



INSTITUTO FEDERAL DE CIÊNCIA E TECNOLOGIA DE PERNAMBUCO

Campus Recife

Departamento Acadêmico de Sistema, Processos e Controles Eletroeletrônicos

Tecnologia em Análise e Desenvolvimento de Sistemas

MICHEL HENRIQUE DA SILVA NASCIMENTO

**ANÁLISE DE DADOS COLETADOS POR DISPOSITIVOS DE INTERNET DAS  
COISAS APLICADA A SISTEMAS EMBARCADOS**

Recife

2020

MICHEL HENRIQUE DA SILVA NASCIMENTO

**ANÁLISE DE DADOS COLETADOS POR DISPOSITIVOS DE INTERNET DAS  
COISAS APLICADA A SISTEMAS EMBARCADOS**

Trabalho de conclusão de curso apresentado ao Departamento acadêmico de sistema, processos e controles eletroeletrônicos do Instituto Federal de Ciência e Tecnologia de Pernambuco, como requisito para obtenção do título de Tecnólogo em Análise e Desenvolvimento de Sistemas.

Orientador: Prof. MsC. Anderson Luiz Souza  
Moreira

Recife

2020

N244a Nascimento, Michel Henrique da Silva.  
2020 Análise de dados coletados por dispositivos de Internet das Coisas aplicada a  
Sistemas Embarcados / Michel Henrique da Silva Nascimento. – Recife: O Autor, 2020.

61f. il. Color.

TCC (Curso Superior Tecnológico em Análise e Desenvolvimento de Sistemas) –  
Instituto Federal de Pernambuco, Departamento Acadêmico de Sistema, Processos e  
Controles Eletroeletrônicos, 2020.

Inclui Referências

Orientador: Prof. M.e. Anderson Luiz Souza Moreira

1. Análise de Dados. 2. Internet das Coisas. 3 Mineração de Dados. 4. Sistemas  
Embarcados. I. Título. II Moreira, Anderson Luiz Souza (orientador). III. Instituto Federal  
de Pernambuco.

CDD 005.13

Trabalho de Conclusão de Curso apresentado por **Michel Henrique da Silva Nascimento** à coordenação de Análise e Desenvolvimento de Sistemas, do Instituto Federal de Pernambuco, sob o título de “**Análise de dados coletados por dispositivos de Internet das Coisas aplicada a Sistemas Embarcados**”, orientado pelo Prof. **Anderson Luiz Souza Moreira** e aprovada pela banca examinadora formada pelos professores:

Recife, 03 de dezembro de 2020.

---

Prof. M. Sc. Anderson Luiz Souza Moreira  
CSIN/DASE/IFPE

---

Prof. M. Sc. Hilson Vilar Gomes Andrade  
DASE/IFPE

---

Prof. Ph.D. Marco Antônio de Oliveira Domingues  
CSIN/DASE/IFPE

## **AGRADECIMENTOS**

Agradeço a orientação do professor Anderson Moreira que me auxiliou na construção desta pesquisa através de sugestões que contribuíram de forma significativa. Obrigado pelos ensinamentos e por ter despertado em mim o pensamento crítico.

Aos docentes do curso de Análise e Desenvolvimento de Sistemas do IFPE que diretamente colaboraram com minha formação acadêmica.

À minha noiva pelo apoio e companheirismo de sempre.

Aos meus pais pelo cuidado e suporte vitalício.

*“Os otimistas são ingênuos, e os pessimistas amargos.*

*Sou um realista esperançoso.”*

*Ariano Suassuna*

## RESUMO

O crescimento significativo do número de sistemas embarcados – principalmente inseridos em um ambiente contextualizado pelo paradigma tecnológico da Internet das Coisas (IoT) – tem sido responsável por uma geração massiva de dados detentores de uma capacidade intrínseca de produção de valor que foi desaproveitada em certa medida até a virada deste século, tendo como causa diversos motivos, muitos deles relacionados às barreiras presentes no desenvolvimento da engenharia microeletrônica vigente na época. Por sua vez, esse valor atualmente pode ser obtido a partir da aplicação de técnicas de análise, possibilitada em grande parte pela utilização de algoritmos de mineração de dados que são responsáveis por extrair informações relevantes em formato descritivo, prescritivo ou preditivo. Neste estudo, uma revisão do estado dos processos de análise de dados em sistemas IoT é apresentada, explicitando o comportamento dos métodos mais comumente utilizados, assim como são feitas considerações sobre os desafios e oportunidades que surgem em sistemas desse tipo. Tais exposições são alicerçadas por uma compilação de trabalhos recentes que focam em temas próximos ao desta pesquisa, além da construção de uma prova de conceito fundamentada na criação de um sistema *web*, sendo esse condicionado pela exploração de dados provenientes de grupos de sensores conectados à Internet.

Palavras-chave: Análise de Dados. Internet das Coisas. Mineração de Dados. Sistemas Embarcados. Sistema *web*.

## **ABSTRACT**

The significant growth in the number of embedded systems – mainly those in an environment contextualized by the technological paradigm of the Internet of Things (IoT) – has been responsible for a massive generation of data with an intrinsic capability to produce value that was largely unused until the turn of this century, for several reasons, many of them related to the barriers present in the development of microelectronic engineering in force at the time. In turn, nowadays the value can be obtained through the application of analysis techniques, which are possible because of the use of data mining algorithms that are responsible for extracting relevant information in a descriptive, prescriptive or predictive format. In this study, a review of the state of the data analysis processes in IoT systems is presented, explaining the behavior of the most commonly used methods, as well as considerations about the challenges and opportunities that arise in similar systems. Such content is supported by a compilation of recent works that focus on topics close to this research, in addition to the construction of a proof of concept based on the creation of a web system, which is conditioned by the exploration of data from groups of sensors connected to the Internet.

**Keywords:** Data Analysis. Internet of Things. Data Mining. Embedded Systems. Web System.

## LISTA DE FIGURAS

Figura 1 – Modelo 3V's.....	22
Figura 2 – Divisão do mercado IoT global.....	23
Figura 3 – Arquitetura geral de um dispositivo IoT.....	28
Figura 4 – Passageiros de voos internacionais por mês.....	31
Figura 5 – Aplicação do algoritmo KNN para $K = 1$ e $K = 100$ .....	34
Figura 6 – Classificação em SVM.....	35
Figura 7 – Agrupamentos K-means para $k = 3$ .....	36
Figura 8 – Visão geral do Modelo Implementado .....	38
Figura 9 – Exemplo de um dispositivo IoT.....	42
Figura 10 – Arquitetura para coleta de dados.....	43
Figura 11 – Estrutura de pré-processamento dos dados .....	45
Figura 12 – Fluxograma da metodologia Box-Jenkins.....	46
Figura 13 – Arquitetura de implementação de modelos do Sistema .....	47
Figura 14 – Tabela de uma estação de medição.....	50
Figura 15 – Dashboard para visualização de aferições .....	51
Figura 16 – Histograma dos valores de temperatura .....	52
Figura 17 – Resumo de elementos da Série Temporal .....	53
Figura 18 – Visualização dos dados em forma de gráfico de linha .....	54
Figura 19 – Gráfico dos valores previstos .....	55

## LISTA DE TABELAS

<b>Tabela 1 – Critérios para levantamento bibliográfico .....</b>	<b>37</b>
<b>Tabela 2 – Descrição da tabela ESTACAO_001 .....</b>	<b>44</b>

## LISTA DE ABREVIATURAS

ARIMA – *Autoregressive Integrated Moving Average*  
CSS – *Cascading Style Sheets*  
CSV – *Comma-Separated Values*  
DLANN – *Deep Learning Artificial Neural Network*  
D2D – *Device to Device*  
D2S – *Device to Server*  
HTML – *Hyper Text Markup Language*  
HTTP – *Hyper Text Transfer Protocol*  
HW – *Holt-Winters*  
IDC – *International Data Corporation*  
IDE – *Integrated Development Environment*  
IEEE – *Institute of Electrical and Electronics Engineers*  
IFPE – *Instituto Federal de Educação, Ciência e Tecnologia de Pernambuco*  
IoT – *Internet of Things*  
ITU – *International Telecommunication Union*  
KNN – *K-Nearest Neighbors*  
LDA – *Linear Discriminant Analysis*  
LDR – *Light Dependent Resistor*  
LGPD – *Lei Geral de Proteção de Dados*  
NB – *Naive Bayes*  
RFID – *Radio-Frequency Identification*  
RNAs – *Redes Neurais Artificiais*  
SVM – *Support Vector Machine*  
S2S – *Server to Server*

## SUMÁRIO

1 INTRODUÇÃO	13
1.1 Justificativa	14
1.2 Objetivos	15
<b>1.2.1 Objetivo Geral</b>	15
<b>1.2.2 Objetivos Específicos</b>	15
1.3 Organização do Trabalho	15
2 ESTADO DA ARTE	17
3 FUNDAMENTAÇÃO TEÓRICA	20
3.1 Internet das Coisas e Análise de Dados	20
<b>3.1.1 IoT como fonte de dados</b>	20
<b>3.1.2 A Importância do Big Data Para a IoT</b>	21
<b>3.1.3 Indústria 4.0</b>	22
<b>3.1.4 Segmentos que se utilizam da IoT</b>	23
3.1.4.1 Cidades Inteligentes	24
3.1.4.2 Transporte	25
3.1.4.3 Saúde	25
3.1.4.4 Indústria	26
3.1.4.5 Domótica	26
3.1.4.6 Agricultura de Precisão	27
3.2 Implementação da Análise de Dados em Sistemas IoT	27
<b>3.2.1 Sistemas Embarcados e Conectividade</b>	28
<b>3.2.2 Coleta e Manipulação dos Dados</b>	29
<b>3.2.3 Implantação de Modelos de Análise</b>	29
3.3 Mineração de Dados no Contexto da IoT	30
<b>3.3.1 Séries Temporais</b>	30
3.3.1.1 ARIMA	32
3.3.1.2 Holt-Winters	33
<b>3.3.2 Classificação</b>	33

