento entre classes, desenvolveu-se o back-end a partir de uma arquitetura formada por três camadas, denominadas apresentação, serviço e dados. Ao receber o pedido pela camada de apresentação (API Rest), a aplicação encaminha a camada de serviço que responsabiliza-se em validar os valores. Após, a camada de dados, direciona as operações ao BD, na qual efetua-se à persistência e/ou consulta das informações.

Caso venha surgir a necessidade de inserções de demais campos, o BeeBr deixa aberto a viabilidade, o que faz necessário a inclusão do novo atributo no corpo da mensagem e tratamento

deste, na lógica do back-end.

Por fim, o código fonte do componente back-end pode ser visualizado em sua íntegra neste link ⁹.

5.2.2 Desenvolvimento Componente - Banco de Dados

O banco de dados enquadra-se em um componente do tipo ferramentas que auxiliam como um todo. A escolha pelo paradigma relacional deu-se pela proposta levantada pelo modelo e familiaridade deste pesquisador. Dado ao cenário de prototipação, resultou a preferência pela aplicação h2 ¹⁰, na qual desfruta das características de baixo consumo de memória, embarcado como uma biblioteca nas dependências do projeto back-end e persistência dos dados em memória.

Definiu-se quatro tabelas para armazenar os dados de monitoramento das colmeias, cadastros de apicultor, apiário e colmeia, nas quais, são informações originadas pelos endpoints implementados no back-end. Na Figura 25 é apresentada a estrutura destas tabelas.

O ponto central para a implementação e suporte da arquitetura multitenancy, é a disposição em cada tabela de uma coluna denominada *idTenant*, conforme destacado na figura. A tabela *DadosApicultor* armazena o cadastro dos apicultores e sua chave primária (pk) constitui a coluna *idTenant*. As demais, desfrutam de uma coluna do tipo chave estrangeira (fk), na qual é responsável por armazenar a pk da tabela *DadosApicultor*, ou seja o *idTenant*. Desta forma, qualquer inserção nestas tabelas, deve-se inserir o identificador do apicultor.

Portanto, ao realizar qualquer operação no banco de dados, faz necessário o emprego do valor *idTenant*, a fim de distinguir o usuário, na qual pertence o registro na tabela. Por este motivo, as requisições por parte do componente back-end, devem ser enviadas com a propriedade *idTenant*.

5.3 Elaboração e treinamento do novo modelo de aprendizado de máquina

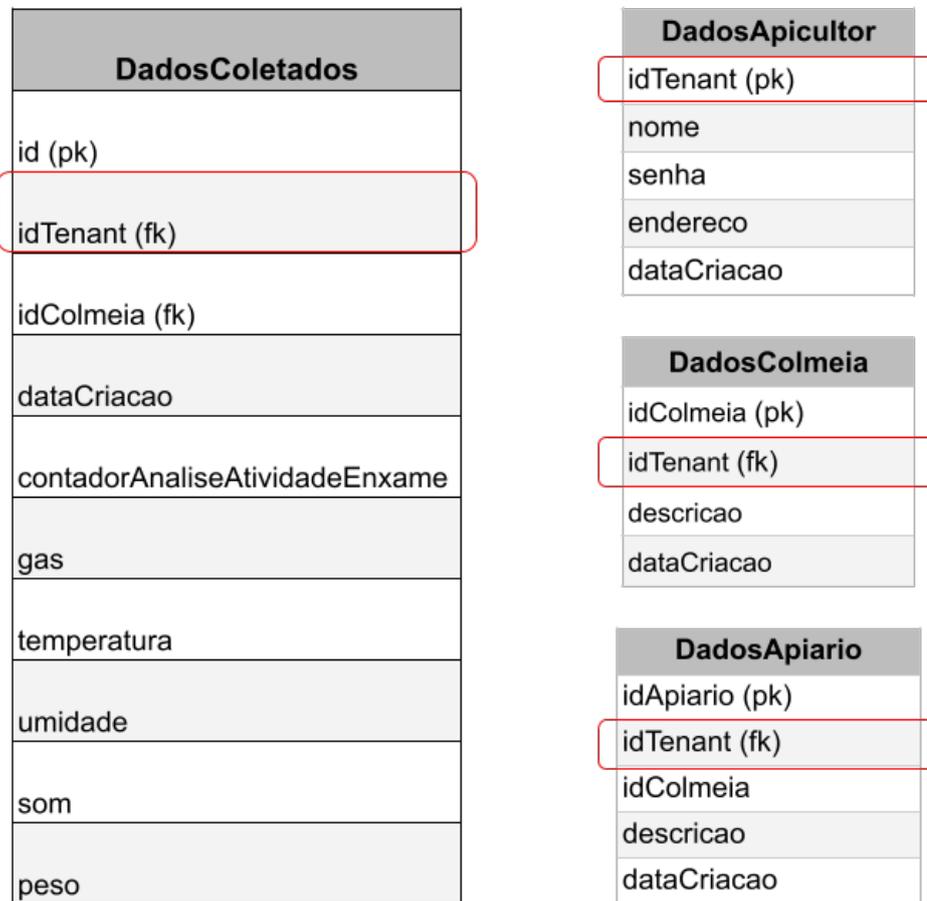
A partir do exposto em seção 4.5 é apresentado na sequência as etapas de modelagem e treinamento do novo modelo de aprendizado de máquina, na qual é responsável pela predição de problemas da enxameação. Para sua implementação utilizou-se linguagem de programação Python e bibliotecas que oferecem suporte aos algoritmos de AM. Baseada nas fases iniciais do microprocessamento, inicia-se o fluxo, conforme exposta na Figura 26.

Nas subseções seguintes são discutidos os estágios que compõem a elaboração e treinamento do novo modelo de aprendizado de máquina.

⁹https://github.com/mbsoaresinfor/mestrado/tree/master/codigo_fonte/componente-back-end

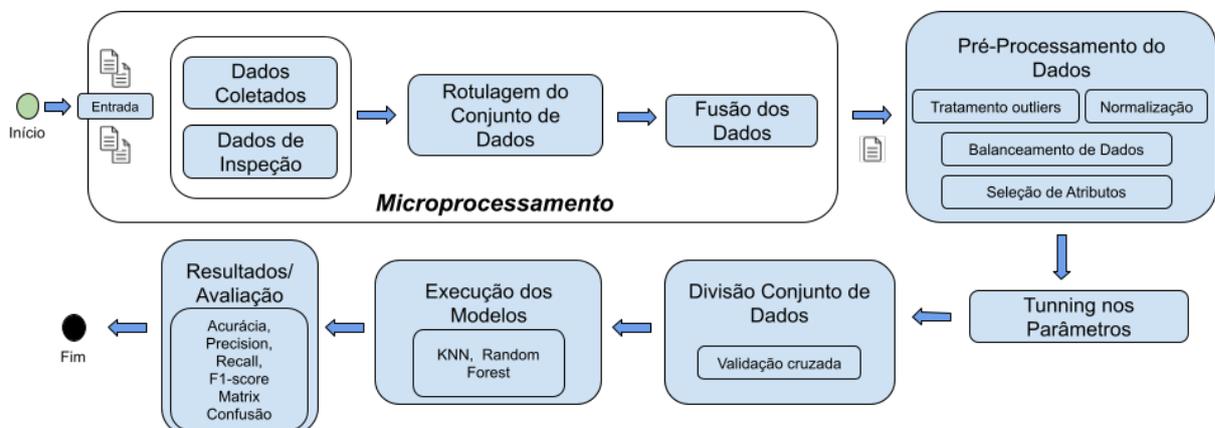
¹⁰<https://www.h2database.com/html/main.html>

Figura 25 – Visão geral das tabelas



Fonte: Elaborado pelo autor. (2021)

Figura 26 – Fases de criação e treinamento do novo modelo de aprendizado de máquina



Fonte: Elaborado pelo autor. (2021)

5.3.0.1 Dados Coletados

Os dados coletados são provenientes dos elementos recebidos pela camada de detecção oriundos da colmeia, discutida nas seções 5.1 e 4.4.6 do modelo BeeBr; sendo o primeiro tipo de dataset de entrada, objetivando na construção dos atributos do novo modelo. Realizou-se leituras das variáveis de temperatura, umidade, gás, ruído, peso e analisador atividade enxame (AAE) em um apiário que encontra-se localizado na cidade de Encruzilhada do Sul/RS pelas coordenadas (Latitude: -30.81065981325515 - Longitude: -52.49520522662678), onde situa-se à uma distância de 250 Km de Porto Alegre/RS (mesmo apiário da Pré-Validação do Protótipo). Conforme Figura 27 é apresentado uma visão ampla da área onde realizou-se as coletas.

Figura 27 – Coletas feitas no Apiário



Fonte: Elaborado pelo autor. (2021)

O apiário em questão é formado por 96 caixas de abelhas. Percebe-se que o gabinete (camada de detecção) encontra-se na parte superior da colmeia. Efetuou-se 2 coletas de dados nos meses de Outubro e Novembro de 2021, nas quais a primeira denomina-se *Coleta ID 1* e a segunda de *Coleta ID 2*. Por meio destas coletas, levantou-se 3x mais amostras na ID 1 em relação a ID 2 (mais detalhes na seção Resultados 6.1.2).

5.3.0.2 Dados de Inspeção

O segundo tipo de dataset de entrada proposto pelo BeeBr, objetiva trazer informações sobre o comportamento das abelhas através de um olhar humano; realizou-se inspeções presenciais

na caixa vigiada em ciclos de 3 dias, sendo no mesmo período das coletas de dados, citada em seção anterior 5.3.0.1. Por este motivo, as amostras de inspeção feitas no período da primeira leitura dos dados coletados, também são chamadas de *Coleta ID 1*, valendo essa regra para a segunda.

As visitas foram feitas pelos próprios responsáveis pelo apiário de maneira neutra e sem interferência por este pesquisador, bem como, algumas perguntas nas quais eles deveriam responder, foram formuladas pelos próprios. Segundo os apicultores, tais indagações propostas destinam-se a saber a saúde da caixa. Assim, ao cumprir as inspeções, os mesmos recebiam um documento, denominado de *Questionário de Inspeção*, na qual deveriam responder 'sim' ou 'não', às perguntas (A,B,C,D,F) e observações climáticas (sol, chuva, nublado) na (E) em cada vistoria, bem como dia e mês. Na sequência são visualizadas as perguntas e no Apêndice C é exposto o documento disposto aos apicultores.

- A - Existem a presença de pragas (traças ou formigas) ?
- B - Dispõem de alimentação suficiente ao redor ?
- C - Há uma circulação fora do comum ?
- D- A caixa está em condição apta ?
- E - Como está a previsão do tempo ?
- F - O sistema está em operação ?

As questões (A,B,C,D) são as elaboradas pelos responsáveis pelo apiário, da mesma forma, para efeito de uso nos algoritmos de aprendizagem de máquina, somente estas farão parte do conjunto de dados de inspeção. No final, usou-se o seguinte *encoder* na transformação da resposta em valor numérico para uso na criação do dataset de entrada; caso o usuário responda 'sim' é atribuído o número '1' e 'não' o conteúdo '0'. Baseado neste levantamento é possível dispor de um entendimento mais amplo da colmeia; entende-se que alguns dos pontos levantados seja mais um insumo de dados, que favoreça na predição da enxameação.

5.3.0.3 Rotulagem do Conjunto de Dados

Adotou-se o procedimento de rotulagem das amostras, a fim de identificar, descrever e caracterizar um aspecto e/ou traço da saúde e bem estar das abelhas, dentro do conjunto de dados coletados. Desta forma, cada coleta em suas respectivas amostras foram dispostas com um rótulo na qual indica o estado da colmeia. Dentre as características que denotam o bem estar das abelhas, cita-se a não existência de pragas, presença da abelha rainha e eventos que possam prejudicar a saúde das abelhas (TECNICO, 2007).

Considerou-se então, que o traço de boa saúde é quando a abelha rainha está presente na caixa e, ruim, quando não a encontra. Esta peculiaridade pode ser vislumbrada na apicultura, quando ocorre a enxameação e morte do inseto. Desse modo, consegue-se caracterizar a enxameação por meio da remoção da abelha mãe. Com o propósito de gerar essa anomalia, necessitou-se retirar a rainha da colmeia. Assim, a Coleta ID 1 dispõem de rótulo *Presença abelha rainha* e a Coleta ID 2 *Sem abelha rainha*, nas quais foram feitas os *encoder* 1 e 0 respectivamente, com intuito de salvar no campo rótulo do dataset. Por tratar-se de uma situação em que prejudica a produção do apiário e impacta a saúde da caixa, o processo realizou-se por um breve período (6 dias); por isso se justifica a menor quantidade de amostras da Coleta ID 2 em relação a ID 1. A fim de um entendimento amplo das nomenclaturas, tipos de dataset(s) e rótulos usados, a Tabela 5 expõe um resumo do que tratou-se até o momento.