3.3.2.1 K-Nearest Neighbors	33
3.3.2.2 Support Vector Machine	34
<b>3.3.3 Agrupamento</b>	<b>35</b>
3.3.3.1 K-means	35
4 Metodologia	37
4.1 Métodos para levantamento e análise da literatura	37
4.2 Métodos para a implementação do protótipo	38
<b>4.2.1 Preparação dos Dados</b>	<b>39</b>
<b>4.2.2 Mineração dos Dados</b>	<b>39</b>
<b>4.2.3 Visualização dos Dados</b>	<b>40</b>
5 Desenvolvimento do Protótipo	41
5.1 Materiais e Ferramentas	41
<b>5.1.1 Sensores e Microcontrolador</b>	<b>41</b>
<b>5.1.2 Servidor</b>	<b>42</b>
<b>5.1.3 Shiny</b>	<b>42</b>
5.2 Dados Coletados	43
<b>5.2.1 Pré-Processamento</b>	<b>44</b>
5.3 Linguagens de Programação	45
5.3.1 Aplicação do Modelo de Previsão e Implementação	46
6 Resultados	48
6.1 Desafios e Oportunidades	48
6.2 Prova de Conceito	49
<b>6.2.1 Cenário e Funcionalidades Implementadas</b>	<b>50</b>
<b>6.2.2 Análise Exploratória dos Dados</b>	<b>51</b>
<b>6.2.3 Aplicação do Modelo Preditivo</b>	<b>54</b>
7 CONSIDERAÇÕES FINAIS	56
7.1 Trabalhos Futuros	57
REFERÊNCIAS	59

## 1 INTRODUÇÃO

A Internet das Coisas (IoT) pode ser entendida como a interconexão de vários dispositivos independentes através da Internet. Em termos práticos, IoT é uma infraestrutura global de informações sociais que permite diversas formas de serviços avançados, que são alcançados através da interação (virtual ou física) de “coisas”, baseando-se nas tecnologias existentes e em seu próprio desenvolvimento. Essa estrutura é percebida como um dos maiores resultados do avanço tecnológico atual, devido ao potencial que atingiu na área da tecnologia da informação (ITU, 2012).

Até o fim de 2025, 41,6 bilhões de dispositivos estarão conectados à Internet, tais conexões possibilitarão cada vez mais o uso de dados para análise e tomada de decisões inteligentes de forma autônoma (IDC, 2019). Essa condição é evidenciada pelo grande número de aplicações em diversas áreas que a IoT está diretamente envolvida como, por exemplo, cidades inteligentes, agricultura de precisão, saúde, logística, transporte e economia. Esses exemplos são bases sólidas que impulsionam o uso e o desenvolvimento dessa tecnologia por causa da necessidade que os próprios ambientes possuem de elementos que requerem processos mais precisos e eficientes.

O crescimento acelerado de tecnologias que envolvem o desenvolvimento de sistemas microprocessados de tarefa específica dentro de um ambiente controlado, conhecidos como sistemas embarcados, fomentado pela importância dada atualmente à Internet das Coisas, traz certos desafios a serem enfrentados com urgência. Assim, surgem diversos aspectos que se fazem presentes no estado contemporâneo da tecnologia e que são resultados do seu rápido avanço: a quantidade massiva de dados gerados por meio desses sistemas, questões de eficiência operacional, capacidade de processamento e acurácia dos dispositivos embarcados, meios de comunicação para transmissão de dados, consumo de energia, etc.

Diante desse cenário, fica evidente a necessidade da busca por algoritmos computacionais e processos que possam transformar essa quantidade enorme de dados, gerados pela intensificação da presença de sistemas inseridos na tecnologia IoT, em informações qualificadas. A análise de dados aplicada no contexto

apresentado se torna bastante desafiadora, tendo em vista os processos incomuns a que é submetida, a escolha de métodos e algoritmos na realização dessa análise adquire outras variáveis a serem consideradas no intuito de garantir resultados adequados ao sistema.

## 1.1 Justificativa

O processo de análise de dados ganha mais complexidade quando está relacionado a sistemas que envolvem o paradigma IoT. Através do estudo de modelos, métodos e técnicas de análise em vários campos, podem ser extraídas informações e conhecimento com maior eficiência e confiabilidade. Dessa forma, é notável a relevância aplicada ao estudo de modelagens de dados com o uso de métodos de classificação, agrupamento e regressão, por exemplo. Isso acontece porque encontrar o modelo de melhor ajuste ao conjunto de dados gerados é uma das tarefas mais importantes na análise de dados.

Para tanto, alguns *softwares* importantes como *R* e *Python* possuem papel fundamental na busca pela solução de problemas circunscritos no tipo de sistema apresentado no início deste tópico. Tais sistemas são bastante usados para lidar com os dados gerados através da tecnologia IoT, pois fornecem ferramentas robustas para a aplicação de modelos que são vitais para a obtenção de respostas em forma de tomadas de decisão e extração de informações, imputando inteligência ao processo.

Desse modo, os benefícios do uso e da compreensão das tecnologias abordadas neste estudo são inúmeros, tendo em vista que indivíduos e companhias podem adquirir vantagens significativas através da análise de grandes volumes de dados, bem como pelo gerenciamento das informações obtidas. Segundo Marjani et al. (2017), o desenvolvimento de análises em sistemas envolvendo IoT pode ajudar organizações e associações a terem um melhor entendimento sobre os dados e assim, tomarem decisões mais eficientes e bem informadas gerando valor social e econômico no ambiente em que estão inseridas.

## 1.2 Objetivos

### 1.2.1 Objetivo Geral

O objetivo geral do estudo é investigar os principais métodos para análise de dados e suas aplicações em sistemas que envolvem conceitos da Internet das Coisas, compreendendo as principais características do sistema e as restrições impostas no contexto em que são inseridos, além de implementar uma prova de conceito que compreende o tema do trabalho apresentado.

### 1.2.2 Objetivos Específicos

Os seguintes objetivos específicos são abordados neste estudo: Indicar os desafios e limitações da análise de dados em sistemas no contexto IoT; apresentar comparativos entre alguns dos algoritmos mais comumente utilizados para mineração de dados, tais como *KNN*, *K-means*, *Holt-Winters*, *ARIMA* e *SVM*; analisar modelos de arquiteturas de sistemas inteligentes que existem hoje; desenvolver um sistema *Web* experimental de baixo custo, usando tecnologias *open source*<sup>1</sup>, para análise de dados em um ambiente dentro do contexto descrito, capaz de permitir a exploração dos dados envolvidos no processo, aplicando técnicas de análise adequadas e fornecendo informações visuais através de uma interface de usuário.

## 1.3 Organização do Trabalho

Este trabalho está organizado da seguinte forma: O Capítulo 2 apresenta o estado da arte no que se refere ao tema abordado neste estudo, trazendo contextualizações de trabalhos correlatos recentes. O Capítulo 3 trata da fundamentação teórica que envolve a discussão de conceitos, ideias e compreensões sobre definições importantes que dão base ao prosseguimento do trabalho. No Capítulo 4 uma metodologia simplificada da pesquisa e da construção

---

<sup>1</sup> Termo que se refere a códigos abertos, livres para receber modificações e redistribuições.

da prova de conceito proposta é descrita. O desenvolvimento de um sistema experimental IoT com serviços de análise de dados é apresentado no Capítulo 5 de forma mais detalhada. O Capítulo 6 apresenta os resultados alcançados e por fim, o Capítulo 7 descreve as considerações finais do estudo e propõe trabalhos futuros.

## 2 ESTADO DA ARTE

Este capítulo tem como objetivo apresentar publicações cujas temáticas se correlacionam à deste estudo. Além disso, intenta-se ressaltar os resultados e os progressos das pesquisas estudadas e relacioná-los com a proposta de apresentação deste trabalho.

A pesquisa de Tsai et al. (2014) trata de responder alguns questionamentos comuns sobre preparação e processamento de dados, tidos como os maiores desafios nas áreas da IoT e análise de dados. Além disso, o autor apresenta de forma concisa uma explicação sobre Internet das coisas, os dados e os desafios que existem na atualidade, também descreve os cuidados e dificuldades ao se estabelecer modelos de algoritmos para mineração de dados. A estrutura do texto apresentado indica três grandes etapas para o desenvolvimento da modelagem em curso: o processamento dos dados, a extração de padrões e características dos dados extraídos e a escolha em si do algoritmo para tratá-lo.

No trabalho de Chen, Deng e Wan (2015) foram observados, de forma detalhada, os desafios enfrentados durante o processo de construção de aplicações que se relacionam com os temas: Internet das coisas e *Big Data*. Foram revisados conceitos, técnicas e aplicações referentes à mineração de dados, bem como algoritmos de classificação, associação, agrupamento e de séries temporais também foram descritos. Ainda nesse estudo foi possível abordar diversos sistemas que geram bases de dados potenciais para a aplicação desses algoritmos, como setores do comércio eletrônico, indústria, saúde e cidades inteligentes. Algumas relações importantes entre métodos de mineração de dados e aplicações mais apropriadas para seu uso também foram observadas.

Em uma abordagem inovadora, Zhou et al. (2014) apresenta uma soma de opiniões e perspectivas sobre novas oportunidades e desafios trazidos pela nova onda da utilização massiva de dados, comum no contexto IoT, relacionando as ideias dos vários autores e aplicações existentes na contemporaneidade que foram construídas sobre esse tema. As concepções destacadas têm o objetivo de provocar discussões pertinentes sobre dados e meios analíticos de manipulá-los, através de técnicas de aprendizado de máquina, formas de otimização, distribuição, etc. Por isso, o trabalho referido é interessante não apenas por produzir uma pesquisa sobre

o tópico, mas por criar um material interdisciplinar gerado pela contribuição de diferentes expertises.

Ahmed et al. (2017) explora em seu estudo os avanços mais recentes referentes a análises em sistemas IoT. Também descreve os requisitos necessários para lidar com essa tecnologia, citando os desafios para que se possa analisar de forma inteligente os dados provindos desses tipos de sistemas.

O estudo de Ge, Bangui e Buhnova (2018) envolve uma pesquisa sobre as tecnologias *Big Data* em diferentes aplicações da IoT. O trabalho ainda compara casos de uso que relacionam Internet das Coisas com novas tecnologias, mostrando como essas formas de se observar os dados podem ser reutilizadas dentre os mais diversos contextos de aquisições de dados automatizados.

Mahdavinejad et al. (2018) apresenta em seu estudo uma compreensão objetiva da relação entre algoritmos de aprendizado de máquina e aplicações IoT, observando os modelos mais frequentemente utilizados que possibilitam a transformação de dados comuns em dados inteligentes. O estudo aponta também as principais características dos dados providos atualmente através desse paradigma, além de trazer um estudo específico sobre um caso de uso referente a cidades inteligentes usando o algoritmo *Support Vector Machine* (SVM) que tem a função de analisar dados para classificação ou regressão.

Marjani et al. (2017) propõe uma nova arquitetura para análise de dados em sistemas IoT. O trabalho discute ainda sobre tipos, métodos e tecnologias de mineração computacional no âmbito de grandes volumes de dados. Também são apresentados diversos casos de uso importantes sobre o tema, além da descrição de novas visões de oportunidades trazidas pela interseção prática dos termos IoT e análise de dados.

Em Nait Malek et al. (2017) é proposta a criação de uma plataforma digital completa que trata combinações diretas entre análise de dados e Internet das Coisas. Esse estudo apresenta relações entre o uso dessas tecnologias em monitoramentos feitos em tempo real, assim como o processamento automatizado dos dados gerados de forma contínua. O trabalho ainda abrange o uso da plataforma desenvolvida em casos de uso reais.

Em Alam et al. (2016) foram feitas comparações entre oito algoritmos mais comuns de mineração de dados no contexto da Internet das Coisas. Em uma base de dados gerada por três sensores e com o auxílio do software R, foram executadas

simulações e comparações entre os algoritmos: *Support Vector Machine* (SVM), *K-Nearest Neighbors* (KNN), *Linear Discriminant Analysis* (LDA), *Naive Bayes* (NB), C4.5, C5.0, Redes Neurais Artificiais (RNAs) e *Deep Learning ANN* (DLANN). Os resultados do estudo indicaram bons índices entre os algoritmos C4.5 e C5.0 no que concerne aos parâmetros mais importantes seguintes: acurácia, eficiência no uso da memória e complexidade computacional.

Qiu et al. (2016) apresenta uma revisão de literatura ampla que compreende os últimos avanços nas pesquisas que se referem ao aprendizado de máquina e processamento de *Big Data*. Nesse trabalho é feita a revisão de técnicas de aprendizado de máquina, abordando também seus métodos, *deep learning*, *parallel learning*, *transfer learning*, *kernel-based learning*, dentre outros. Isso permite a compreensão de como sistemas aplicados em contextos múltiplos podem ter mudanças consideráveis em termos da eficiência nos usos específicos desses métodos, especialmente quando relacionados a sistemas que envolvem aspectos da Internet das Coisas. Ainda é possível extrair desse estudo os desafios e as possíveis soluções para esses tipos de análises de dados, além de alguns problemas importantes a serem resolvidos em pesquisas futuras.

O presente trabalho, além de realizar um estudo teórico envolvendo métodos e algoritmos relacionados a procedimentos de análise de dados, os quais são descritos nos objetivos específicos desta pesquisa, no capítulo anterior, também apresenta o desenvolvimento de um sistema IoT com serviços para a implementação dessa análise, capaz de demonstrar relações empíricas importantes percebidas na construção do mesmo, assim como destaca os desafios e ligações intrínsecas nesse tipo de arquitetura. Tal estudo adquire importância por fornecer uma compilação de uma base teórica relativa a temas que estão no auge de seu desenvolvimento, dando uma visão geral dos desafios e caminhos a serem seguidos, com o intuito de se ampliar o conhecimento e de se extrair novas informações e ideias a respeito do tema. Isso pode servir de guia para aplicações específicas de análise de dados do mundo real, assim como para a construção de modelos experimentais *open source* de pesquisa que estejam inseridos no contexto abordado.

## 3 FUNDAMENTAÇÃO TEÓRICA

### 3.1 Internet das Coisas e Análise de Dados

Este tópico busca descrever e relacionar conceitos que abrangem os temas da Internet das Coisas e da análise de dados, discutindo a importância de suas tecnologias na atualidade, bem como as oportunidades e desafios inerentes aos processos imersos nesse contexto.

#### 3.1.1 IoT como fonte de dados

A IoT surge com uma ideia central de que todas as coisas podem estar conectadas à internet, essas “coisas” podem ser quaisquer dispositivos ligados a sensores. Tais conexões são em sua maioria realizadas seguindo os padrões: D2D (*entre dispositivos*), D2S (*entre dispositivo e servidor*) e S2S (*entre servidores*). Isso permite que a coleta de informações ganhe alcance proporcional ao número de dispositivos espalhados.

A internet das Coisas refere-se à integração de objetos físicos e virtuais em redes conectadas à Internet, permitindo que “coisas” colem, troquem e armazenem informações a respeito do meio que estão inseridas, isso implica em uma enorme quantidade de dados gerados, em que uma vez processados e analisados, geram informações e serviços em escala inimaginável. Apontada como uma revolução tecnológica iminente e com mercado mundial estimado em 1,7 trilhão de dólares em 2020, a IoT gera impacto em todas as áreas, incluindo eletrônica de consumo, saúde, e de maneira transversal, na forma como a sociedade consome informação (LIMA; CUNHA, 2018, p.3).

Segundo previsão realizada pela IDC (2019), estima-se que até 2025, 41,6 bilhões de dispositivos, desde *smartphones*, *laptops*, câmeras, sensores e consoles de *videogame* estarão conectados à Internet por meio de várias redes de acesso, gerando uma soma em torno de 79,4 *zettabytes*<sup>2</sup> de dados. Essa comunicação acontece por meio de tecnologias que comumente envolvem redes de sensores sem fio. Os recentes avanços nesse tipo de tecnologia posicionam a IoT como a próxima

---

<sup>2</sup> Um *zettabyte* equivale a 1.000 *exabytes*, 1.000.000 *petabytes* ou  $10^{21}$  *bytes*.

tecnologia revolucionária, pois aproveita muitas das oportunidades oferecidas pela Internet.

Decerto, é possível observar que as tecnologias IoT são incorporadas em vários segmentos importantes da nossa vida. Nos últimos anos, alguns segmentos tradicionais, como manufatura, saúde e energia, tornaram-se baseados em IoT e obtiveram a capacidade de realizar a comunicação entre máquinas e humanos, gerando a criação de um modelo de vida moderno, inteligente e autônomo (GE; BANGUI; BUHNOVA, 2018).

### **3.1.2 A Importância do Big Data Para a IoT**

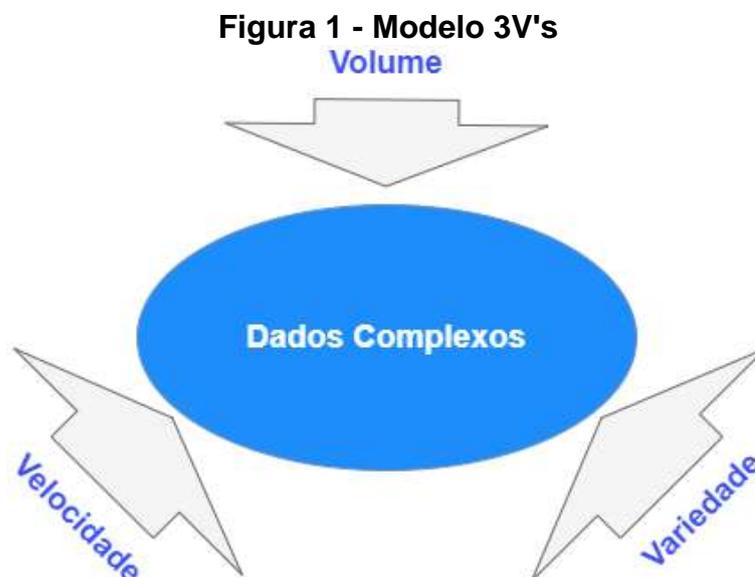
A utilização de sensores, dispositivos eletrônicos, mídias sociais, aplicativos de assistência médica e tantos outros dispositivos digitais, que continuamente geram grandes quantidades de dados estruturados, não estruturados ou semiestruturados, está aumentando fortemente. Essa geração massiva de dados e sua necessidade de tratamento fez com que sistemas antigos de banco de dados entrassem em desuso, pois não eram adequados o suficiente para processá-los, resultando na utilização moderna do termo conhecido como *Big Data* (MARJANI et al., 2017).

O termo Big Data pode ser definido como um vasto conjunto de dados, armazenados de forma estruturada ou não, que requer ferramentas especiais para que informações específicas sejam localizadas e possam ser analisadas e combinadas em tempo real, de modo que sejam tomadas as melhores decisões (JESUS et al., 2016, p. 2-3).