Tabela 5 – Rótulos usados nos conjuntos de dados

Nomenclatura	Data	Tipo de DataSet	Rótulo
Coleta ID 1	30/10/2021-13/11/2021	Dados Coletado / Inspeção	Presença abelha rainha (1)
Coleta ID 2	14/11/2021-20/11/2021	Dados Coletado / Inspeção	Sem abelha rainha (0)

Fonte: Elaborado pelo autor. (2021)

5.3.0.4 Fusão dos Dados

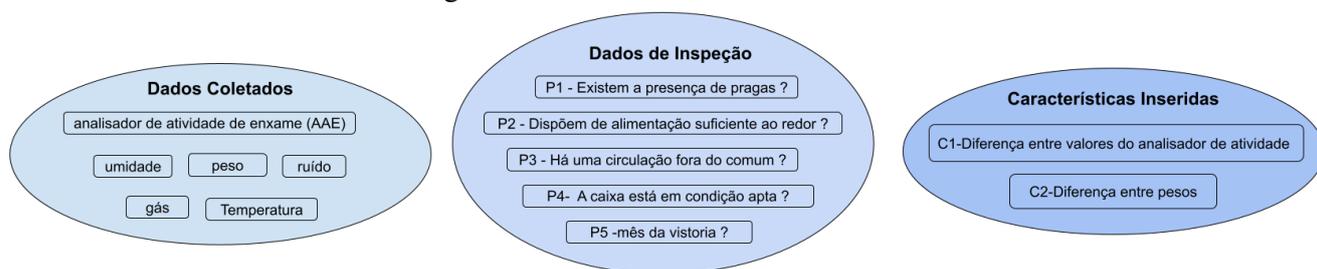
Em virtude de dispor de elementos nas quais estão armazenadas em locais distintos, essa etapa apresenta a metodologia usada para unir as informações dos Dados Coletados em conjunto com as levantadas pelas inspeções realizadas pelos apicultores, denominada de Dados de Inspeção, bem como a remoção e junção de novas características. Como resultado, é concebido um dataset com a máxima quantidade de atributos, a fim de realizar experimentos por meio dos algoritmos de ML. Para atingir este objetivo, faz necessário a seguinte sequência:

- 1 - Concatenação dos dados coletados com as de inspeção.
- 2 - Inserção de novas features e remoções de classes dos dados coletados.

Primeiramente é feita a fusão das informações coletadas com as de inspeção. Como as averiguações ocorriam em janelas de 3 dias e os elementos levantados na colmeia em tempos de 5 minutos, adotou-se o seguinte critério, exposto na Figura 28.

Percebe-se no topo da Figura 28 as características existentes, nas quais devem ser inseridas no dataset. Pelo lado dos dados coletados estão (*AAE, temperatura, umidade, peso, ruído e gás*) e nas inspeções às respostas das perguntas (*A,B,C,D*) do questionário de inspeção que consta no Apêndice C, bem como o mês da vistoria. No exemplo apresentado, visualiza-se as

Figura 28 – Fusão dos Dados



1 - Concatenação dos dados coletados com as de inspeção.

Dia	Amostras	analisador atividade	umidade	temperatura	peso	ruído	gás	P1	P2	P3	P4	P5	
1	1	5	65	23	23,5	480	270	0	0	1	0	11	
	1	329	6	65	23,5	480	265	0	0	1	0	11	
	2	330	8	65	24	23,5	485	270	0	0	1	0	11
2	2	889	60	65	25	22,5	485	270	0	0	1	0	11
	3	890	170	65	24,5	24,5	485	280	0	0	1	0	11
	3	1199	180	65	24	25	480	270	0	0	1	0	11
4	4	1200	230	65	23	23,5	480	270	1	1	1	0	11
	4	1349	240	66	24	22	490	270	1	1	1	0	11
	5	1450	380	67	25	23,5	490	270	1	1	1	0	11
	5	1799	390	66	23	24,3	490	270	1	1	1	0	11
	6	1800	600	65	23,5	23,5	485	270	1	1	1	0	11
	6	1801	650	65	25	23,5	490	270	1	1	1	0	11

2 - Inserção de novas features e remoções de classes dos dados coletados.

Amostras	umidade	temperatura	ruído	gás	P1	P2	P3	P4	P5	C1	C2
329	65	23,5	480	265	0	0	1	0	11	1	0
890	65	24,5	485	280	0	0	1	0	11	2	-1
1800	65	23,5	480	270	1	1	1	0	11	5	1

DataSet final com suas entradas

data/hora	umidade	temperatura	ruído	gás	P1	P2	P3	P4	P5	C1	C2	Rótulo
-----------	---------	-------------	-------	-----	----	----	----	----	----	----	----	--------

Fonte: Elaborado pelo autor. (2021)

amostras, nas quais estão dispostas pelo intervalo das coletas da colmeia. Toda vez que é feita a averiguação da caixa, os valores são planilhados e distribuídos às linhas dos dias (atual, atual-1, atual-2). A tabela T1 na Figura 28 ilustra este cenário, em que é terminado a vistoria no dia 3, e as informações atribuídas aos dias (3,2,1), revista no dia 6, dados concedidos aos dias (6,5,4).

Por fim, a última sequência trata da inserção de novas features e remoções de classes dos dados coletados. Entendeu-se que às características coletadas no apiário (analisador atividade enxame e peso), não trariam respostas significativas aos algoritmos, pois são classes em que seus valores alteram como a forma de um contador numérico. Este aspecto ocasiona uma desproporção na medida de variância em relação aos demais atributos, impactando no processo de generalização dos algoritmos.

A fim de aproveitar o potencial que estes valores expõem, adotou-se a seguinte política: a cada amostra do AAE e peso, seus conteúdos são subtraídos pelo valor da amostra anterior. A partir destes resultados são realizados os seguintes encoder, na qual é visualizado na Tabela 6.

Em que $R = (Ax) - (Ax - 1)$, consiste em:

- R = Resultado

Tabela 6 – Encoder - analisador de enxame e peso

AAE		Peso	
$R = (Ax) - (Ax - 1)$	Valor Atribuído	$R = (Ax) - (Ax - 1)$	Valor Atribuído
≤ 50	1	< 0	-1
≤ 100	2	$= 0$	0
≤ 150	3	> 0	1
≤ 200	4	-	-
≤ 250	5	-	-
≤ 300	6	-	-
≤ 350	7	-	-
≤ 400	8	-	-
≤ 450	9	-	-
> 400	10	-	-

Fonte: Elaborado pelo autor. (2021)

- (Ax) = Amostra atual
- $(Ax - 1)$ = Amostra anterior

A tabela T2 na Figura 28 apresenta o dataset parcial, na qual é removido os campos (AAE e peso) e inseridos C1 e C2 que dispõem deste cálculo.

Por fim, a tabela T3 visualiza-se o dataset final, com as alterações propostas, bem como, a inclusão da data/hora e o campo rótulo. Reforça-se que este conjunto de dados pode sofrer alterações (remoção de atributos), a fim de realizar experimentos por meio dos algoritmos de aprendizado supervisionado.

5.3.0.5 Pré-Processamento dos Dados

O pré-processamento dos dados trata-se de um importante instrumento, na qual bem aplicado, traz bons frutos nos resultados de AM. O processo usado nesta implementação valeu-se das etapas de tratamento de outliers, normalização, balanceamento de dados e seleção de atributos.

O primeiro passo é conhecer e entender a coleção de dados em mãos e conduzir uma análise inicial com as técnicas estatísticas básicas (BRAGA, 2020). A partir deste exame inaugural possibilitou inicialmente verificar os outliers e assim os removê-los.

Na sequência, aplicou-se a normalização, que visa ajustar os valores numéricos do conjunto em uma escala comum (XU; WUNSCH, 2005). Dado que, os atributos coletados em sua grande parte, dispõem de faixas que impactam nos resultados dos algoritmos. Um exemplo deste cenário dá-se à variável *ruído* na qual opera na casa das centenas (460..500), e no oposto, *temperatura* entre (0 .. 45). Diante disso, foi utilizado a técnica *MinMax* em que transforma os valores discrepantes em uma escala de 0,0 à 1,0. Sua fórmula consiste em:

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

- x = valor da classe
- $\min(x)$ = menor valor da classe
- $\max(x)$ = maior valor da classe

Conforme citado no estágio Dados Coletados 5.3.0.1 desta seção, as Coletas ID 1 e 2 gozavam de um total de amostras desproporcionais entre si, ou seja, há mais amostras com rótulos de *Presença abelha rainha* do que *Sem abelha rainha*. A fim de contornar essa situação na qual impactaria nos resultados do novo modelo de aprendizado de máquina, aplicou-se a etapa de balanceamento de dados. Usou-se a técnica estatística SMOTE (Synthetic Minority Over-sampling Technique) para aumentar o número de casos do conjunto de menor quantidade de instâncias, sendo neste cenário a Coleta ID 2. Por fim, selecionou-se os atributos que pudessem trazer os melhores resultados para o novo modelo, a partir da variância da classe (estratégia conhecida por Seleção de Características ou Feature Selection).

5.3.0.6 Tuning nos Parâmetros e Divisão Conjunto de Dados

Os algoritmos de AM dispõem de parâmetros de configuração na qual é possível alterar os seus comportamentos e assim conseguir melhores resultados. Neste sentido, utilizou-se a biblioteca em Python GridSearchCV em que automatizou este processo. Seu trabalho é combinar diversos valores dos parâmetros e encontrar qual melhor adequa-se ao dataset. Após realizada, chegou-se aos seguintes números:

- $KNN = n_neighbors=3, metric="minkowski", p = 1$
- $Random\ Forest = criterion = 'gini', min_samples_leaf = 1, min_samples_split=2, n_estimators = 10$

Na fase de Divisão Conjunto de Dados utilizou-se a técnica k-fold Cross-Validation, também conhecida em português de validação cruzada, que tem como objetivo dividir os dados de treinamento e teste em grupos, conforme discutido na fundamentação teórica 2.2.1. Realizou-se a configuração com $k = 10$.

5.3.0.7 Execução dos Modelos e Resultados/Avaliação

Nesta etapa concentra-se a execução dos algoritmos de AM. Inicialmente, trabalhou com o KNN e Random Forest. O funcionamento destes encontram-se reportados em seção 2.2.1. As implementações dos algoritmos estão disponíveis na biblioteca scikit-learn do Python, nas quais foram executados a partir dos dados introduzidos pela técnica de validação cruzada. Com

base nestas ações são dispostos os resultados amostrais, nas quais foram avaliados por meio das métricas de (1) *Accuracy*, (2) *Precision*, (3) *Recall*, (4) *F1-score* e (5) (5) *Matrix de confusão*.

5.4 Considerações

Dentre os objetivos definidos nesta dissertação, entendeu-se que o baixo custo é um dos itens de suma importância para o sucesso do modelo proposto, dado ao cenário exposto da apicultura brasileira, formada em sua grande parte por pequenas famílias em que a atividade complementa seu sustento. Em face a esta conjuntura, o modelo propôs trabalhar com elementos de hardware e software que buscassem minimizar os gastos financeiros, a fim de viabilizar a arquitetura na apicultura do Brasil.

Ao elaborar a escolha dos equipamentos da camada de detecção, pretendeu-se focar no menor preço, a fim de atingir o foco de baixo custo. Desta forma atingiu-se o valor de R\$ 916,89 na compra de todos os itens para a construção da prova de conceito. Deste valor, mais de 60% é investimento em peças do módulo de alimentação (placa solar, baterias e controle de carga), ficando o restante para demais componentes. No Apêndice E é possível visualizar em detalhes os números.

Ao trazer os valores de custos, este pesquisador objetivou mostrar números que trouxessem uma noção para a implantação do modelo. Sendo assim, não houve esforço em evidenciar tais variáveis de forma precisa, visto que o presente estudo apresenta uma prova de conceito baseado no modelo. Feita estas considerações, a quantia evidencia que o grande custo gerado na camada de detecção advém do módulo de alimentação, disposto de placas solares e baterias. Tais equipamentos dispõem de alto valor de compra, o que consecutivamente aumenta os valores na implantação. Na descrição do modelo, é proposto soluções para contornar esta dificuldade, de acordo com o exposto em subseção 4.2.4, na qual propõe-se que o módulo de alimentação seja compartilhado entre os nós sensores (colmeia).

Considera-se que a escolha da CPU, a partir de modelos mais acessíveis do ponto de vista financeiro e específicos no emprego de IoT, tende a assegurar ganhos na redução energética e proporcionar características focadas na internet das coisas. Dentro dos itens da camada de detecção, entende-se que é possível empregar a compra em grande escala destes, o que impactaria na redução dos custos.

Por fim, conforme exposto em subseção 5 não houve implementação na camada de comunicação, o que impede de mensurar o valor de custo e segundo subseção 5.2, a programação, testes e simulações das aplicações foram executadas em máquina local, não fazendo uso de serviços em nuvem, o que impossibilita medir as despesas.

6 RESULTADOS E DISCUSSÕES

Neste capítulo são apresentados os resultados e discussões obtidas a partir da avaliação dos temas: Sustentabilidade e eficiência energética, Dados coletados na colmeia, Predição do novo modelo de aprendizado de máquina, nas quais englobam as conclusões referentes aos objetivos desta dissertação. Para fins de organização, primeiramente são exibidos os resultados de cada estágio dentro da subseção 6.1 e posteriormente suas discussões em cima destes frutos em anexo a subseção 6.2.

6.1 Resultados

6.1.1 Sustentabilidade e eficiência energética

O planejamento e desenvolvimento de soluções que atendam a área de sustentabilidade e eficiência energética são elementos que objetivaram este estudo. Diante disto, esta seção expõe os resultados referente a área energética do modelo BeeBr, por parte da camada de detecção. Para fins de disposição, dividiu-se em dois temas este tópico: eficiência e sustentabilidade energética.

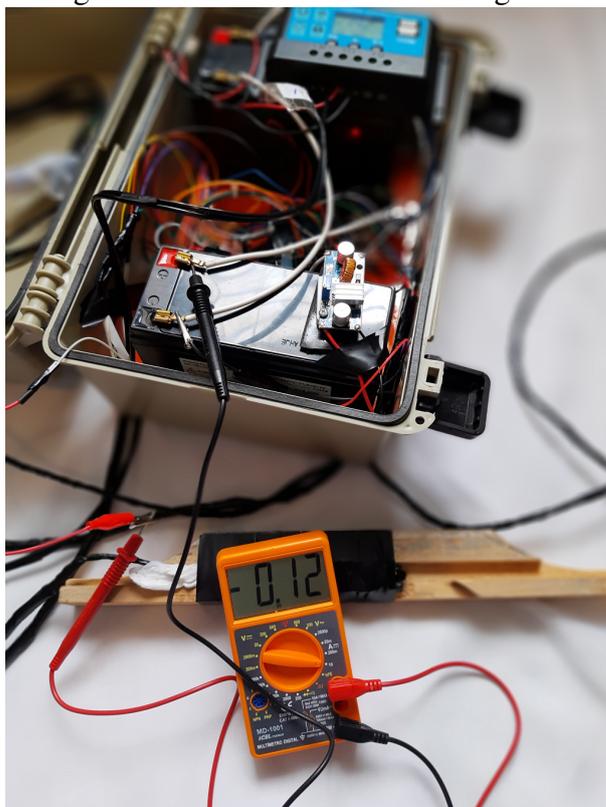
Com intuito de verificar de fato o quão houve de eficácia energética no projeto, calculou-se inicialmente a soma total gasta pelo conjunto de itens que formam a camada de detecção, sendo elas: aplicação embarcada, microprocessador, sensores e regulador de tensão; na qual este valor surge a partir do resultado de soluções desenvolvidas no modelo BeeBr, discutida na subseção 5.1.3.2. Posteriormente descobre-se o quão é utilizada energeticamente sem as soluções sugeridas. Dentro desta etapa, o algoritmo implementado a fim de comparação, não desenvolveu otimizações que objetivaram economizar energia, valendo-se de leituras dos sensores a cada 1 segundo. Tal implementação encontra-se neste link ¹

por meio de um repositório do GitHub. As leituras foram feitas através de um multímetro, conforme visualizado na Figura 29 e os resultados são apresentados na Tabela 7.

Percebe-se na Figura 29 que o multímetro está configurado para leitura em amperes (A), sendo que uma das pontas conecta-se ao positivo da bateria e a outra extremidade liga-se ao Arduino. As medições apresentam o consumo do conjunto em miliampere/hora. Cabe mencionar que devido ao multímetro usado para as aferições não constar centésimos na parte decimal, faz necessário multiplicar por 10; desta forma os valores apresentados estarão convertidos. As medições foram executadas em laboratório por um período de 31 horas ininterruptas, sendo que a cada 60 minutos pegava-se o gasto energético e armazenava em um arquivo a fim de computar as informações. Desta forma, obteve-se um tamanho amostral de 31 elementos. É possível

¹https://github.com/mbsoaresinfor/mestrado/blob/master/codigo_fonte/modulo_IoT-leituraVariaveisColmeia_sem_otimizacao/modulo_IoT-leituraVariaveisColmeia_sem_otimizacao.ino

Figura 29 – Leitura Consumo Energético



Fonte: Elaborado pelo autor. (2021)

visualizar as leituras por meio deste link ².

Tabela 7 – Resultados do consumo energético

Com soluções/melhorias desenvolvidas ?	n	$\bar{X} = \sum X_i/n$	$D_P = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n}}$
Sim	31	0.120 (mAh)	0,003738782506
Não	31	0.140 (mAh)	0,004472135955

Fonte: Elaborado pelo autor. (2021)

Os dados na tabela apresentam os seguintes resultados. Com as melhorias implementadas, obteve uma média de 0.120 (mAh) de consumo e um desvio padrão de 0,003738782506. Sem as otimizações, 0.140 (mAh) de gasto médio e desvio padrão de 0,004472135955. A partir destes valores, calcula-se um ganho em eficiência energética de 16.67% com as implementações desenvolvidas.

Em relação a sustentabilidade, na qual valeu-se da energia solar este feito, foram realizadas alterações nas configurações, reportadas a partir dos resultados da pré-validação do protótipo discutido na subseção 5.1.6, obteve-se os seguintes resultados. Durante os 14 dias que esteve em execução o protótipo, a fim de levantar as primeiras amostras das variáveis da colmeia,

²https://github.com/mboaresinfor/mestrado/blob/master/dataset/consumo_energetico/consumo_energetico.txt

verificou-se nas inspeções no apiário que o sistema parou de responder pela falta de energia por duas ocasiões. Neste período houve dias com pouca irradiação solar e muita chuva, conforme consta nas observações dos apicultores, reportadas no questionário de inspeção do Apêndice C. O que impactou na produção energética, tendo em vista que as placas solares não atenderam a demanda de carregar as baterias suficientemente. No segundo levantamento, executou-se em um intervalo de 6 dias, na qual constatou que o sistema travou em uma ocasião, o que fez também a não leitura dos dados da colmeia. Os frutos obtidos pelas leituras feitas pelo conjunto exposto (camada de detecção) do modelo BeeBr, são mostradas na seção seguinte.

6.1.2 Dados coletados na colmeia

Nesta seção é exibido os resultados referente aos dados coletados na colmeia, efetuado por parte do protótipo de hardware desenvolvido na subseção 5.1 e executada na seção 5.3.0.1, bem como, o desempenho dos dispositivos nele constituído. Os valores aqui reportados, advém de objetivos concluídos por esta pesquisa, tais como, o desenvolvimento de protótipo de hardware que coleta às variáveis do apiário; implantação de componentes de softwares que processam as informações levantadas em campo; validações destas implementações junto ao apiário real, assim como a oferta de mecanismos que facilitem o apicultor usufruir das programações realizadas, mesmo não dispondo de comunicação.

Conforme discutido na Pré-Validação do Protótipo em 5.1.6, realizou-se alterações que objetivaram solucionar problemas encontrados na pré-homologação. Após realizadas, foram feitas as coletas dos dados no apiário localizado na cidade de Encruzilhada do Sul/RS, conforme citado na seção 5.3. Foram efetuadas 2 coletas de dados nos meses de Outubro e Novembro de 2021. A primeira amostra consta com um total de 3036 e a segunda 1212 registros. As leituras foram feitas na mesma caixa de abelha. Uma visão geral dos elementos coletados são apresentados na Tabela 8

Tabela 8 – Resultados das coletas na colmeia

Coleta ID	Amostras	Dias	Datas	Rótulo
1	3036	14	30/10/2021-13/11/2021	Presença abelha rainha (1)
2	1212	6	14/11/2021-20/11/2021	Sem abelha rainha (0)

Fonte: Elaborado pelo autor. (2021)

Percebe-se na Tabela 8 que as amostras constam com rótulos, a fim de identificar a característica disposta na coleta, conforme citado na seção 5.3.0.3. A Coleta ID 1 dispõe de um número significativo maior em relação a Coleta ID 2, haja vista, que fez-se necessário a remoção da abelha rainha, a fim de gerar uma anomalia e consecutivamente a enxameação na caixa. Este processo de retirada do inseto é malicioso e prejudicial para a produtividade, desta forma, a última coleta, apresenta menor quantidade de dados.

Segundo a seção Algoritmos 5.1.3.2, cada sensor apresenta um tempo para realizar a leitura,

mas as gravações são feitas em períodos de 5 minutos, o que proporciona um total de 288 amostras por dia. Pelo exposto nos resultados no item Sustentabilidade e eficiência energética 6.1.1, a aplicação parou de responder por falta de energia em alguns momentos, por este motivo, o total das amostras não são proporcionais ao valor da fórmula (Dias x Amostras). Nos links Coleta ID 1 ³ e Coleta ID 2 ⁴ é possível visualizar os dataset coletados.

A Tabela 9 apresenta os dados estatísticos descritivos (desvio padrão e moda) relacionadas às leituras feitas pelos sensores de temperatura, umidade, ruído, gás e peso.

Tabela 9 – Descrições Estatísticas da Coleta ID 1 e ID 2

Sensores	Coleta ID 1		Coleta ID 2	
	Desvio Padrão	Moda	Desvio Padrão	Moda
Temperatura (°C)	6.63	32	4.46	17.1
Umidade (%)	11.64	99.9	10.92	99.9
Ruído	78.94	309	103.05	521
Gás	10.20	519	4.48	521
Peso (Kg)	0.69	48.38	0.23	49.64

Fonte: Elaborado pelo autor. (2021)

Efetuando-se uma análise baseada na comparação dos dados, referente a cada elemento coletado, oriundos das *Coleta ID 1* e *Coleta ID 2*, apresentado na Tabela 9, percebe-se que nos valores de umidade e gás que em suas *moda* situam-se numa mesma faixa numérica. Em relação aos seus *DP* (*Desvio Padrão*), a umidade encontrou-se valores muito próximos, o que não ocorreu com a propriedade gás.

No que refere-se a temperatura, verificou-se o *DP* dados semelhantes, o que não aconteceu nas *moda* nais quais constam números discrepantes. Tratando-se da variável ruído, os dados apresentam valores diferenciados em cada coleta. O campo peso constatou-se *DP* e *moda* equivalentes. Por fim, o sensor de analisador atividade enxame (AAE), realiza uma contagem em cima das abelhas, desta forma, entendeu-se que apresentar o *DP* e *moda* não objetivaria dados significantes, visto que o valor é um contador numérico.

Na sequência, são exibidas as Figuras 30, 31 e 32 nais quais apresentam dados estatísticos descritivos (quartil, mediana, mínimo, máximo e outliers) representados em forma de gráficos, oriundos das leituras da *Coleta ID 1* e *Coleta ID 2*.

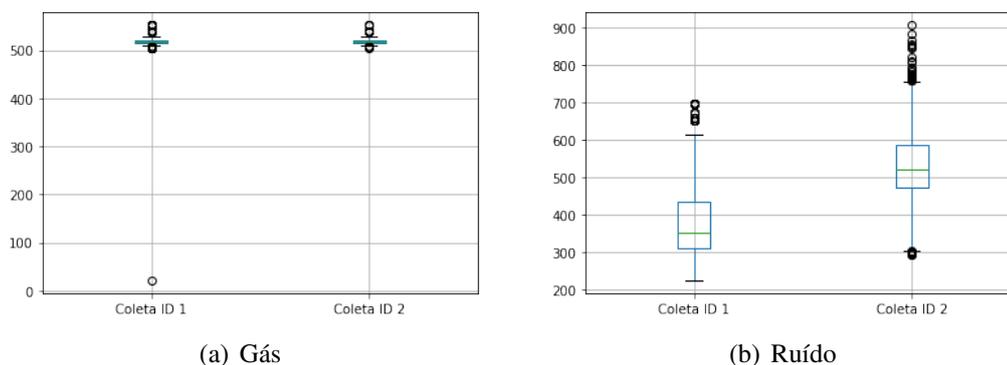
Na Figura 30(a), percebe-se no intervalo interquartilico e na amplitude pouca dispersão e tênue presença de outliers nos dados relacionados aos gases coletados. Na Figura 30(b) constata diferenças nas duas coletas em relação às suas dispersões e amplitudes, acrescido à uma maior quantidade de outliers por parte da Coleta ID 2.

A Figura 31(a) expõe as leituras da umidade, nais quais os gráficos apresentam grande quantidade de outliers e mínimas visualizações do intervalo interquartilico e amplitude. Verifica-se

³https://github.com/mboaresinfor/mestrado/blob/master/dataset/abelhas/1_amostra/DATASET.txt

⁴https://github.com/mboaresinfor/mestrado/blob/master/dataset/abelhas/2_amostra/DATASET.txt

Figura 30 – Dados estatísticos descritivos das variáveis de Gás e Ruído

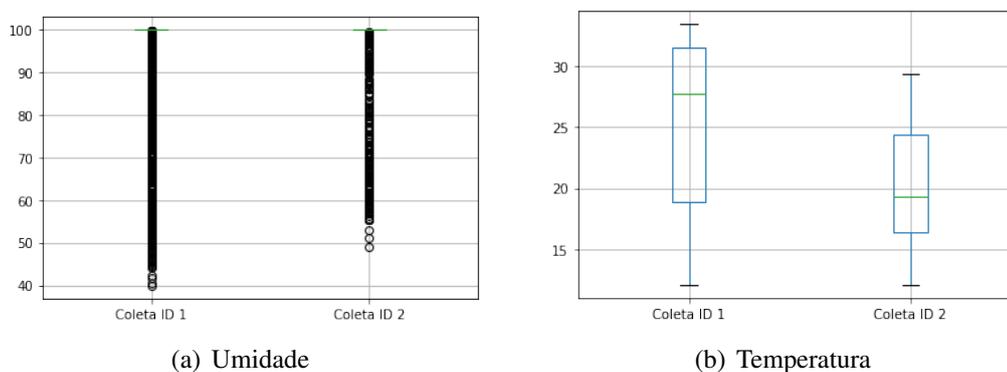


(a) Gás

(b) Ruído

Fonte: Elaborado pelo autor. (2021)

Figura 31 – Dados estatísticos descritivos das variáveis de Umidade e Temperatura



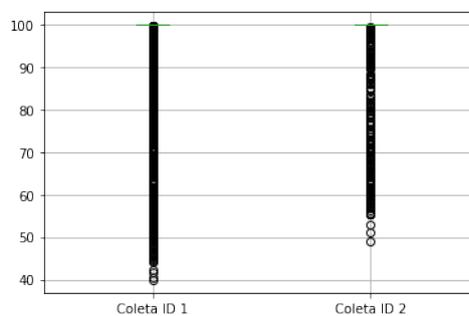
(a) Umidade

(b) Temperatura

Fonte: Elaborado pelo autor. (2021)

na Figura 31(b) relacionada as coletas das temperaturas, uma maior amplitude e dispersão da parte da Coleta ID 1.