De modo geral, a principal representação do *Big Data* se associa à caracterização de um conjunto de dados em enorme quantidade. Segundo Mahdavinejad et al. (2018), esse termo pode ser definido como uma referência a dados com grande volume, velocidade e variedade que demandam formas avançadas de processamento para que sejam possibilitados o ganho de informação sobre eles, conhecimento, automação de processos e tomada de decisão.

Assim, a definição do termo supracitado compreende também o modelo 3V's (volume, variedade, velocidade), esse modelo é próprio de sistemas de gerenciamento de dados no comércio eletrônico, que enfrenta diversos desafios para gerenciar grandes quantidades, diferentes fontes de dados e velocidade na criação. Outros pesquisadores ainda introduzem três características adicionais, a

veracidade, o valor e a variabilidade, transformando esse modelo em um modelo 6V's (MARJANI et al., 2017).



Fonte: Adaptado de Jesus et al. (2016)

A Figura 1 apresenta um esquemático da ideia que envolve o *Big Data* e como os atributos estão diretamente relacionados à sua complexidade.

### 3.1.3 Indústria 4.0

A concepção da indústria 4.0 é consequência do desenvolvimento da Internet das Coisas, juntamente com o termo *Big Data*. É sabido que a terceira revolução industrial foi fortemente baseada em condições como automação e informatização de processos, o modelo de indústria referido faz menção a uma espécie de quarta revolução industrial, focada na utilização de modelos digitais de produtos, com suas produções relacionadas a fábricas formadas por procedimentos inteligentes (WITKOWSKI, 2017).

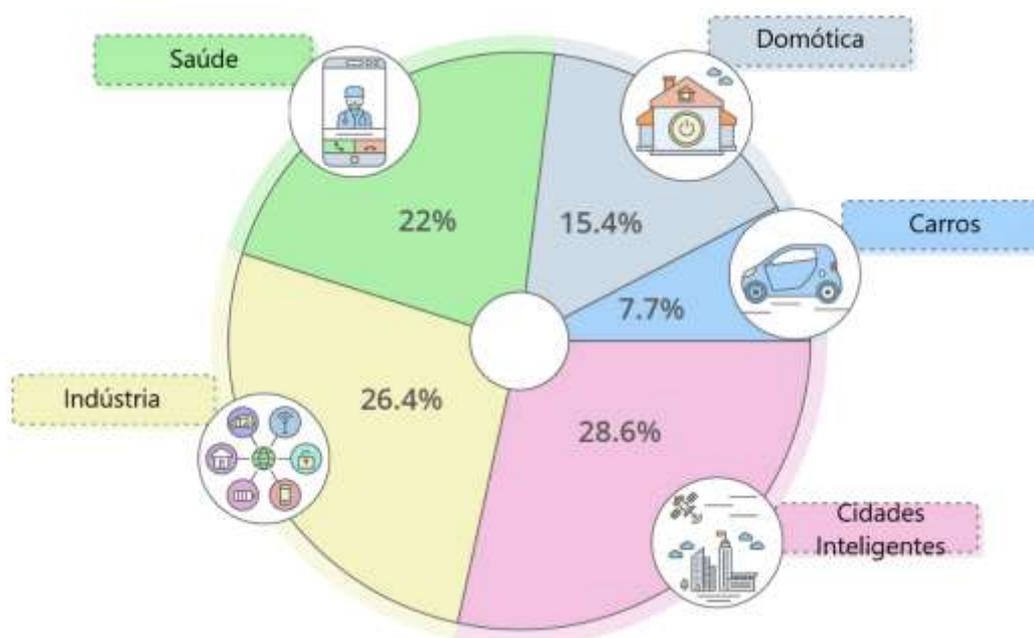
Esse novo conceito na indústria provoca um grande impacto em termos de produção, contribuindo com o aumento da eficiência em diversos aspectos referentes a métodos e processos das companhias de modo geral. As vantagens comerciais alcançadas não estão apenas diretamente ligadas à possibilidade de acesso imediato a dados ou a um melhor controle de produtos, mas também à geração de informações essenciais que guiam as empresas para uma melhor

compreensão do ambiente em que está inserida, o que torna os processos internos de tomada de decisão mais acurados através da aplicação de algoritmos de mineração de dados que conseguem obter uma menor taxa de erro, conforme possuam como entrada um maior volume de dados.

### 3.1.4 Segmentos que se utilizam da IoT

Neste tópico serão apresentados alguns segmentos da sociedade impactados pela união das tecnologias IoT e de análises de dados.

**Figura 2 – Divisão do mercado IoT global**



Fonte: Adaptado de Bustamante (2019)

Como pode ser observado, a Figura 2 apresenta uma divisão do mercado global em relação às áreas que possuem envolvimento com a Internet das Coisas. A partir do gráfico, é possível identificar os principais segmentos em que a tecnologia está inserida.

### 3.1.4.1 Cidades Inteligentes

A proposta da criação de cidades inteligentes é melhorar o modo de vida dos cidadãos, fornecendo aplicativos inteligentes em várias áreas. Para tanto, a cidade inteligente adota tecnologias IoT para otimizar diferentes sistemas e serviços públicos, como estacionamento de automóveis, limpeza urbana, gestão de resíduos, iluminação pública e controle de emergência. Alguns projetos desse tipo encontram-se em desenvolvimento em cidades importantes do mundo demonstrando que a coleta de dados é uma iniciativa fundamental para obter experiência e conhecimento necessários para incentivar o desenvolvimento desse segmento (GE; BANGUI; BUHNOVA, 2018).

Existem alguns indicadores importantes que determinam o nível de inteligência de uma cidade, são eles: mobilidade, urbanismo, meio ambiente, tecnologia e inovação, empreendedorismo, educação, saúde, segurança, energia, governança e economia. Dentro dessas características a cidade de Barcelona na Espanha é tida hoje como uma das mais importantes nesse processo de evolução, um dos principais responsáveis por esse desempenho é o conjunto de plataformas disponibilizado pela prefeitura, como a *decidim* e a *sentillo*<sup>3</sup>, que são plataformas de código aberto com acesso a milhares de sensores espalhados pela cidade, utilizando linguagens como *Java*, *Ruby*, *JavaScript*, dentre outras.

O enorme conjunto de sensores atua em diversas topologias, atendendo as áreas mais distintas e fornecendo a qualquer cidadão a capacidade de extrair valores importantes dentro da arquitetura de dados abertos produzida. Cidades como Singapura, Helsinque e Zurique possuem as melhores performances no *Smart City Index 2020*<sup>4</sup> e também possuem plataformas *opensource* arquitetadas com a distribuição em massa de sensores em várias localidades, o sucesso das cidades envolve primordialmente o compromisso das prefeituras juntamente com uma participação ativa de seus cidadãos, esses casos de sucesso possuem as tecnologias de IoT e análise de dados como matérias indispensáveis de política pública.

---

<sup>3</sup> Repositórios de mesmo nome são disponibilizados na plataforma de versionamento de código GitHub.

<sup>4</sup> IMD. Smart City Index, 2020. Disponível em: <https://www.imd.org/smart-city-observatory/smart-city-index>. Acesso em: 22 de out. de 2020.

O cenário no Brasil em termos de cidades inteligentes ainda é bem incipiente, basta que se note o *Ranking Connected Smart Cities 2020*<sup>5</sup>, que mapeia as cidades com maior capacidade de conectividade no país, São Paulo que aparece como a mais bem pontuada possui a nota 37,901 em um ranking com nota máxima de 70.

#### 3.1.4.2 Transporte

O desenvolvimento em tecnologias IoT permitiram diversos avanços referentes ao tráfego, já que essa atividade ocorre todos os dias em todo o mundo, produzindo uma grande quantidade de dados, suficiente para dar inteligência ao processo se tratadas com algoritmos específicos. Planejamento de rotas, previsão e controle de trânsito intenso são atividades que podem ser atribuídas a aplicações desse tipo.

Modernamente, as aplicações ligadas a transportes são focadas em controle de frotas e monitoramento dos dispositivos eletromecânicos presentes nos automóveis. A cooperativa de ônibus escolares KWRL usa diversos serviços de controle de rotas através do estudo analítico do caminho seguido pelos seus diversos veículos desde sua origem até o destino, além de oferecer plataformas onde os pais dos estudantes podem acompanhar o traslado em tempo real.

Ainda na área de transporte, um dos grandes avanços está relacionado aos carros autônomos, a Tesla é uma das empresas que está na liderança nesse setor que promete ser o futuro dos veículos pessoais, todos os carros da empresa são construídos com foco em tecnologias IoT.

#### 3.1.4.3 Saúde

Serviços relativos aos cuidados de saúde possuem grandes oportunidades de implementação por meio de aplicações IoT. Através do monitoramento constante de diversas condições do indivíduo em tempo real, é possível adquirir informações referentes a sinais importantes, como variações no batimento cardíaco, movimentação corporal, temperatura do corpo e até sinais elétricos do cérebro. Os

---

<sup>5</sup> URBANSYSTEMS, 2020. Disponível em: <https://ranking.connectedsmartcities.com.br>. Acesso em 22 de out. de 2020.

dados podem ser tratados e processados fornecendo conhecimento e suporte na detecção de anormalidades no corpo humano.

Atualmente, avanços na área da saúde se tornaram ainda mais urgentes por um fator externo que afeta o mundo, a pandemia de coronavírus<sup>6</sup>. O assistente *nCapp* surgiu dessa demanda, através da coleta de informações pessoais via *smartphones* e dados na *cloud*<sup>7</sup> em tempo real, ele consegue diagnosticar pacientes e classificar graus de pneumonia baseando-se em tecnologia médica e em conceitos da Internet das Coisas (LI, 2020).

#### 3.1.4.4 Indústria

Soluções IoT que envolvem automação de processos, controle e monitoramento de uma forma geral são capazes de gerar demasiados ganhos nas indústrias. A redução do custo operacional das atividades e o controle de gastos de forma minuciosa garantem um desenvolvimento econômico mais eficaz e são condições essenciais e desejadas em todos os setores industriais.

A área de aplicação da IoT industrial hoje está ligada a uma série de projetos desenvolvidos dentro e fora das fábricas. Internamente alguns projetos de automação e controle são muito comuns, fazendo com que o monitoramento da produção seja feito de modo mais inteligente. Os projetos de fora das fábricas estão relacionados a plataformas de controle remoto de maquinários e operações que podem ser executadas e monitoradas a distância no intuito de reduzir o tempo de atividade operacional e custos (SCULLY, 2020).

#### 3.1.4.5 Domótica

A conectividade entre sistemas embarcados, possibilitada por tecnologias IoT, permite um modo de vida diferente das pessoas dentro de suas residências. A evolução da domótica – também referenciada como automação residencial – hoje garante condições de segurança através de dispositivos de alerta e monitoramento,

---

<sup>6</sup> BRASIL, Ministério da Saúde. COVID-19. Disponível em: <https://coronavirus.saude.gov.br/sobre-a-doenca>. Acesso em: 5 de nov. de 2020.

<sup>7</sup> Sistema de armazenamento e processamento de dados que funciona através da internet, sem a necessidade do gerenciamento direto pelo usuário.

além de maior conforto, possibilitado pela operação de funções da casa através de sistemas geridos em computadores ou celulares de qualquer lugar do mundo.

#### 3.1.4.6 Agricultura de Precisão

Vale ainda apontar um outro segmento importante que surge no cenário global, a agricultura de precisão é um exemplo desse tipo de setor onde se usa a análise de dados em conjunto com tecnologias *Big Data* e IoT. Os sensores nesse caso são instalados em lugares estratégicos para obter dados sobre o nível de umidade do solo, diâmetro do tronco das plantas, condição do microclima e níveis de umidade, bem como para previsão do tempo. Os sensores transmitem dados obtidos usando dispositivos em rede, esses dados passam por um *gateway*<sup>8</sup> e depois são processados a fim de gerir atuadores<sup>9</sup>. O controle automático do clima de acordo com as necessidades de colheita, a irrigação oportuna e controlada e o controle de umidade para a prevenção de fungos são fatores de extrema relevância nessa área e só são possíveis graças ao desenvolvimento de análises de dados em sistemas incluídos no contexto abordado (MARJANI et al., 2017). Segundo Pivoto et al (2018), a introdução dessas tecnologias no sistema de produção agrícola, aumentando a eficiência dos equipamentos e de produção, é representada pelo conceito de *Smart Farming* que se apresenta hoje como o estado da arte na gestão de negócios da agricultura.

### 3.2 Implementação da Análise de Dados em Sistemas IoT

Segundo Forbes (2017), companhias mercadológicas que usam processos de análise de dados mais apurados geram três vezes mais valor em sistemas IoT se comparado às que não investem nessa metodologia. A aplicação de métodos que possibilitam essa análise adquire uma série de implicações quando ela é relacionada a sistemas imersos no contexto da Internet das Coisas, por diversos motivos que serão apresentados nesta seção.

---

<sup>8</sup> Dispositivo capaz de gerenciar o fluxo de dados entre redes diferentes.

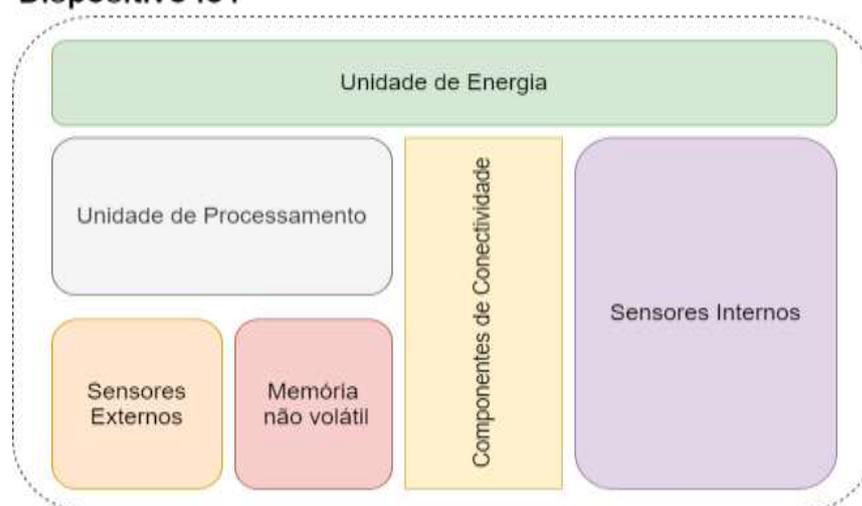
<sup>9</sup> Dispositivos responsáveis por mover e controlar um mecanismo ou sistema.

### 3.2.1 Sistemas Embarcados e Conectividade

Há um ponto de distinção crucial entre sistemas embarcados convencionais e os embarcados IoT, a diferença está exclusivamente na conectividade. A Internet das Coisas é uma infraestrutura global que compreende numerosas aplicações e é por meio da comunicação e transmissão de dados que as interações e a inteligência do processo podem acontecer. Nesse contexto, sistemas desse tipo devem incorporar no mínimo um protocolo de comunicação sem fio e um dispositivo programável. Uma complexidade enorme de análises e decisões estruturais é levada em conta quando estão presentes estes termos: o estudo da tecnologia de comunicação a ser utilizada, o nível de eficiência energética do sistema e as perdas nas taxas de transmissão de dados.

Atualmente a interconexão de diversos dispositivos acontece através de uma variedade de tecnologias como *Bluetooth*, LoRa<sup>10</sup>, SigFox<sup>11</sup>, dentre outras. Tal abrangência na diversidade implica na necessidade de flexibilização de modelos para adquirir valor em cada um dos possíveis grupos de elementos utilizados em uma rede IoT, seja pelas características específicas dos tipos de *hardware* envolvidos no sistema ou da sua própria tecnologia de comunicação.

**Figura 3 – Arquitetura geral de um dispositivo IoT**  
Dispositivo IoT



Fonte: Adaptado de Samie (2016).

<sup>10</sup> Tecnologia de comunicação de longo alcance e baixo consumo de energia, informações específicas sobre o funcionamento dessa tecnologia são obtidas em: <https://lora-alliance.org>.

<sup>11</sup> Protocolo bastante usado em aplicações IoT para transmissão de dados em longas distâncias, mais detalhes podem ser acessados em: <https://www.sigfox.com>.

A Figura 3 apresenta uma arquitetura geral para dispositivos embarcados IoT, tais componentes estão quase sempre presentes nas aplicações inseridas nesse contexto.

### **3.2.2 Coleta e Manipulação dos Dados**

O elemento mais importante para a implementação da análise de dados é o conjunto de dados em si. Em um ambiente que envolve Internet das Coisas esses dados estudados possuem certas características intrínsecas à arquitetura referida. Diferentes dispositivos geram dados distintos, esparsos e geralmente em frequências arbitrárias em sua rede de interconexão.

Combinar informações coletadas por diversas fontes e tratá-las em um ambiente único a fim de que através de suas relações próprias produzam informações qualificadas é uma tarefa que demanda certa complexidade e exige a aplicação de técnicas como a fusão de sensores.

A fusão de sensores envolve a combinação de informações provenientes de diferentes sensores para captar dados do ambiente cuja obtenção está além da capacidade isolada de cada um dos sensores envolvidos, principalmente quando são considerados confiabilidade e precisão. Além disso, a fusão de sensores tanto torna o sistema mais tolerante a falhas, como também pode fornecer informações novas, que nenhum dos sensores isoladamente consegue obter (FACELI, 2001, p.3).

Sensores evoluem constantemente, seja em suas características físicas como na redução de tamanho, seja nos seus protocolos de comunicação, permitindo que a geração de dados seja processada gradualmente de forma mais precisa e em uma velocidade maior.

### **3.2.3 Implantação de Modelos de Análise**

Modelos de análise de dados aplicados em sistemas IoT possuem metodologias diferentes dos aplicados em um ambiente tradicional por alguns fatores, dentre eles, a frequência na captura de dados, a variedade das fontes desses dados e a velocidade de processamento necessária.

Quando se trata de observar o modelo de análise, considera-se tanto a criação do modelo quanto à implantação desse, tecnologias como computação na nuvem e computação de borda são comumente usadas em sistemas de Internet das Coisas, assim que o modelo é elaborado de modo a funcionar para resolver o problema em questão, a implantação do modelo pode seguir em correspondência com a arquitetura que deve estar alicerçada por um projeto específico e completo, tendo em vista a sua complexidade.

### 3.3 Mineração de Dados no Contexto da IoT

Este tópico visa descrever as técnicas e algoritmos de mineração de dados mais comuns aplicados em ambientes que envolvem a Internet das Coisas.