Figura 32 – Dados estatísticos descritivos da variável Peso



(a) Peso

Fonte: Elaborado pelo autor. (2021)

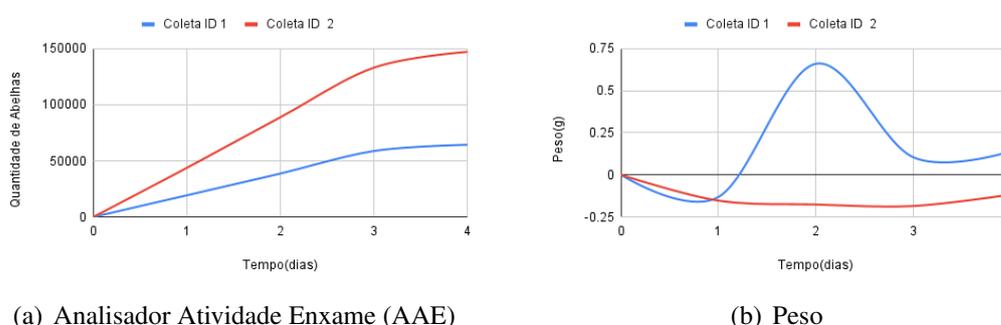
Por fim, o gráfico apresentado pela Figura 32 mostra os pesos das Coletas ID 1 e 2, sendo

que na primeira visualiza superior dispersão e amplitude.

Dando seguimento aos resultados referente às leituras dos sensores, são apresentados no curso, gráficos lineares a partir de séries temporais diárias, que objetivam visualizar mais traços comparativos que denotam características encontradas na Coleta ID 1 e Coleta ID 2. Estes valores tendem elucidar o CdC (Comportamento da Colmeia) quando dispõe de boa ou má saúde, na qual esse aspecto é descrito e rotulado na Tabela 8.

Finalmente, em virtude das coletas gozarem de distintas quantidades de amostras (Coleta ID 1 = 3036 | Coleta ID 2 = 1212), optou-se em apresentar as séries temporais pelo período de 0 à 4 dias, na qual consiste um total de 1000 registros. Sendo assim, é exibido nos gráficos o comportamento da colmeia, baseado nas primeiras 1000 amostras da Coleta ID 1 e 2.

Figura 33 – Comportamento da Colmeia em relação às variáveis AAE e Peso



(a) Analisador Atividade Enxame (AAE)

(b) Peso

Fonte: Elaborado pelo autor. (2021)

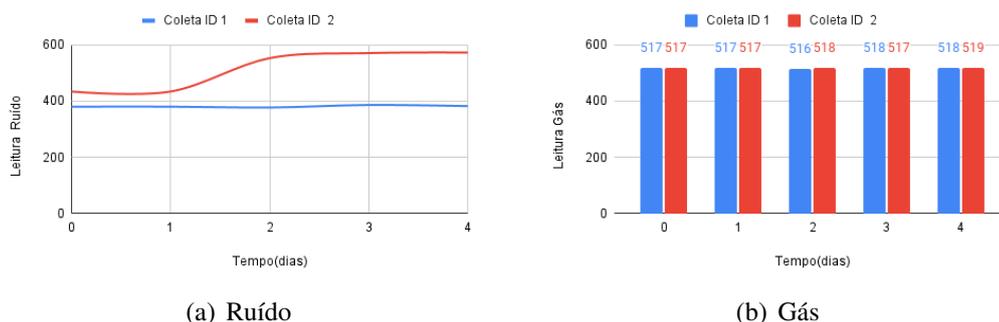
Na Figura 33(a) ilustra o comportamento da colmeia em relação ao sensor infravermelho que efetua as leituras do analisador de atividade do enxame. Percebe-se pela linha destacada em vermelho (Coleta ID 2), que a atuação da colmeia tende a ser mais ativa, agitada quando não há a presença da abelha rainha. Essa expressiva movimentação dá-se desde as primeiras horas da retirada da abelha mãe, na qual a partir deste evento é crescente o número que expressa a atividade da caixa.

O gráfico apresentado na Figura 33(b) destaca o comportamento do enxame em relação ao peso, quando a caixa dispõe de traços de boa saúde (presença abelha rainha) ou alguma anomalia, má saúde (sem abelha rainha). Em linha azul (Coleta ID 1) nota-se variações de aumento e diminuição na produção do mel, mas que na grande parte do período analisado, obteve-se ganho de peso. No oposto, em linha vermelha (Coleta ID 2), verifica-se perda de peso, quando a abelha rainha não encontra-se na colmeia.

O comportamento do enxame em relação às leituras do ruído, pode-se visualizar na Figura 34(a), na qual percebe-se que após a retirada da abelha rainha, houve um aumento do zumbido na caixa; sendo que este som é ampliado após o 1 dia. A Coleta ID 1 apresenta um ruído constante em toda a série temporal analisada.

As leituras dos gases trouxeram comportamentos parecidos, conforme mostra a Figura 34(b). Por fim, variáveis como temperatura e umidade são sensíveis aos aspectos do clima,

Figura 34 – Comportamento da Colmeia em relação às variáveis Ruído e Gás



Fonte: Elaborado pelo autor. (2021)

por este motivo, optou-se em não dispor comparativos de seus comportamentos.

6.1.2.1 Desempenho dos Dispositivos

No tocante ao desempenho dos sensores, compreendeu-se que os equipamentos atuaram de forma aceitável pelo olhar deste pesquisador. Observa-se dentro deste quesito que mesmo diante as proteções mais resistentes ao redor dos sensores (implantadas em face aos resultados da Pré-Validação do Protótipo, descrito na subseção 5.1.6), às abelhas permaneceram a propoliza-los, conforme ilustra Figura 35

Os sensores de infravermelho, temperatura/umidade e ruído identificado pelos números 1, 2 e 3 respectivamente, visualizados na Figura 35, estão protegidos e propolizados.

Referente aos demais dispositivos que formam o protótipo de hardware, percebeu-se que o microprocessador Arduino Uno apresentou performance satisfatória, atendendo ao que foi proposto. Em relação ao SD Card, que compõe o (camada de detecção -> módulo de saída -> componente local) da camada de detecção, também mostrou desempenho adequado, não havendo nenhum problema operacional com tal dispositivo. Os resultados do módulo de alimentação são apresentados na seção 6.1.1. A partir dos dados expostos e avaliados nesta seção, somados com os de inspeção, passa-se a trabalhar nos valores alcançados pelos algoritmos de aprendizagem de máquina, usados no novo modelo de aprendizado de máquina.

6.1.3 Predição dos dados pelo novo modelo de aprendizado de máquina

Nesta seção são exibidos os resultados obtidos pelo novo modelo de aprendizado de máquina apresentado em 5.3. Conforme citado em 5.3.0.7 foram usados os algoritmos KNN e Random Forest como implementação, baseando-se nos dataset(s) dispostos pelos dados coletados/inspeção, provenientes das etapas reportadas em seção 5.3. Para avaliá-los, usou-se as métricas: (1) *Accuracy*, (2) *Precision*, (3) *Recall*, (4) *F1-score* e (5) *Matrix de confusão*.

A fim de buscar os melhores resultados para o novo modelo elaborado, verificou-se os va-

Figura 35 – Sensores Propolizados



Fonte: Elaborado pelo autor. (2021)

lores da variância da classe de cada atributo trabalhado neste estudo, conforme exposto nas Tabelas 10, 11, 12. A partir dos números obtidos, aplicou-se a técnica de seleção das características (*Feature Selection*), resultando em três conjuntos, nominados por este pesquisador de *experimentos*, na qual cada um dispõe de uma coleção de atributos, conforme mostra a Figura 36.

Tabela 10 – Variância dos dados coletados

	Atributos			
	umidade	temperatura	ruído	gás
Valor Variância	122.54	38.48	14517.49	60.30

Fonte: Elaborado pelo autor. (2021)

Tabela 11 – Variância dos dados de inspeção

	Atributos				
	P1	P2	P3	P4	P5
Valor Variância	0	0	0.25	0	0.06

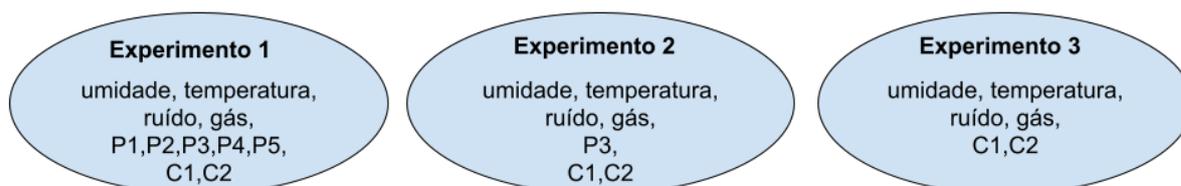
Fonte: Elaborado pelo autor. (2021)

Tabela 12 – Variância das características inseridas

	Atributos	
	C1	C2
Valor Variância	3.02	0.73

Fonte: Elaborado pelo autor. (2021)

Figura 36 – Experimentos a partir da seleção das características



Fonte: Elaborado pelo autor. (2021)

Percebe-se que os *dados coletados* expostos na Tabela 10, apresentam números muito variados e com grande distância da média. Na contramão, observa-se na Tabela 11 que os resultados atingidos pelas variâncias das *classes de inspeção*, dispõem de números iguais ou próximo a zero, demonstrando uma proximidade junto a média. Por fim, as *características inseridas* encontram-se com grandezas diversas em relação à média, conforme mostra a Tabela 12.

Sendo assim, para a concepção dos experimentos, elaborou-se critérios a partir das observações prévias dos resultados atingidos; constatou-se que as escalas (0.25 - 0.50), dispõem e abrangem uma gama de cenários, bem como a junção de todas as características, nas quais assim, possibilita-se visualizar o comportamento dos algoritmos. Na sequência são expostos os experimentos elaborados.

- Experimento 1: todos os atributos coletados por esta pesquisa
- Experimento 2: atributos de variância acima ou igual ao valor 0.25
- Experimento 3: características com variância acima ou igual ao valor 0.50

Por fim, conforme reportado em seção Pré-Processamento dos Dados 5.3.0.5, aplicou-se a técnica SMOTE (Synthetic Minority Over-sampling Technique), a fim de balancear os dados. Após aplicado o procedimento, obteve-se 6072 amostras, das quais, 4554 foram reservadas para o treinamento do modelo e 1518 para os testes.

6.1.3.1 Experimento 1

Nesta subseção são dispostos os resultados obtidos pelo experimento 1, em que avaliou-se a precisão dos algoritmos, levando em consideração todos os atributos, umidade, temperatura,

ruído, gás, P1, P2, P3, P4, P5, C1, C2. A Tabela 13 dispõe das métricas Acurácia, Precision, Recall e F1-score dos modelos usados.

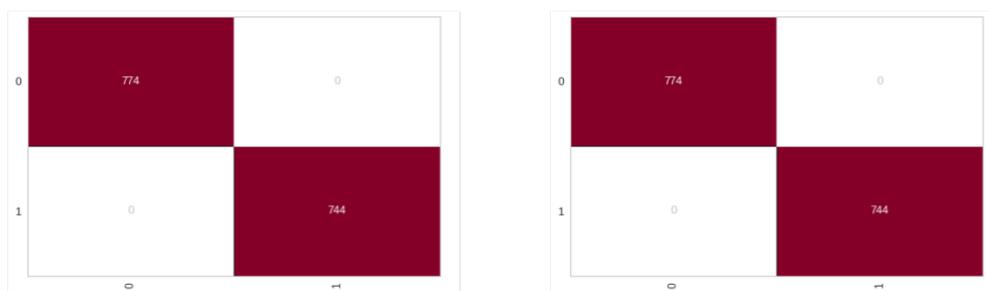
Tabela 13 – Métricas do Experimento 1

	Métricas			
Algoritmo	Acurácia	Precision	Recall	F1-score
KNN	1.00	1.00	1.00	1.00
Random Forest	1.00	1.00	1.00	1.00

Fonte: Elaborado pelo autor. (2021)

Percebe-se que os algoritmos atingiram números idênticos em todas as métricas. Na Figura 37(a) e 37(b) são mostradas as matrizes de confusão dos algoritmos KNN e Random Forest respectivamente.

Figura 37 – Experimento 1: Matrizes de confusão



(a) KNN

(b) Random Forest

Fonte: Elaborado pelo autor. (2021)

Na Figura 37 os valores (1) e (0) representam os rótulos do dataset, na qual (1 = presença da abelha rainha) e (0 = ausência da abelha). Tal legenda estende-se às demais Figuras 38(a), 38(b). Por fim, cabe notar que as 1518 amostras de testes foram previstas pelos modelos.

6.1.3.2 Experimento 2

Nesta subseção são exibidos os resultados obtidos pelo experimento 2, constituída pelas características umidade, temperatura, ruído, gás, P3, C1, C2, nas quais detêm variância acima ou igual ao valor 0.25. A Tabela 14 mostra as métricas dos modelos.

Visualiza-se o mesmo comportamento do Experimento 1, com valores idênticos em todas as métricas trabalhadas. Na Figura 38(a) e 38(b) são mostradas as matrizes de confusão dos algoritmos KNN e Random Forest respectivamente.

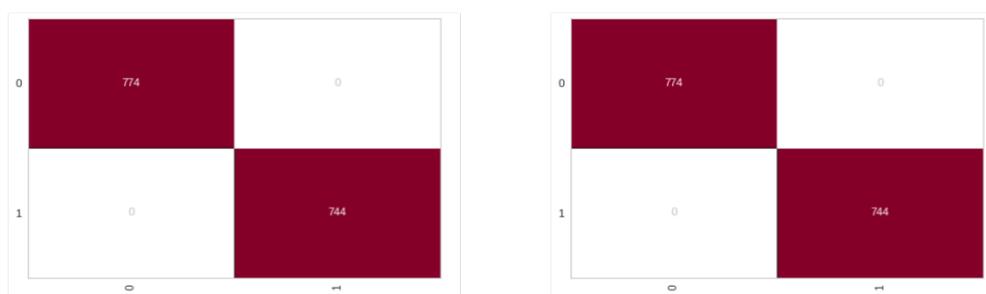
Percebe-se na matriz de confusão que houve o mesmo número de acertos que no Experimento 1.

Tabela 14 – Métricas do Experimento 2

	Métricas			
Algoritmo	Acurácia	Precision	Recall	F1-score
KNN	1.00	1.00	1.00	1.00
Random Forest	1.00	1.00	1.00	1.00

Fonte: Elaborado pelo autor. (2021)

Figura 38 – Experimento 2: Matrizes de confusão



(a) KNN

(b) Random Forest

Fonte: Elaborado pelo autor. (2021)

6.1.3.3 Experimento 3

Por fim, nesta subseção são revelados os resultados obtidos pelo experimento 3, composto pelos atributos umidade, temperatura, ruído, gás, C1, C2, em que dispõem de variância acima ou igual ao valor 0.50. A Tabela 15 apresenta os números das métricas colhidas.

Tabela 15 – Métricas do Experimento 3

	Métricas			
Algoritmo	Acurácia	Precision	Recall	F1-score
KNN	0.93	0.94	0.93	0.93
Random Forest	0.95	0.95	0.95	0.95

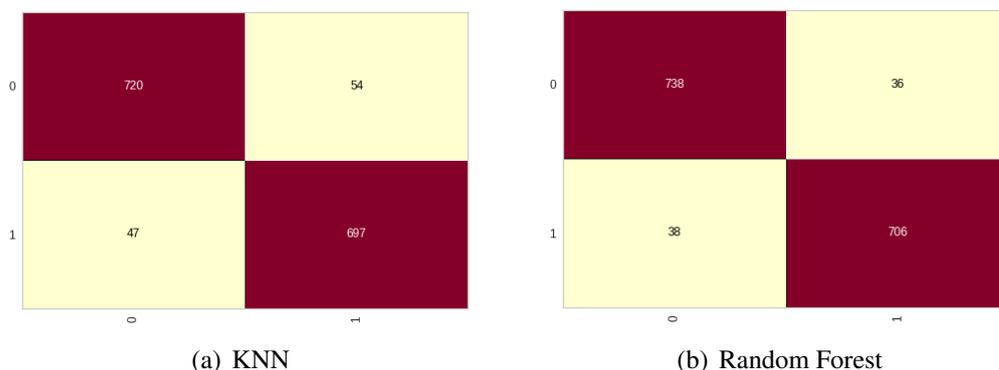
Fonte: Elaborado pelo autor. (2021)

Percebe-se números mais realistas nos resultados das métricas do Experimento 3. Na Figura 39(a) e 39(b) são mostradas as matrizes de confusão dos algoritmos KNN e Random Forest respectivamente.

Visualiza-se na matriz de confusão do Experimento 3 que houve 720 e 697 acertos no algoritmo KNN, rotulados respectivamente por (0 = ausência da abelha rainha) e (1 = presença da abelha rainha). Já para o Random Forest, obteve 738 e 706 acertos nos rótulos citados.

Portanto, baseado nos resultados expostos nas subseções Sustentabilidade e eficiência energética 6.1.1, Dados coletados na colmeia 6.1.2 e Predição dos dados coletados pelo novo modelo

Figura 39 – Experimento 3: Matrizes de confusão



Fonte: Elaborado pelo autor. (2021)

de aprendizado de máquina 6.1.3 são feitas na seção seguinte as suas discussões.

6.2 Discussões

6.2.1 Sustentabilidade e eficiência energética

Dentro do tema eficiência energética, alcançou-se 16.67% de economia energética com as implementações desenvolvidas. A partir deste número consegue-se realizar cálculos que demonstram o quão esta diminuição representa de fato. A Tabela 16, apresenta uma síntese comparativa do ganho na eficácia energética.

Tabela 16 – Comparativo de ganho na eficiência energética

Com soluções/melhorias desenvolvidas ?	1h (mAh)	24h (Ah)	30d (Ah)	Apiário c/ 100 colmeias (Ah)	Apiário c/ 100 colmeias (W)
Sim	0.120	2.88	86.4	8.640	103.680
Não	0.140	3.36	100.8	10.080	120.960

Fonte: Elaborado pelo autor. (2021)

O gasto apresentado na Tabela 16 pelo conjunto otimizado ficou em 0.120 (mAh). Em 24 horas de consumo o valor mantém em 2,88 (Ah) ($consumo\ diário = (mAh) * 24$) e em 30 dias o resultado permanece na faixa de 86,4 (Ah) ($consumo\ mês = consumo\ diário * 30$). Buscando como parâmetro e exemplo o apiário na qual realiza-se esta pesquisa, em que dispõe em torno de 100 caixas de abelhas, o consumo total estaria na faixa de 8.640 (Ah) ($apiário\ c/ 100\ colmeias = consumo\ mês * 100$), e por fim, transformando em Watts, o valor localiza em 103.680, trabalhando em tensão de 12V ($W = apiário\ c/ 100\ colmeias * V$). Em cima desta economia apresentada, consegue-se diminuir a quantidade de elementos energéticos da estrutura, como baterias e placas solares, ou em ambientes que fazem uso de energia cabeada, o custo operacional diminui-se significativamente.

Entende-se que estes valores possam ser aperfeiçoados com alterações no algoritmo 5.1.3.2, valendo-se de bibliotecas de baixo nível das quais não foram utilizadas na implementação, e que objetivam a otimizar o uso da CPU. Outro ponto a aprimorar são o emprego de microprocessadores com ênfase em IoT. Este tipo potencializa cenários em que há escassa energia oferecida. E por fim, pode-se buscar sensores na indústria que disponham de aprimoramentos em seu consumo.

Finalmente concluiu-se alguns pontos em relação à sustentabilidade do projeto. Primeiramente sobre o uso das baterias. Como utilizou-se as usadas em moto, na qual sua construção foca-se em energizar equipamentos, e, necessariamente, deve-se realizar no mesmo período seu carregamento por parte do alternador, fazendo-a sempre estar no estado de carregada. Devido aos dias nublados, as placas solares não dispuseram suprir a demanda do conjunto, o que acarretou um descarregamento rápido, uma característica marcante de baterias automotivas. Para o autor COPETTI e MACAGNAN (2007) faz necessário a aplicação de baterias do tipo estacionária pois com as suas características construtivas fazem com estas baterias possam suportar as condições de ciclagem e, portanto, se comportem melhor que as demais em sistemas fotovoltaicos (energia solar).

No que tange a energia solar, observa-se as seguintes questões. A placa usada disponha de 30W, um número suficiente para atender a demanda do conjunto, levando em consideração seu consumo apresentado. Devido a pouca irradiação solar na região onde encontra-se os apiários, este dispositivo não supriu a exigência necessária. Conforme discutido na subseção 5.1.6 a solução a esse tipo de dificuldade é a inserção de mais baterias e placas solares, o que acarreta impacto no custo. A depender do território onde estão as abelhas, os problemas de pouca luminosidade são menores, como é o caso do nordeste brasileiro. Visualiza-se na Figura 40, a diferença de irradiação solar de dois pontos do Brasil.

Figura 40 – Comparativo Irradiação Solar - Sul x Nordeste - Brasil

Médias do Total Diário da Irradiação Global Horizontal para o Estado do RIO GRANDE DO SUL

(Wh/m².dia)

Mostrar registros

Procurar:

---- Inserir ID ----

ID	Lon	Lat	Anual	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
863	-52,549	-30,5005	4457	6431	5811	4888	3817	2780	2335	2562	3232	3723	4938	6221	6748

Médias do Total Diário da Irradiação Global Horizontal para o Estado da BAHIA

(Wh/m².dia)

Mostrar registros

Procurar:

---- Inserir ID ----

ID	Lon	Lat	Anual	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
31217	-42,049	-10,9005	5876	6372	6209	5977	5356	4977	4852	5074	5870	6507	6495	6383	6436

Fonte: Elaborado pelo autor. (2021)

Na Figura 40 retirada do site Laboratório de Modelagem e Estudos de Recursos Renováveis de Energia - (PEREIRA et al., 2017)⁵ é possível encontrar a irradiação solar de qualquer ponto do Brasil. A fim de comparação buscou-se a cidade de Encruzilhada do Sul/RS (ID= 863) e Jussara/BA (ID= 31217) situada no nordeste, região na qual está em 2^o posição na produção de mel no Brasil (ABELHAS, 2021e). Percebe-se que na imensa parte do ano que o território nordestino, dispõe mais luz solar que a região Sul, o que melhora o desempenho das placas solares. Em cima destes conhecimentos e bem como aos resultados obtidos, pode-se concluir os seguintes pontos sobre energia solar. Sua viabilidade dentro modelo é válido, desde que seja feito a aquisição de mais equipamentos, nas localidades com menor irradiação solar, o que por consequência torna nestas áreas seu custo maior em relação às demais.

6.2.2 Dados coletados na colmeia

A partir dos resultados obtidos nas leituras realizadas pelos sensores e apresentados na subseção 6.1.2, percebeu-se que houve alterações significativas no comportamento da colmeia, quando a mesma constava de má saúde; sendo que este estado foi por consequência da retirada da abelha rainha, na qual é identificado no decorrer da subseção Resultados - Dados coletados na colmeia 6.1.2 por Coleta ID 2.

Destaca-se, expressiva movimentação no enxame, diminuição na produção de mel e aumento do ruído na caixa nas primeiras horas da remoção do inseto. Em contrapartida, não captou-se tais hábitos com a presença da abelha mãe, conforme gráficos comparativos dispostos nas Imagens 33(a), 33(b) e 34(a). Baseada nas observações feitas em cima dos resultados apresentados, nota-se por parte deste pesquisador, possíveis correlações entre as variáveis de atividade do enxame com as de peso e zumbido da caixa, ou seja, o fluxo na colmeia, influência o peso e ruído.

Diante destas percepções, é realizada inicialmente uma análise de correlação linear entre as variáveis quantitativas (atividade enxame (X), peso (Y)) e (atividade enxame (X), ruído (Y)), na qual objetiva a verificar e entender as viáveis relações entre (X,Y), nas duas coletas de dados existentes. Os exames foram baseado em 152 amostras de cada Coleta (ID 1 e 2), sendo que, de forma aleatória, buscou-se os registros 158 à 309 para a associação (atividade enxame (X), peso (Y)) e 1 à 152 para relação (atividade enxame (X), ruído (Y)).

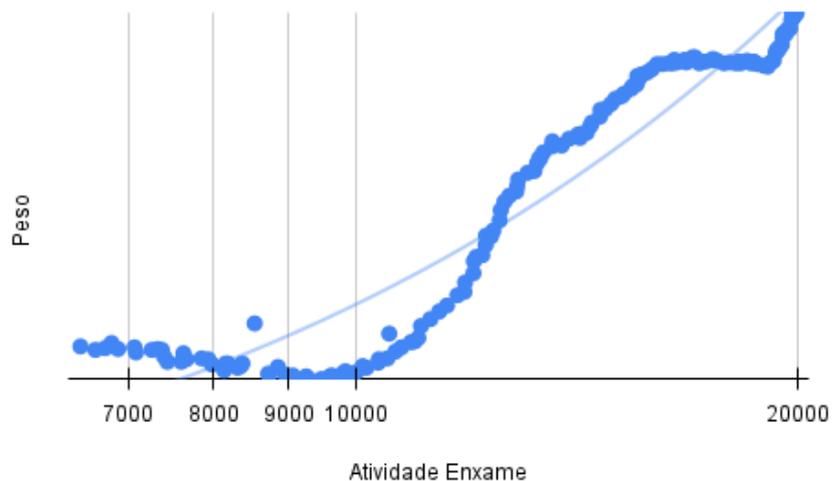
Primeiramente, é apresentado nas imagens seguintes os gráficos de dispersão das variáveis quantitativas (Atividade Enxame X Peso) e (Atividade Enxame X Ruído) a fim de ter uma melhor ideia das associações entre (X,Y).

Percebe-se no gráfico da Figura 41 uma tendência positiva, sendo que na Figura 42 dispõe de negativa, entretanto, nos gráficos da Coleta ID 2, Figura 43 desfruta de tendência negativa e na subsequência, 44 ocorre associação oposta a 43.