#### 3.3.1 Séries Temporais

Diversas aplicações IoT são desenvolvidas hoje através da utilização de séries temporais, compreendendo domínios importantes como a meteorologia, a economia e a biomedicina. Uma série temporal é uma sequência de medidas históricas de uma variável observada em determinados intervalos de tempo. Assim, uma série temporal  $S$  com  $n$  observações, pode ter a seguinte formulação:

$$S = (s_1, s_2, s_3, \dots, s_n)$$

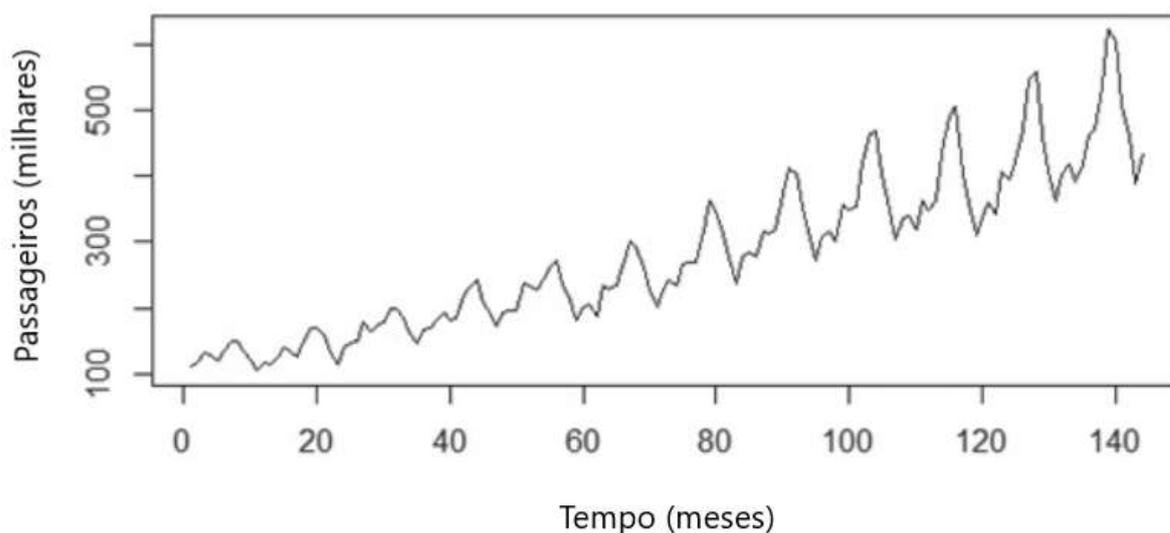
onde  $s_t \in \mathbb{R}$ , dado que  $s$  é uma observação no instante de tempo  $t$ .

Dentre os principais motivos para o estudo de séries temporais está a atividade de previsão, pois conseguir prever comportamentos futuros baseando-se em características históricas tem um potencial bastante relevante em diversos campos da ciência.

A Figura 4 mostra um exemplo de uma série temporal, refere-se a um gráfico que representa o número de passageiros por mês que fizeram voos internacionais durante um período de 12 anos.

No exemplo, tem-se a exposição da ideia vista na definição de série temporal, ou seja, nesse caso, é definida como uma sequência ordenada de 144 valores que representam os meses durante os 12 anos.

**Figura 4 – Passageiros de voos internacionais por mês**



Fonte: Adaptado de Dietrich (2015)

Uma série temporal é composta por quatro componentes característicos, são eles: tendência, sazonalidade, ciclicidade e aleatoriedade. A tendência se refere às inclinações ou pontos de inflexão que a série pode apresentar, tais movimentos indicam a direção de seu comportamento no tempo, o exemplo da Figura 4 mostra uma tendência de aumento em termos gerais no número de passageiros dos valores observados no tempo.

Sazonalidade é um elemento que descreve os registros de mudanças em certo período, ou seja, a forma como o comportamento da série se aproxima de flutuações em tempos determinados, como estações do ano, feriados, etc.

Outro componente importante é a ciclicidade, esse é diretamente relacionado à sazonalidade, porém não possui a característica de período determinado, um exemplo comum são os ciclos econômicos, que possuem condições geradas por uma série de situações distintas.

A aleatoriedade é também um componente presente e que exige maior cuidado quando as análises são feitas com o intuito de se obter resultados futuros.

### 3.3.1.1 ARIMA

O modelo *Auto Regressive Integrated Moving Average* (ARIMA) é resultado da combinação de três diferentes processos estatísticos, são eles: Auto regressão (AR), Integração (I) e Médias Móveis (MA).

Problemas envolvendo análise de séries temporais utilizam tradicionalmente técnicas lineares, como a ARIMA. Essa técnica é aplicada muito comumente na implementação de aplicações com esse tipo de comportamento, inclusive quando o objetivo é prever condições futuras através de dados obtidos no passado.

Ainda que na prática muitas das séries envolvidas nas aplicações sejam não estacionárias, o que impede a aplicação de modelos auto regressivos de forma direta, em muitos casos se usa modelos de diferenciação para torná-las estacionárias (NALLAKARUPPAN; KURAMAN, 2019).

Uma série temporal  $S_t$ , para  $t = 1, 2, 3, \dots, n$ , é dita estacionária se essa atende três condições: a média de  $S_t$  é constante para todos os valores de  $t$ , a variância de  $S_t$  é finita e a covariância de  $S_t$  e  $S_{t+h}$  depende apenas dos valores de  $h=0, 1, 2, \dots$  para todo  $t$ .

A covariância de  $S_t$  e  $S_{t+h}$  é a medida de como duas variáveis,  $S_t$  e  $S_{t+h}$ , variam juntas. Podendo ser expressa pela equação:

$$cov(S_t, S_{t+h}) = E[(S_t - \mu_t)(S_{t+h} - \mu_{t+h})]$$

Assim, relacionando os três diferentes processos estatísticos característicos desse modelo, esse pode ser descrito através da equação:

$$S_t = \delta + \theta_1 S_{t-1} + \theta_2 S_{t-2} + \dots + \theta_p S_{t-p} + \varepsilon_t + \varphi_1 \varepsilon_{t-1} + \dots + \varphi_q \varepsilon_{t-q}$$

Onde  $\delta$  é uma constante que assume valores para séries não centralizadas no ponto zero,  $\theta_i$  é uma constante para  $i = 1, 2, \dots, p$ ,  $\theta_p \neq 0$ ,  $\varphi_j$  é uma constante para  $j = 1, 2, \dots, q$ ,  $\varphi_q \neq 0$  e  $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$  para todo  $t$ . A partir daí também pode-se considerar para séries temporais não estacionárias a diferenciação dada por:

$$d = S_t - S_{t-1}, \text{ onde } t = 2, 3, \dots, n.$$

### 3.3.1.2 Holt-Winters

O modelo *Holt-Winters* (HW) é um modelo de suavização recomendado para séries temporais com tendência e sazonalidade, esse possui dois tipos: aditivo e multiplicativo.

O modelo HW aditivo tem maior capacidade de explicação em séries que possuem tendência e sazonalidade aditiva, podendo ser representado pela equação a seguir:

$$Z_t = \mu_t + T_t + F_t + \varepsilon_t,$$

onde  $t = 1, \dots, n$ , sendo  $n$  o número de observações da série temporal.

Da mesma forma, o modelo HW multiplicativo é usado para atender séries com tendência e sazonalidade multiplicativa, podendo assim ser formulada:

$$Z_t = \mu_t F_t + T_t + \varepsilon_t$$

Outro modelo importante trata da suavização exponencial. Segundo Becker (2010), essa técnica possui uma vantagem importante por não precisar armazenar dados, pelo fato de que a atualização da previsão realizada através desse método tem como variável um atributo da observação mais recente e da última realizada, por outro lado previsões além de um passo à frente não podem ser obtidas.

Tal modelo é representado pela seguinte equação:

$$\bar{Z}_t = \alpha Z_t + \alpha(1 - \alpha)Z_{t-1} + \alpha(1 - \alpha)^2 Z_{t-2} + \dots,$$

onde  $\bar{Z}_t$  é o valor suavizado e  $0 \leq \alpha \leq 2$  é a constante de suavização.

### 3.3.2 Classificação

Nesta seção são apresentados os algoritmos de mineração de dados mais utilizados em aplicações IoT, de acordo com a literatura pesquisada.

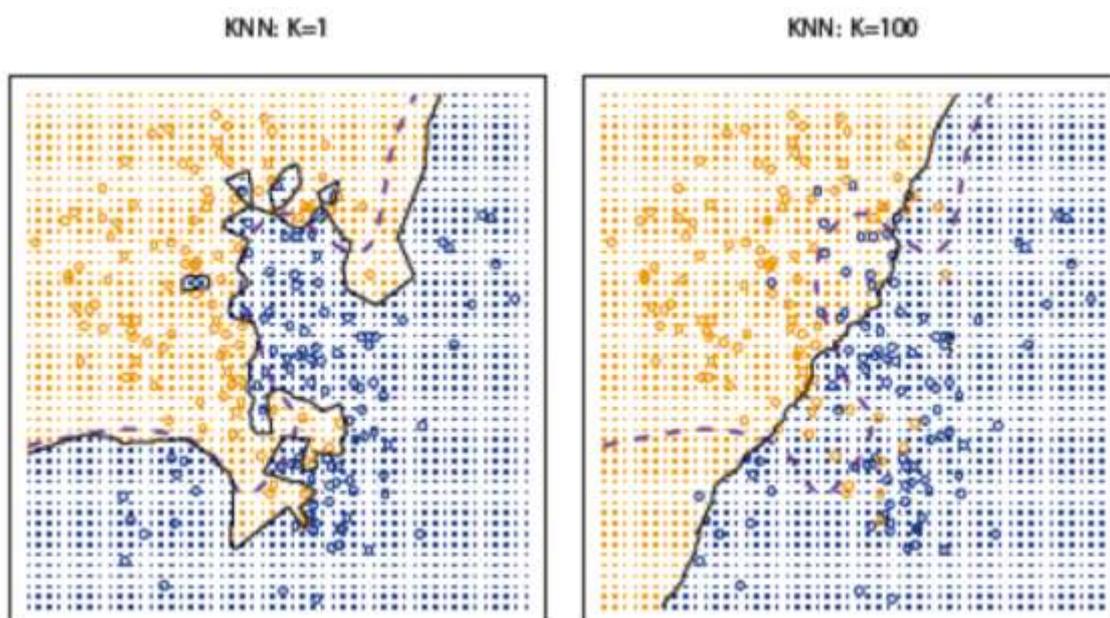
#### 3.3.2.1 K-Nearest Neighbors

O algoritmo *K-Nearest Neighbors* ou KNN, tem o objetivo de classificar um novo atributo de dados observando os K dados mais próximos obtidos na base de dados usada como treinamento. A parametrização dos K pontos vizinhos (mais

próximos) é dada por uma escolha de como será condicionada a distância entre pontos, dentre essas métricas, a distância euclidiana é a mais utilizada.