Seguindo a análise, são realizados os cálculos para medição das relações entre às variáveis

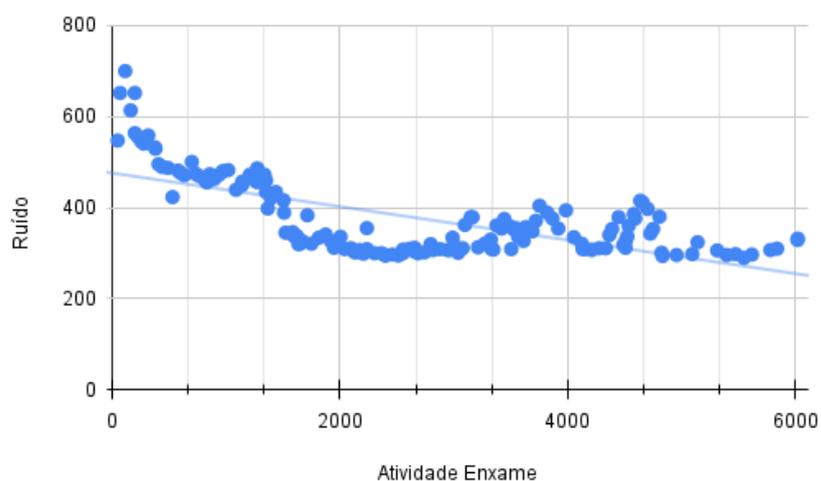
⁵http://labren.ccst.inpe.br/atlas_2017.html

Figura 41 – Coleta ID 1: Correlação entre Atividade Enxame X Peso



Fonte: Elaborado pelo autor. (2021)

Figura 42 – Coleta ID 1: Correlação entre Atividade Enxame X Ruído



Fonte: Elaborado pelo autor. (2021)

e expostas na Tabela 17; utilizando para este feito o coeficiente de correlação linear de Pearson, a partir da fórmula:

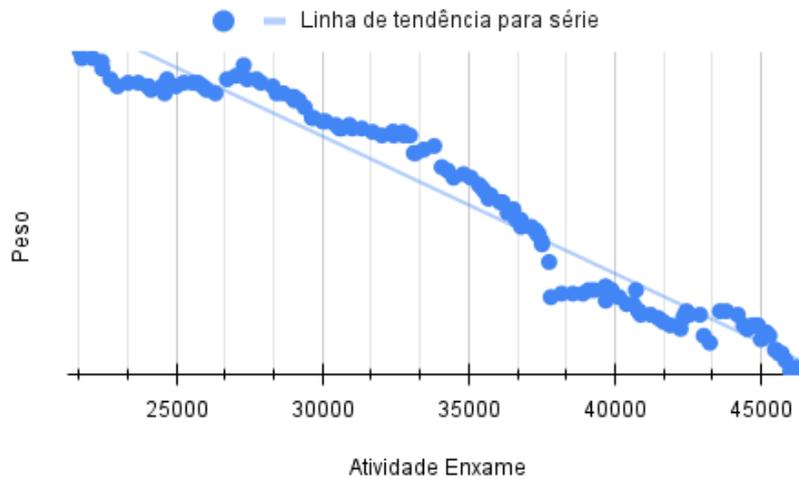
$$r = \frac{S_{xy}}{\sqrt{(S_{xx})(S_{yy})}}$$

Tabela 17 – Coeficiente de correlação de Pearson

Coleta ID 1		Coleta ID 2	
Correlação	<i>r</i>	Correlação	<i>r</i>
Ativ. enxame(X), peso(Y)	0.95 (Forte)	Ativ. enxame(X), peso(Y)	-0.97 (Forte)
Ativ. enxame(X), ruído(Y)	-0.68 (Boa)	Ativ. enxame(X), ruído(Y)	0.62 (Boa)

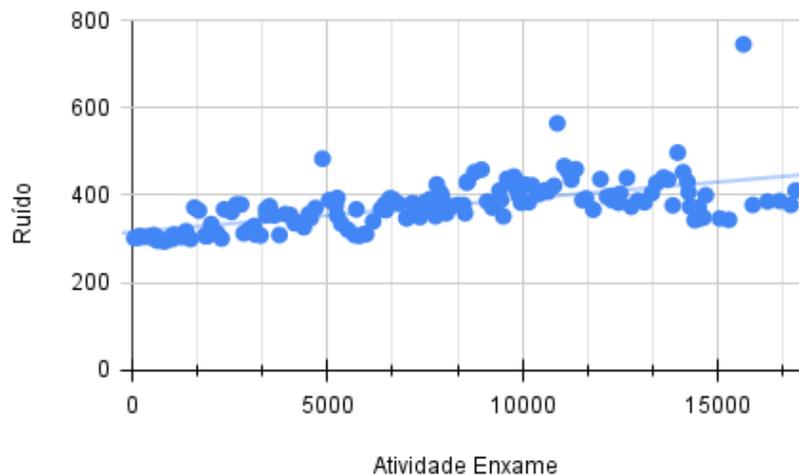
Fonte: Elaborado pelo autor. (2021)

Figura 43 – Coleta ID 2: Correlação entre Atividade Enxame X Peso



Fonte: Elaborado pelo autor. (2021)

Figura 44 – Coleta ID 2: Correlação entre Atividade Enxame X Ruído



Fonte: Elaborado pelo autor. (2021)

Os números apresentados na Tabela 17, exibem coeficientes de Pearson com alto grau de intensidade nas associações, na qual realizando interpretações nestes valores, torna as correlações como Boa e Forte. Em destaque, a relação (atividade enxame (X), peso (Y)), exibiu para as duas coletas, um desenvolvido valor r (acima de 0.90 = Forte), em que na Coleta ID 1, apresentou correlação positiva em suas amostras e negativa em Coleta ID 2. Na associação (atividade enxame (X), ruído (Y)) mostrou Boa intensidade nas duas relações analisadas.

Dado que o coeficiente de correlação não mede a relação causa e efeito entre as variáveis e, conforme visualizada na Tabela 17, há (Boas e Fortes) indícios de que a associação possa estar presente dentro do escopo de amostras testadas. Diante do cenário exposto e sendo de interesse deste pesquisador confirmar a presença das correlações, são formuladas as seguintes hipóteses

a serem testadas:

HIPÓTESE 1:

A hipótese da existência de uma relação entre as variáveis (ativ. enxame (X), peso (Y)).

- $H_0 : p = 0$ (não existe relação)
- $H_1 : p \neq 0$ (existe relação)

HIPÓTESE 2:

A hipótese da existência de uma relação entre as variáveis (ativ. enxame (X), ruído (Y)).

- $H_0 : p = 0$ (não existe relação)
- $H_1 : p \neq 0$ (existe relação)

Parâmetros

As hipóteses foram testadas, usando-se os seguintes parâmetros nos cálculos:

- $r =$ (Verificar Tabela 17)
- $\alpha = 5\%$ (Nível de significância)
- $n = 152$ (Número de amostras)
- $t_{gl} = 150$ (Grau de liberdade)
- $t_{tab} = 1,9759$ (Distribuição T Student)
- $t_c = r \cdot \sqrt{\frac{n-2}{1-r^2}}$ (Teste estatístico: T Student)

Utilizou-se 5% como nível de significância do teste, ou erro tipo I; às 152 amostras das Coletas ID 1 e 2 reportadas no início desta subseção; 150 o grau de liberdade ($n - 2$); usou-se T Student, dado a presença de mais 30 amostras ($n > 30$) e a não existência do desvio padrão da população.

Resultados da Hipótese 1

* Resultados relacionados a Coleta ID 1:

- $t_c = 39.13$
- $t_c > t_{tab} \rightarrow 39.13 > 1,9759$
- Conclusão: t_c (39.13) está dentro de H_1 , logo rejeita-se H_0 e pode-se afirmar, ao nível de 5% de significância que existe relacionamento entre as variáveis.

* Resultados relacionados a Coleta ID 2:

- $t_c = (-48.87) \rightarrow 48.87$
- $t_c > t_{tab} \rightarrow 48.87 > 1,9759$
- Conclusão: t_c (48.87) está dentro de H_1 , logo rejeita-se H_0 e pode-se afirmar, ao nível de 5% de significância que existe relacionamento entre as variáveis.

Resultados da Hipótese 2

* Resultados relacionados a Coleta ID 1:

- $t_c = (-11.36) \rightarrow 11.36$
- $t_c > t_{tab} \rightarrow 11.36 > 1,9759$
- Conclusão: t_c (11.36) está dentro de H_1 , logo rejeita-se H_0 e pode-se afirmar, ao nível de 5% de significância que existe relacionamento entre as variáveis.

* Resultados relacionados a Coleta ID 2:

- $t_c = 9.68$
- $t_c > t_{tab} \rightarrow 9.68 > 1,9759$
- Conclusão: t_c (9.68) está dentro de H_1 , logo rejeita-se H_0 e pode-se afirmar, ao nível de 5% de significância que existe relacionamento entre as variáveis.

6.2.3 Predição dos dados pelo novo modelo de aprendizado de máquina

A partir dos feitos obtidos nos experimentos 1 e 2 reportados na seção 6.1.3, visualizou-se números idênticos para todas as métricas reportadas pelos algoritmos testados; nas quais entende-se que são resultados extraordinários, mas, muito atípicos da realidade. Para averiguar tal comportamento entre as experiências, na qual levantou curiosidade por parte deste pesquisador, realizou-se diversas execuções no modelo, a fim de confirmar se os números permaneceram o mesmo. No final, os dados confirmaram-se, ou seja, conclui-se que os resultados atingidos pelas experiências 1 e 2 são verídicas.

Neste sentido, constata-se as seguintes observações relacionadas a este feito. Primeiramente, levanta-se as hipóteses que os algoritmos tenham atingido tais valores devido a grande quantidade de atributos expostos pelo modelo (dados coletados/inspeção e características inseridas), bem como tenha sofrido o sintoma de problema de Overfitting. Mesmo diante de variâncias iguais ou próximas a zero dos atributos dos dados de inspeção, verificou-se que o excesso de classes sobressaiu, influenciando de forma positiva nos resultados dos experimentos 1 e 2.

Apesar dos frutos conseguidos pelos experimentos 1 e 2 trazerem confiança, otimismo e uma importante contribuição científica, compreende-se que os números conquistados são utópicos, dado que, existem inviabilidade técnica e operacional, no que tange às coletas dos dados de inspeção em um cenário cotidiano.

Neste estudo, conseguiu-se dispor apicultores que realizam as coletas das informações relacionadas aos questionamentos das perguntas dos dados de inspeção *P1*, *P2*, *P3*, *P4*, *P5*, conforme seção 5.3.0.2. Compreende-se do ponto de vista prático, que tais leituras são impraticáveis em um contexto do dia a dia; pois sugere-se pôr uma pessoa captando os dados do ambiente, a fim de preencher o dataset dos dados de inspeção para o modelo.

Tais observações não inviabilizam os resultados oferecidos pelos experimentos 1 e 2, visto que é exposto uma grande conquista científica. Mas, entende-se que o novo modelo de aprendizado de máquina necessita de um aperfeiçoamento e reavaliação em relação à etapa de modelagem dos dados de inspeção. Tal análise objetiva poder dispor de forma prática das características oriundas das pesquisas feitas em campo e assim, gozar dos excelentes resultados aqui reportados.

Finalmente o experimento 3 obteve feitos mais consistentes da realidade. Tais números advêm das coletas de dados feitas por meio dos dispositivos IoT embarcados na caixa de abelhas. Percebe-se na Tabela 15 que todos os valores são superiores a 0.93%, sendo que o modelo Random Forest obteve uma ligeira vantagem em relação ao KNN. Tais superioridades encontram-se em quaisquer das métricas analisadas, na qual pode-se confirmar também nas matrizes de confusão reportada nas Figuras 39.

Estes números apontam que de fato, mesmo sem o conhecimento dos dados de inspeção, os modelos estudados dispõem de frutos satisfatórios. Tais feitos são reforçados em seção 6.2.2 em que visualiza-se a existência de atributos do grupo dos dados coletadas nas quais detêm correlações entre si, favorecendo neste sentido positivamente os resultados, conforme comprova-se nos números expostos.

Cabe reforçar que os experimentos foram aplicados em uma única caixa de abelhas (vide seção 6.1.2). Entende-se que seja relevante aumentar o número de amostras a fim de ampliar o grau de confiabilidade dos resultados, em particular ao experimento 3. Também observa-se que as coletas dos dados da colmeia monitorada foram feitas nos meses de Outubro e Novembro de 2021, conforme reportada na Tabela 8, na qual contempla a estação da primavera. Segundo o autor Maciel (2018) as estações do ano influenciam na rotina e eficiência das abelhas; isto sugere a importância de novos levantamentos de dados nos demais meses do ano.

Mesmo diante aos pontos expostos, compreende-se que as características englobadas no experimento 3 na qual inclui o revolucionário AAE (Analisador Atividade Enxame), oferecem em seus resultados, grandes feitos científicos; neste sentido, entende-se que o novo modelo de aprendizado de máquina, exposto na seção 4.5 apresenta elementos que de fato demonstram que o modelo dispõem da capacidade de realizar seu objeto principal, a predição da enxameação; o que acarreta na diminuição do manejo e consecutivamente o bem estar das abelhas.

6.3 Considerações

Cabe mencionar neste item que os Dados de Inspeção originados pelas pesquisas feitas em campo pelos apicultores, nas quais realizadas em períodos de 3 dias, durante a mesma fase de coleta de dados na colmeia, originou como resultado final um total de 7 exames, gerando consecutivamente sete questionários de inspeção preenchidos. No Apêndice D é exposto a compilação dos dados levantados pelos apicultores.

7 CONSIDERAÇÕES FINAIS

7.1 Conclusão

Este trabalho objetivou na construção de um modelo de arquitetura de software para a apicultura. Denominada de BeeBr, o modelo buscou oferecer tecnologia ao segmento, redução do manejo e dispor de melhorias no bem estar das abelhas. Diante as lacunas apontadas pelos trabalhos relacionados em capítulo 3 e somada às necessidades dos apicultores, procurou-se na concepção do BeeBr, uma estrutura que atendesse às metas de sustentabilidade e eficiência energética; captação das variáveis da colmeia através de dispositivos IoT; elaboração e treinamento de um novo modelo de aprendizado de máquina na qual é responsável pela predição de problemas da enxameação. Com a intenção de atingir tais objetivos expostos, examinou-se soluções nos campos de estudos da *IoT*, *comunicação de dados móveis* e *aprendizado de máquinas*, nais quais são as bases para a formação do BeeBr.

Diante às metas estabelecidas, elaborou-se um modelo de arquitetura de software composta por três camadas, denominadas de detecção, comunicação e serviço; em que cada uma detém a responsabilidade de atingir os objetivos apresentados. Destaca-se as seguintes propostas na modelagem do BeeBr: (1) AAE (Analisador Atividade Enxame) que tem o intuito de verificar o fluxo de entrada e saída das abelhas no alvado da colmeia, em que oferta dados importantes aos algoritmos de ML; (2) DDS (Detecção-Direto-Serviço) que oferece aos apicultores que encontram-se em localidades distantes, sem conectividade de redes móveis, uma alternativa para fazer uso do BeeBr; (3) infraestrutura completa de hardware e software dispostas nos apiários, na qual goza de sustentabilidade energética; (4) uso da tecnologia (SaaS - multitenancy) em que disponibiliza um ambiente na nuvem de forma compartilhada, favorecendo assim a redução de custos; (5) por fim, um moderno modelo de aprendizado de máquina, na qual diferencia-se do estado da arte, a partir do momento que retrata o objetivo de predição da enxameação, baseando-se na inclusão de classes oriundas de dados coletados da colmeia, bem como de pesquisas efetuadas no apiário.

O capítulo 5 expôs a metodologia empregada ao modelo BeeBr. Ressalta-se na camada de detecção o desenvolvimento do protótipo, usando conceitos de IoT, na qual trabalhou-se na inserção de seis sensores, implementação de algoritmo que ofertasse ao máximo a diminuição do consumo energético, desenvolvimento do DDS (Detecção-Direto-Serviço), inclusão de placas solares e sua pré-validação. Na sequência, a camada de serviço apresentou os componentes de back-end, front-end e a metodologia usada ao novo modelo de aprendizado de máquina; sendo este último formado por dados coletados e inspeção. Tais informações serviram de subsídios aos algoritmos KNN e Random Forest, usados pelo moderno modelo de aprendizado de máquina. Por fim, o BeeBr foi testado e validado em uma caixa de abelha de um apiário na cidade de Encruzilhada do Sul/RS nos meses de Outubro e Novembro de 2021, por um período de 20 dias.

Os resultados apresentados pelo modelo BeeBr foram divididos em três temas, nas quais integravam as respostas referentes aos objetivos deste estudo; abordou-se a sustentabilidade e eficiência energética como primeiro ponto, na qual verificou-se ganhos de 16.67% com as implementações feitas pelo algoritmo em que visava a diminuição energética, bem como o funcionamento da camada de detecção por intermédio das placas solares.

Tratou-se no segundo assunto os dados coletados na colmeia, na qual visualiza-se os seguintes feitos: (1) o protótipo de hardware desenvolvido demonstrou-se satisfatório seu funcionamento, mesmo diante as adversidades encontradas no apiário, conseguindo coletar as variáveis por meio dos sensores e na sequência armazená-los localmente; (2) foram feitas duas coletas pelo protótipo de HW, na qual diferenciou-se pelo fato da primeira dispor da abelha rainha e a última sem sua presença. Fez-se esta divisão a fim de gerar uma anomalia (má saúde na colmeia) e consecutivamente originar a enxameação na caixa; em que a partir deste feito, conseguiu-se rotular as informações coletadas. (3) Percebeu-se que houve alterações significativas no comportamento da colmeia, quando a mesma constava de má saúde; sendo que este estado foi por consequência da retirada da abelha rainha.

O último tema abordado relacionou-se aos resultados da predição do novo modelo de aprendizado de máquina, em que baseado nos dados coletados/inspeção, aplicou-se a técnica de seleção das características; por meio dos números obtidos por tal método, este pesquisador dividiu os atributos em três grupos, nominados de experimentos, a fim de serem treinados e testados pelos algoritmos KNN e Random Forest. No final, verificou-se que as características oriundas das pesquisas feitas pelos apicultores, originaram resultados muito atípicos da realidade e, aplicando somente os dados originários pelos sensores, os números são mais consistentes da realidade, girando em torno de 0.93 à 0.95 de acerto na predição da enxameação.

Portanto, a partir dos elementos expostos, entende-se que foi possível responder a questão de pesquisa, citada na introdução desta dissertação.

7.2 Contribuições

Este estudo concretizou as seguintes contribuições científicas e tecnológicas. No campo científico pode-se destacar a elaboração e treinamento de um novo modelo de aprendizado de máquina; tal modelo, objetivou na predição de eventos referentes a problemas na caixa, em particular a enxameação. Sua concepção foi disposta por elementos não encontrados no estado da arte; cita-se o analisador de atividade do enxame (AAE) e os dados provenientes dos apiários (dados de inspeção). Conseguiu-se demonstrar que o modelo dispõe de resultados que satisfaçam seu objeto principal, a predição da enxameação.

Na sequência, buscou-se soluções energeticamente sustentável e eficiente no que tange ao consumo energético por parte dos equipamentos IoT; essa procura projetou recursos de hardware e software para a camada de detecção, na qual resultou no uso do modelo BeeBr em condições adversas no âmbito energético, e, melhorias na eficiência energética de 16.67%.

Por fim, além das contribuições científicas reportadas, houve-se ganhos na área tecnológica; a partir do BeeBr obteve-se uma solução completa de arquitetura computacional de baixo custo para o segmento da apicultura. Por meio do modelo, possibilitou-se leituras, monitoramento e processamento inteligente das informações colhidas dos apiários.

7.3 Limitações

Expõem-se na sequência as seguintes limitações encontradas neste trabalho. Primeiramente encontra-se a não implementação da camada de comunicação do modelo BeeBr. Conforme citado no capítulo 5 houve grandes dificuldades de conseguir sinal de internet móvel no apiário usado. Desta forma, não houve a possibilidade de desenvolver e concluir a prova de conceito da camada de comunicação; o que consecutivamente impactou na realização de um teste completo, fazendo uso de todas as camadas planejadas (detecção, comunicação e serviço) do modelo BeeBr.

Por fim, o analisador de atividade do enxame (AAE) apresenta o recurso de contabilizar a quantidade de abelhas, baseada na movimentação dos insetos na entrada da colmeia (alvado), conforme citado na seção 4.2.1. Na visão deste pesquisador, esta funcionalidade é uma limitação, dado que o estado da arte seria dispor da informação de contagem separada de entrada e saída de insetos, ou seja, o número real de abelhas que de fato saíram e outro que entraram na colmeia.

7.4 Trabalhos Futuros

Para trabalhos futuros, pode-se focar nos temas comunicação de dados e analisador de atividade do enxame (AAE), abordados na seção Limitações 7.3, bem como melhorias no novo modelo de aprendizado de máquina.

Dentro do campo de transmissão de dados, sabe-se que a rede móvel 6G, prevista para o ano de 2030, está em fase de pesquisas, visões, tendências e oportunidades para a futura geração. Neste sentido, oportuniza-se estudos com a tecnologia, a fim de dispor de uma nova alternativa na comunicação de dados ao modelo BeeBr. Uma atualização ao (AAE), apresenta-se com uma possibilidade de estudo; na qual propõem-se buscar um controle melhor e mais preciso da movimentação de entrada e saída da colmeia.

Por fim, entende-se que um aperfeiçoamento em relação à etapa de modelagem dos dados de inspeção do novo modelo de aprendizado de máquina, mostra-se como uma alternativa de trabalho futuro dado aos resultados apresentados por estes atributos no moderno modelo.

REFERÊNCIAS

- ABELHAS, A. B. de Estudo das. **Historico apicultura**. <https://abelha.org.br/historico/>. Acesso em: outubro. 2021.
- ABELHAS, A. B. de Estudo das. **Localização dos apiários**. <https://abelha.org.br/localizacao-dos-apiarios/>. Acesso em: agosto. 2021.
- ABELHAS, A. B. de Estudo das. **Manejo produtivo**. <https://abelha.org.br/manejo-produtivo/>. Acesso em: outubro. 2021.
- ABELHAS, A. B. de Estudo das. **Própolis das abelhas**. <https://abelha.org.br/propolis/>. Acesso em: outubro. 2021.
- ABELHAS, A. B. de Estudo das. **Produção de mel no brasil em 2019**. <https://abelha.org.br/producao-de-mel-no-brasil-cresceu-85-em-2019/>. Acesso em: outubro. 2021.
- Ahmad Saeed, K. et al. Perfil da apicultura no nordeste brasileiro. In: 2014. **Anais...** [S.l.: s.n.], 2014.
- AMMAR et al. The makers' beehives: smart beehives for monitoring honey-bees' activities. In: INTERNATIONAL CONFERENCE ON THE INTERNET OF THINGS, 9., 2019, New York, NY, USA. **Proceedings...** Association for Computing Machinery, 2019. (IoT 2019).
- APICULTURA, C. B. de. **Xv seminário piauiense de apicultura, i seminário piauiense de meliponicultura**. <https://abelha.org.br/historico/>. Acesso em: outubro. 2021.
- APRENDIZAGEM RURAL, S. S. N. de. **Mel: manejo de apiário para produção do mel**. <https://wp.ufpel.edu.br/apicultura/files/2010/05/Manejo-do-Mel.pdf>. Acesso em: outubro. 2021.
- ARRUDA, H. et al. Identifying bee species by means of the foraging pattern using machine learning. In: INTERNATIONAL JOINT CONFERENCE ON NEURAL NETWORKS (IJCNN), 2018., 2018. **Anais...** [S.l.: s.n.], 2018. p. 1–6.
- Ashton, K. That 'internet of things' thing. **RFiD Journal**, [S.l.], 2008.
- BARROS, B. de; GEDRA, R. **Eficiência energética: técnicas de aproveitamento, gestão de recursos e fundamentos**. [S.l.]: Saraiva Educação S.A., 2015.
- BENCSIK, M. et al. Identification of the honey bee swarming process by analysing the time course of hive vibrations. **Computers and Electronics in Agriculture**, [S.l.], v. 76, n. 1, p. 44–50, 2011.
- BERALDO, R. M.; BARBOZA, R. A. B. **Apicultura orgânica - dossiê técnico**. <http://respostatecnica.org.br/dossie-tecnico/downloadsDT/NTQzMg==>. Acesso em: outubro. 2021.
- BEZEMER, C.-P. et al. Enabling multi-tenancy: an industrial experience report. In: IEEE INTERNATIONAL CONFERENCE ON SOFTWARE MAINTENANCE, 2010., 2010. **Anais...** [S.l.: s.n.], 2010. p. 1–8.

Bi, Q. Ten trends in the cellular industry and an outlook on 6g. **IEEE Communications Magazine**, [S.l.], v. 57, n. 12, p. 31–36, 2019.

BRAGA, A. R. **Modelos de classificação para predição do bem estar de colônias da abelha apis mellifera**. 2020. Tese (Doutorado em Ciência da Computação) — Centro de Tecnologia, Universidade Federal do Ceará, Fortaleza, 2020.

CISCO. **Cisco visual networking index: global mobile data traffic**. Disponível em: <<https://s3.amazonaws.com/media.mediapost.com/uploads/CiscoForecast.pdf>>. Acesso em: maio. 2021.

CISCO. **Reimagining the end-to-end mobile network in the 5g era**. Disponível em: <<https://www.cisco.com/c/en/us/solutions/service-provider/mobile-internet/reimagining-mobile-network.html>>. Acesso em: junho. 2021.

COPETTI, J. B.; MACAGNAN, M. H. Baterias em sistemas solares fotovoltaicos. In: I CBENS - I CONGRESSO BRASILEIRO DE ENERGIA SOLAR, FORTALEZA, 8 A 11 DE ABRIL DE 2007, 2007, Fortaleza. **Anais...** Congresso Brasileiro de Energia Solar, 2007.

BUYYA, R.; Vahid Dastjerdi, A. (Ed.). **Internet of things, principles and paradigms**. [S.l.]: Morgan Kaufmann, 2016.

DIETTERICH, T. G. Machine learning. In: _____. **Encyclopedia of computer science**. GBR: John Wiley and Sons Ltd., 2003. p. 1056–1059.

ELABORADO pelo autor. 2021.

Elmeadawy, S.; Shubair, R. M. 6g wireless communications: future technologies and research challenges. In: INTERNATIONAL CONFERENCE ON ELECTRICAL AND COMPUTING TECHNOLOGIES AND APPLICATIONS (ICECTA), 2019., 2019. **Anais...** [S.l.: s.n.], 2019. p. 1–5.

Escovedo, T.; Koshiyama, A. **Introdução a data science — algoritmos de machine learning e métodos de análise**. 1. ed. São Paulo/SP: Casa do Código, 2020.

Ferreira, L. S.; Correia, L. M. Evolução e desafios das redes de comunicações móveis. , [S.l.], 2018.

FITZGERALD, D. et al. Design and development of a smart weighing scale for beehive monitoring. In: 2015. **Anais...** [S.l.: s.n.], 2015. p. 1–6.

GOOGLE. **Pesquisa realizada com o termo iot no google**. Disponível em: <<https://trends.google.com.br/trends/?geo=BR>>. Acesso em: abril. 2021.

GOOGLE. **Pesquisa realizada com o termo machine learning no google**. Disponível em: <<https://trends.google.com.br/trends/?geo=BR>>. Acesso em: maio. 2021.

GUBBI, J. et al. Internet of things (iot): a vision, architectural elements, and future directions. **Future Generation Computer Systems**, [S.l.], v. 29, n. 7, p. 1645–1660, 2013. Including Special sections: Cyber-enabled Distributed Computing for Ubiquitous Cloud and Network Services & Cloud Computing and Scientific Applications — Big Data, Scalable Analytics, and Beyond.