A quantidade de vizinhos é dada pelo valor atribuído à letra  $K$ , desse modo, quando  $K = 1$ , a classe será escolhida avaliando a classe da amostra mais próxima dentro do conjunto de treino. A variação nesse valor pode causar diversas consequências na busca pela classificação da amostra.

**Figura 5 – Aplicação do algoritmo KNN para  $K = 1$  e  $K = 100$**



Fonte: Langer (2018)

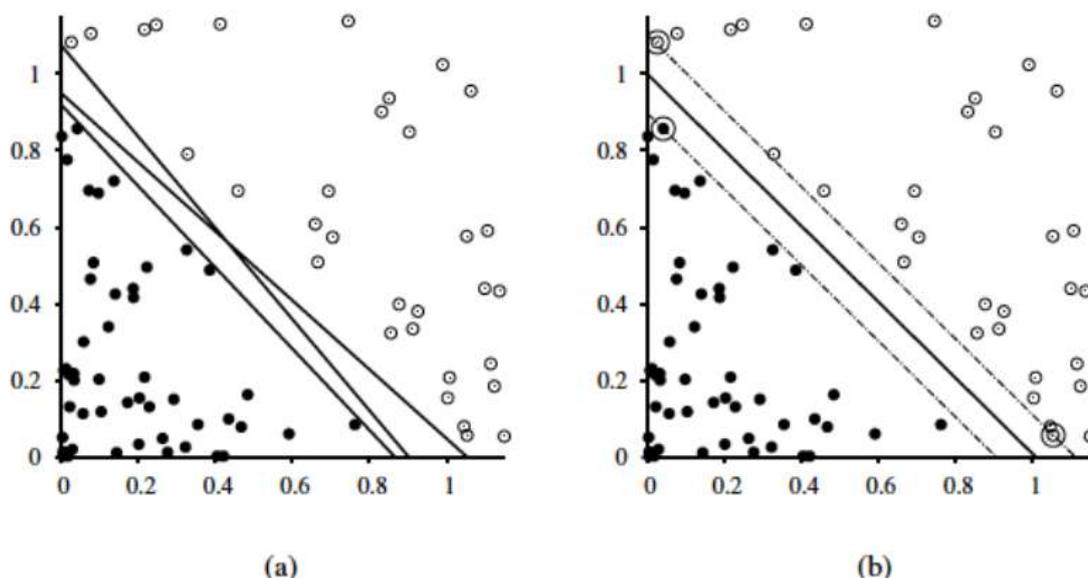
A Figura 5 mostra uma diferença clara na consequência das diferentes aplicações do valor de  $K$ , de um lado, baixo viés e alta variância, do outro o inverso. Uma significativa limitação do *KNN* é o fato de que esse método requer o armazenamento de todo conjunto de dados de treino, trazendo assim problemas de escalabilidade em sistemas com uma grande quantidade de dados.

### 3.3.2.2 Support Vector Machine

O modelo *Support Vector Machine* (SVM), ou máquina de vetores de suporte, é um tipo de algoritmo de mineração de dados que tem como entrada um conjunto de dados, e para cada uma delas consegue prever dentre duas classes possíveis,

a classe a que essa entrada pertence. Assim, tal algoritmo é capaz de produzir um modelo que atribui categorias aos conjuntos de dados, esse também é percebido como um classificador binário não probabilístico pelos motivos descritos no tópico (SCHMITT, 2013).

**Figura 6 - Classificação em SVM. (a) Espaço com três separadores. (b) Separador com margem**



Fonte: Langer (2018)

Como pode ser percebido através da Figura 6 (b), os vetores de suporte constroem um separador de margem máxima, auxiliado também pelos candidatos a separadores como visto em (a).

### 3.3.3 Agrupamento

Algoritmos de agrupamento dividem o conjunto de dados em grupos específicos que respeitam determinado padrão e por consequência permanece distinto de outros grupos.

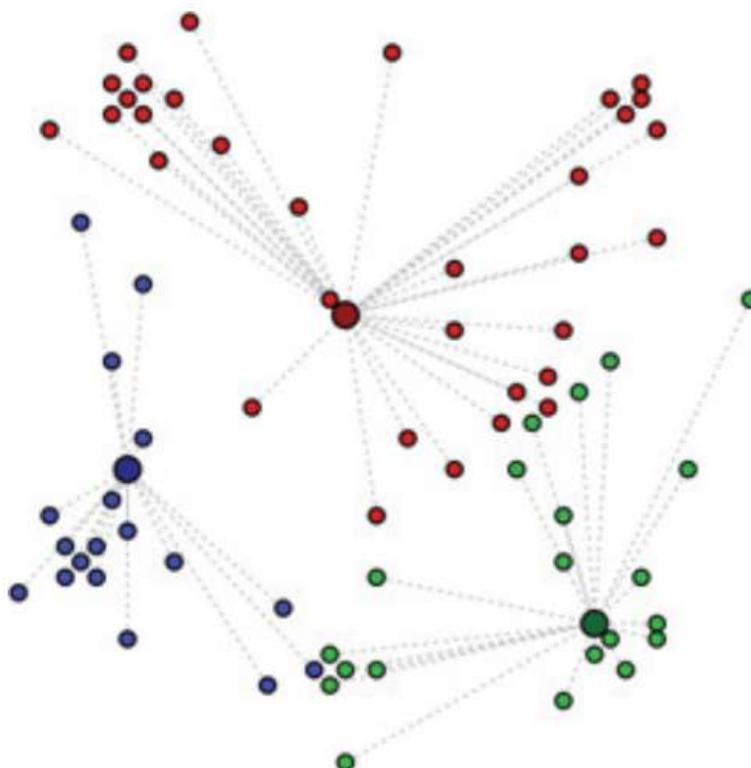
#### 3.3.3.1 K-means

O objetivo do algoritmo *K-Means* é agrupar conjuntos de dados em  $k$  grupos diferentes organizados pelas similaridades entre esses dados. Esse algoritmo é

considerado rápido e escalável, característica importante na maioria dos sistemas ambientados pela tecnologia Internet das Coisas.

Essa técnica de análise, a partir de uma coleção de  $n$  atributos e da escolha de um valor para  $k$ , identifica um número  $k$  de agrupamentos de elementos baseado na proximidade com o centro dos grupos. O centro por sua vez, é determinado pela média aritmética dos atributos de cada grupo.

**Figura 7 - Agrupamentos K-means para  $k = 3$**



Fonte: Dietrich (2015)

A Figura 7 ilustra três agrupamentos distinguidos por pequenas circunferências de cores diferentes, onde as de tamanho maior representam os centros de seus grupos, formados em conjunto com as de tamanho menor, ligadas por uma linha pontilhada.

## 4 METODOLOGIA

Este capítulo trata da descrição dos métodos utilizados na pesquisa que se subdivide em duas partes, onde a primeira apresenta a pesquisa bibliográfica sobre os algoritmos, modelos e arquiteturas IoT, enquanto a segunda parte aborda a implementação de um sistema *Web* envolvendo as características da temática estudada neste trabalho.

### 4.1 Métodos para levantamento e análise da literatura

Este trabalho possui como principal foco teórico a revisão bibliográfica, na qual são analisados artigos, livros, dissertações e teses que dialogam com assuntos ligados ao tema. A seleção do material a ser analisado foi reflexo de sua relevância para o estudo, assim também como foi considerado o período em que esse foi disponibilizado, de forma que artigos mais atuais foram prioritários na pesquisa, já que este estudo em sua essência envolve o surgimento de novas tecnologias. Além disso, foram realizadas pesquisas baseadas em estudos que têm como foco aplicações IoT e análise de dados. A base de dados da pesquisa eletrônica é referenciada com maior foco pelas bibliotecas digitais<sup>12</sup>: ACM Digital Library, IEEE Xplore, ScienceDirect – Elsevier e SpringerLink. A Tabela 1 descreve os critérios para inclusão e exclusão de estudos na pesquisa.

**Tabela 1 - Critérios para levantamento bibliográfico**

<b>Critérios de Inclusão</b>	<b>Critérios de Exclusão</b>
Estudos que focam seus temas em IoT e Análise de Dados	Estudos que não estão relacionados aos objetivos da pesquisa
Artigos, Dissertações, Teses e Livros	<i>blogs</i> e sites informais
Trabalhos escritos em inglês ou português	Trabalhos publicados antes de 2014

Fonte: Próprio autor (2020)