- Guth, J. et al. Comparison of iot platform architectures: a field study based on a reference architecture. In: CLOUDIFICATION OF THE INTERNET OF THINGS (CIOT), 2016., 2016. **Anais...** [S.l.: s.n.], 2016. p. 1–6.
- Herreira, C. C.; Ferreira, E. R. A. Um modelo matemático fuzzy aplicado no crescimento populacional das abelhas. In: 2018. **Anais...** [S.l.: s.n.], 2018.
- Huang, X. et al. Airplane-aided integrated networking for 6g wireless: will it work? **IEEE Vehicular Technology Magazine**, [S.l.], v. 14, n. 3, p. 84–91, 2019.
- Islam, S. M. R. et al. The internet of things for health care: a comprehensive survey. **IEEE Access**, [S.l.], v. 3, p. 678–708, 2015.
- KHAIRUL ANUAR, N. H. et al. Iot platform for precision stingless bee farming. In: IEEE INTERNATIONAL CONFERENCE ON AUTOMATIC CONTROL AND INTELLIGENT SYSTEMS (I2CACIS), 2019., 2019. **Anais...** [S.l.: s.n.], 2019. p. 225–229.
- Kitchenham, B. Kitchenham, b. 2004. procedures for performing systematic review. **Keele University**, [S.l.], p. 1–26, 2004.
- Kocakulak, M.; Butun, I. An overview of wireless sensor networks towards internet of things. In: IEEE 7TH ANNUAL COMPUTING AND COMMUNICATION WORKSHOP AND CONFERENCE (CCWC), 2017., 2017. **Anais...** [S.l.: s.n.], 2017. p. 1–6.
- KRIDI et al. A predictive algorithm for mitigate swarming bees through proactive monitoring via wireless sensor networks. In: ACM SYMPOSIUM ON PERFORMANCE EVALUATION OF WIRELESS AD HOC, SENSOR; UBIQUITOUS NETWORKS, 11., 2014, New York, NY, USA. **Proceedings...** Association for Computing Machinery, 2014. p. 41–47. (PE-WASUN '14).
- Lee, H.; Ke, K. Monitoring of large-area iot sensors using a lora wireless mesh network system: design and evaluation. **IEEE Transactions on Instrumentation and Measurement**, [S.l.], v. 67, n. 9, p. 2177–2187, 2018.
- Li, S.; Xu, L.; Zhao, S. The internet of things: a survey. **Information Systems Frontiers**, [S.l.], v. 17, p. 243–259, 2015.
- LORENZON, M. C. Carga térmica de radiação de dois apiários de abelhas africanizadas dispostos ao sol e à sombra. In: CONGRESSO BRASILEIRO DE MELIPONICULTURA, 1, 2004, Natal. **Anais...** Congresso brasileiro de apicultura: 15., 2004.
- LU, Y.; ZHENG, X. 6g: a survey on technologies, scenarios, challenges, and the related issues. **Journal of Industrial Information Integration**, [S.l.], v. 19, p. 100158, 2020.
- MACIEL, F. A. O. **Reconhecimento de padrões sazonais em colônias de abelhas apis mellifera**. 2018. Dissertação (Mestrado em Ciência da Computação) — Centro de Tecnologia, Universidade Federal do Ceará, Fortaleza, 2018.
- MAHAMUD, M. S. et al. Mouchak - an iot basted smart beekeeping system using mqtt. In: INTERNATIONAL CONFERENCE ON ROBOTICS AND AUTOMATION ENGINEERING (ICRAE), 2019., 2019. **Anais...** [S.l.: s.n.], 2019. p. 84–88.
- Mesquita, A. R. de et al. Ferramenta de monitoramento de temperatura com iot para colmeias de abelhas melíponas. **Brazilian Journal of Development**, [S.l.], 2020.

- Miladinovic, I.; Schefer-Wenzl, S. Nfv enabled iot architecture for an operating room environment. In: IEEE 4TH WORLD FORUM ON INTERNET OF THINGS (WF-IOT), 2018., 2018. **Anais...** [S.l.: s.n.], 2018. p. 98–102.
- Mitchell, T. M. **Machine learning**. 1. ed. USA: McGraw-Hill, Inc., 1997.
- Nawaz, S. J. et al. Quantum machine learning for 6g communication networks: state-of-the-art and vision for the future. **IEEE Access**, [S.l.], v. 7, p. 46317–46350, 2019.
- Oliveira, L. A. N.; Alencar, M. S.; Lopes, W. T. A. Evolução da arquitetura de redes móveis rumo ao 5g. **Revista de Tecnologia da Informação e Comunicação**, [S.l.], v. 8, n. 2, p. 43–50, 2018.
- PEREIRA, E. B. et al. **Atlas brasileiro de energia solar. 2.ed. são José dos campos: inpe**, 2017. 80p. [S.l.]: INPE, 2017.
- Perera, C. et al. Context aware computing for the internet of things: a survey. **IEEE Communications Surveys Tutorials**, [S.l.], v. 16, n. 1, p. 414–454, 2014.
- RODRIGUES Ícaro et al. Detecção de anomalias em padrões acústicos, de temperatura e umidade sazonais para abelhas melíferas (apis mellifera l.). In: XII WORKSHOP DE COMPUTAÇÃO APLICADA à GESTÃO DO MEIO AMBIENTE E RECURSOS NATURAIS, 2021, Porto Alegre, RS, Brasil. **Anais...** SBC, 2021. p. 69–78.
- Roman, R.; Lopes, J. Integrating wireless sensor networks and the internet: a security analysis. **Emerald Group Publishing Limited**, [S.l.], v. 19, p. 246–259, 2009.
- Sekaran, R. et al. Survival study on blockchain based 6g-enabled mobile edge computation for iot automation. **IEEE Access**, [S.l.], v. 8, p. 143453–143463, 2020.
- SHETH, K. et al. A taxonomy of ai techniques for 6g communication networks. **Computer Communications**, [S.l.], v. 161, p. 279 – 303, 2020.
- Tanenbaum, A. S. **Computer networks 5rd**. [S.l.]: Prentice-Hall, 2011.
- TANG, Y. et al. A survey on the 5g network and its impact on agriculture: challenges and opportunities. **Computers and Electronics in Agriculture**, [S.l.], v. 180, p. 105895, 2021.
- TECNICO, C. D. **Manual de sanidade apícola: sintomas – profilaxia – controlo**. Disponível em: <https://fnap.pt/web/wp-content/uploads/documento_cnt_projectos_127.pdf>. Acesso em: set. 2021.
- VAN GOETHEM, S. et al. An iot solution for measuring bee pollination efficacy. In: IEEE 5TH WORLD FORUM ON INTERNET OF THINGS (WF-IOT), 2019., 2019. **Anais...** [S.l.: s.n.], 2019. p. 837–841.
- VAPNIK, V. N. **The nature of statistical learning theory**. [S.l.]: Springer-Verlag New York, Inc., 1995.
- Vidrascu, M. G.; Svasta, P. M. Maintenance-free iot gateway design for bee hive monitoring. In: IEEE 23RD INTERNATIONAL SYMPOSIUM FOR DESIGN AND TECHNOLOGY IN ELECTRONIC PACKAGING (SIITME), 2017., 2017. **Anais...** [S.l.: s.n.], 2017. p. 189–193.

Vidrascu, M. G.; Svasta, P. M. Embedded software for iot bee hive monitoring node. In: IEEE 23RD INTERNATIONAL SYMPOSIUM FOR DESIGN AND TECHNOLOGY IN ELECTRONIC PACKAGING (SIITME), 2017., 2017. **Anais...** [S.l.: s.n.], 2017. p. 183–188.

XU, R.; WUNSCH, D. Survey of clustering algorithms. **IEEE Transactions on Neural Networks**, [S.l.], v. 16, n. 3, p. 645–678, 2005.

Yaacoub, E.; Alouini, M. A key 6g challenge and opportunity—connecting the base of the pyramid: a survey on rural connectivity. **Proceedings of the IEEE**, [S.l.], v. 108, n. 4, p. 533–582, 2020.

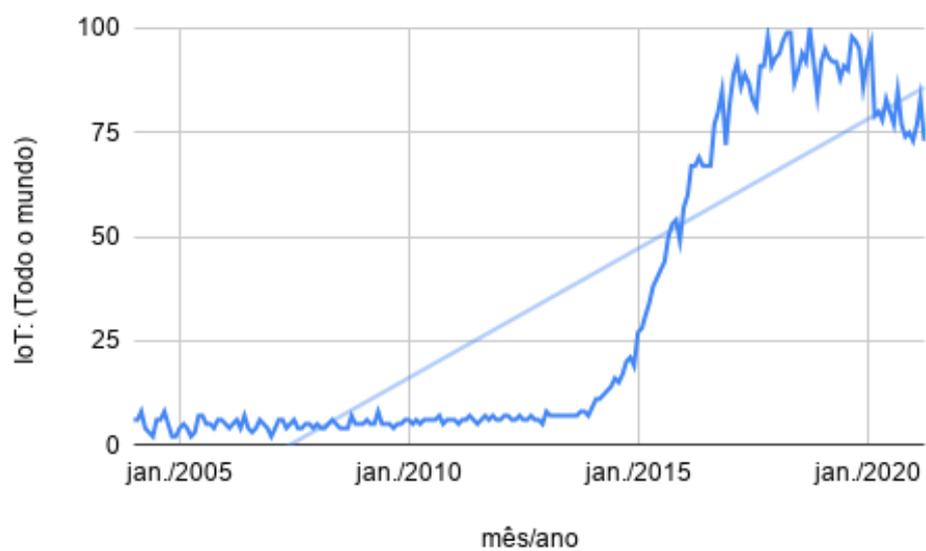
ZABASTA, A. et al. Iot approach application for development of autonomous beekeeping system. In: INTERNATIONAL CONFERENCE IN ENGINEERING APPLICATIONS (ICEA), 2019., 2019. **Anais...** [S.l.: s.n.], 2019. p. 1–6.

Zhang, S. et al. Envisioning device-to-device communications in 6g. **IEEE Network**, [S.l.], v. 34, n. 3, p. 86–91, 2020.

Zong, B. et al. 6g technologies: key drivers, core requirements, system architectures, and enabling technologies. **IEEE Vehicular Technology Magazine**, [S.l.], v. 14, n. 3, p. 18–27, 2019.

APÊNDICE A – PESQUISA FEITA COM O TERMO IOT NO GOOGLE

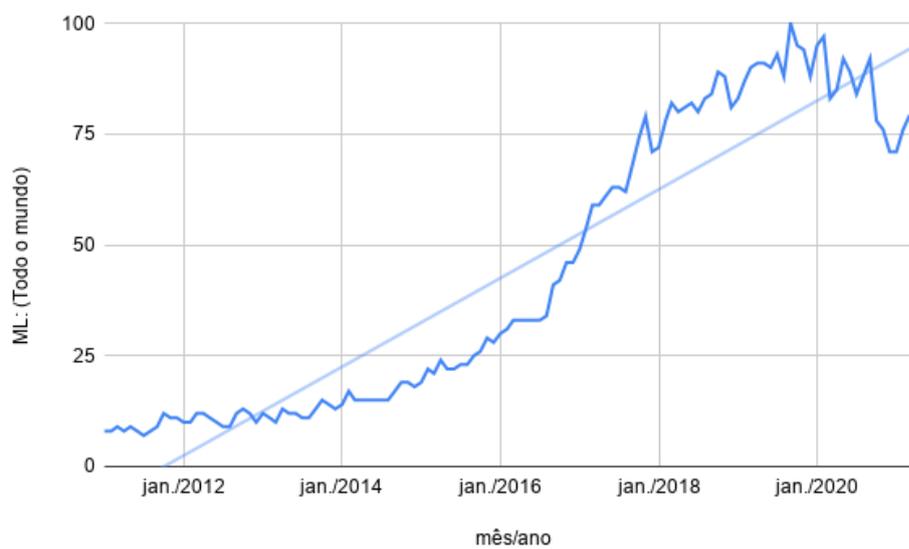
Figura 45 – Pesquisa feita com o termo IoT no Google



Fonte: Google (2021a)

APÊNDICE B – PESQUISA FEITA COM O TERMO ML NO GOOGLE

Figura 46 – Pesquisa feita com o termo Machine Learning no Google



Fonte: Google (2021b)

APÊNDICE C – QUESTIONÁRIO DE INSPEÇÃO

Figura 47 – Questionário de Inspeção

Nome Apicultor: _____
Data (dd/mm): _____

Responda as perguntas abaixo, em cada inspeção no apiário.

Pergunta	Resposta (sim/não) e/ou observações
A - Existem a presença de pragas (traças ou formigas) ?	
B - Dispõem de alimentação suficiente ao redor ?	
C - Há uma circulação fora do comum ?	
D- A caixa está em condição apta ?	
E - Como está a previsão do tempo ?	
F - O sistema está em operação ?	

Fonte: Elaborado pelo autor. (2021)

APÊNDICE D – RESULTADOS DO QUESTIONÁRIO DE INSPEÇÃO

Figura 48 – Resultados do Questionário de Inspeção

Pergunta	Coleta ID 1					Coleta ID 2	
	1/11/2021	4/11/2021	7/11/2021	10/11/2021	13/11/2021	16/11/2021	19/11/2021
A	não	não	não	não	não	não	não
B	sim	sim	sim	sim	sim	sim	sim
C	não	não	não	não	não	sim	sim
D	sim	sim	sim	sim	sim	sim	sim
E	sol	sol	nublado	chuva	chuva	chuva	nublado
F	sim	sim	sim	sim	sim	sim	sim

Fonte: Elaborado pelo autor. (2021)

APÊNDICE E – CUSTOS DA CAMADA DE DETECÇÃO

Figura 49 – Custos da Camada de Detecção

Custos Camada de Detecção	
Componente	Valor
cpu	60,50
sensor peso	25,90
sd card	8,90
sensor movimento	16,90
sensor gas	19,90
modulo coversor peso	14,49
caixa proteção	120,00
sensor peso	16,00
sensor ruido	15,50
sensor temp/umidade	18,00
cartão sd	39,00
cabeamento	10,00
controlar carga	61,30
baterias	270,00
placa solar	220,50
	916,89

Fonte: Elaborado pelo autor. (2021)