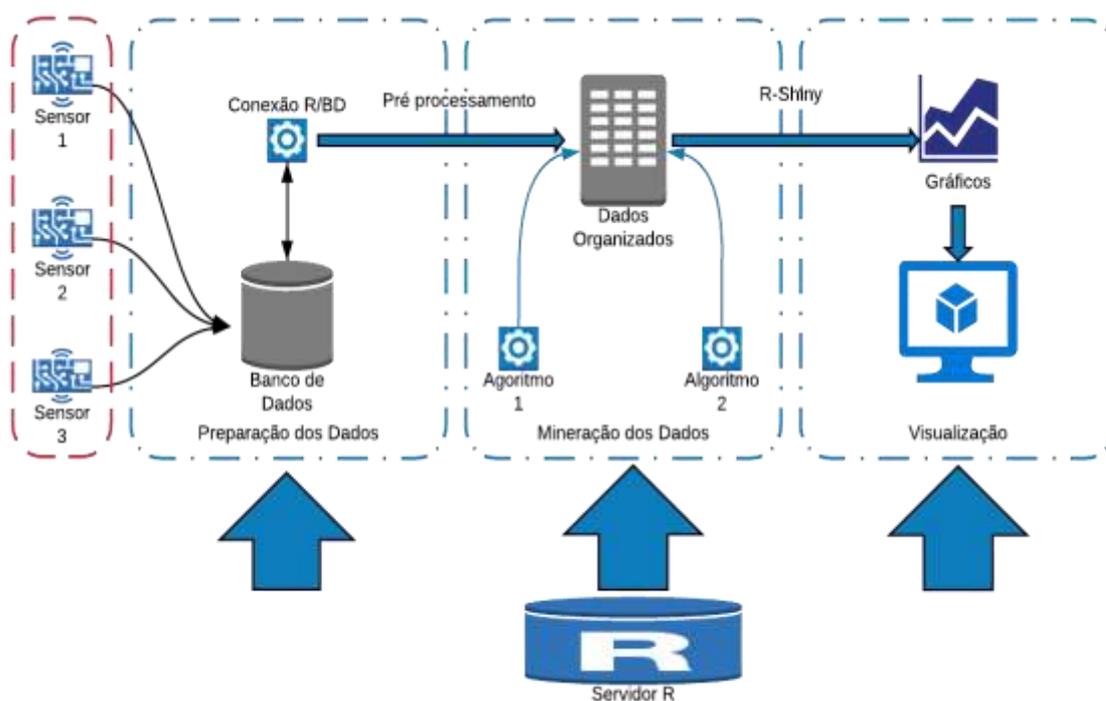
<sup>12</sup> Bibliotecas que possuem inúmeros e relevantes trabalhos, disponíveis em: <https://dl.acm.org>, <https://ieeexplore.ieee.org> e <https://www.sciencedirect.com>.

A partir disso, foram feitos estudos comparativos entre alguns algoritmos de análise de dados comumente utilizados em sistemas baseados em conceitos da IoT. Para realizar esse estudo, foi realizada também uma pesquisa quantitativa para selecionar os algoritmos mais comuns encontrados. Dentre os algoritmos destacam-se o *Support Vector Machine (SVM)*, *K-Neares Neighbors (KNN)*, para classificação, K-Means para agrupamento, ARIMA e Holt-Winters para séries temporais.

## 4.2 Métodos para a implementação do protótipo

Além da parte teórica construída através do levantamento bibliográfico, também foi realizada a implementação do protótipo baseado no tema discutido na pesquisa. Nesse sentido, propõe-se desenvolver e implementar um sistema *Web* para análise de dados de variáveis ambientais, tais dados são produzidos por sensores ligados a seus dispositivos microcontroladores. Já o gerenciamento e a interface do sistema foram feitos através do *software* estatístico R, possibilitando a extração de informações e padrões relevantes dos dados coletados.

**Figura 8 - Visão geral do Modelo Implementado**



Fonte: Próprio autor (2020)

A Figura 8 mostra as principais seções no desenvolvimento de um modelo geral para a implementação do caso de uso proposto neste estudo.

Os *softwares* livres mais utilizados para análise de dados são: *Python* e *R*. Para esse trabalho decidiu-se pela utilização do *R* por ser uma linguagem de propósito específico, mais diretamente ligada à estatística e por possuir um *framework* para criação de um sistema web sem a necessidade de usar outras linguagens no modelo de arquitetura, assim como não há a necessidade de se usar um grande número de bibliotecas externas. Em todo o sistema, o *software R* se utiliza de seus próprios serviços, desde a preparação dos dados em si até a visualização dos dados pelo usuário final.

#### **4.2.1 Preparação dos Dados**

A fonte de dados utilizada nesta aplicação é constituída de um conjunto de sensores que capturam dados como temperatura, luminosidade e umidade a cada dez minutos. A estrutura detalhada do *hardware* microcontrolador do sistema foi abstraída, tendo em vista que o objetivo da pesquisa compreende unicamente a parte relativa à análise de dados dentro da aplicação. O sistema gerenciador do banco de dados utilizado foi o *MySQL* e para atuar junto ao banco foi utilizada a biblioteca do *software R*, *RMySQL*. Dessa forma o servidor consegue trabalhar com os dados persistidos no banco utilizado. Assim, pode-se aplicar métodos característicos da própria linguagem padrão *R*, para fazer o pré-processamento dos dados.

#### **4.2.2 Mineração dos Dados**

Após o pré-processamento dos dados, os algoritmos de mineração de dados podem ser aplicados de forma direta no mesmo ambiente para o tratamento desses modelos, o ambiente utilizado foi o *RStudio*, atualmente a *IDE* mais utilizada para programar na linguagem escolhida. Na implementação proposta foram utilizados dois algoritmos para previsão em séries temporais, *Holt-Winters* e *ARIMA*.

### **4.2.3 Visualização dos Dados**

A visualização dos dados, após a aplicação dos modelos de séries temporais, pode ser gerada através do uso do *framework R-Shiny*. Dessa forma pode-se compilar elementos *HTML* e *CSS* automaticamente, gerando assim páginas *web* através de códigos em R, possibilitando a inserção de gráficos de forma dinâmica e sem recursos externos, permitindo ao usuário observar os dados tratados no servidor, além de poder exportar informações, fazer *download* de gráficos, receber informações via correio eletrônico, dentre outras funcionalidades.

## 5 DESENVOLVIMENTO DO PROTÓTIPO

Este capítulo visa descrever o desenvolvimento do protótipo proposto neste trabalho, abrangendo informações sobre materiais, métodos mais detalhados e implantação do sistema.

### 5.1 Materiais e Ferramentas

Esta seção descreve os materiais e ferramentas utilizados na construção do sistema *Web* desenvolvido como parte de estudo do trabalho em questão, destacando os principais componentes físicos e de *software* responsáveis pelo funcionamento da aplicação.

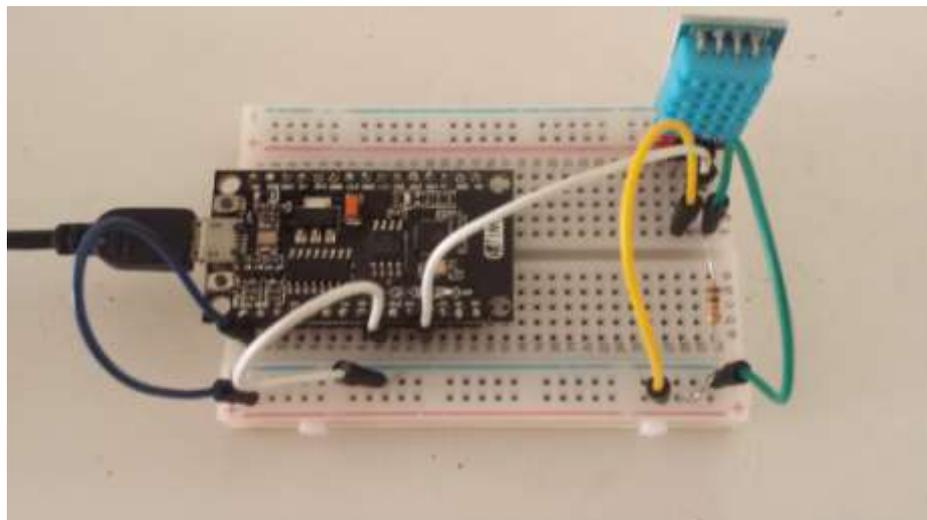
#### 5.1.1 Sensores e Microcontrolador

Uma característica importante da aplicação *Web* para análise de dados, desenvolvida no contexto da Internet das coisas, é o fato de que os dados são gerados através de medidas de sensores reais, tal condição tem influência em todas as partes do sistema.

Os sensores utilizados no protótipo desenvolvido foram divididos em dois grupos. Cada um desses grupos possui o sensor AM2302-DHT22/11, capaz de medir temperatura e umidade do ambiente, bem como o LDR (*Light Dependent Resistor*), responsável por medir o índice de luminosidade. Tais grupos, dispostos em lugares diferentes, se conectam de forma independente ao sistema, transmitindo seus valores de temperatura em graus Celsius (°C), umidade e luminosidade em porcentagem.

No sistema desenvolvido, as unidades de microcontroladores são compostas por placas *NodeMCU*, tal placa de desenvolvimento é produzida com um chip *ESP8266*, permitindo a criação de uma comunicação com outros sistemas via *Wi-Fi*. A placa descrita ainda possui uma interface *USB-Serial* que permite que desenvolvedores programem o comportamento do microcontrolador através dessa porta. Nesse protótipo as unidades foram programadas utilizando-se a interface de desenvolvimento de ambiente *Arduino IDE* versão 1.8.1.

**Figura 9 - Exemplo de um dispositivo IoT**



Fonte: Próprio autor (2020)

A Figura 9 exibe um exemplo de uma estação que nesse caso captura dados de temperatura e umidade, esse dispositivo possui conectividade *Wi-Fi* e através dela envia os dados dos sensores ao servidor em questão.

### **5.1.2 Servidor**

O sistema reside em um servidor Ubuntu 17.04 responsável por hospedar os elementos de funcionamento da aplicação, os componentes que viabilizam o desenvolvimento desta aplicação *Web* são baseados na *web stack*<sup>13</sup> *LAMP* (*Linux Apache MySQL PHP*).

### **5.1.3 Shiny**

O *framework Shiny* é responsável por permitir o desenvolvimento de aplicações *Web* utilizando a linguagem *R*, dessa forma é possível combinar o processamento de modelo de dados utilizando essa linguagem de forma alinhada com a apresentação de páginas *web* modernas. No protótipo desenvolvido, os elementos *HTML* (*Hypertext Markup Language*), *CSS* (*Cascading Style Sheets*) e *JavaScript* são gerados diretamente através de código em *R*, executado na *IDE RStudio* através do pacote *Shiny*.

---

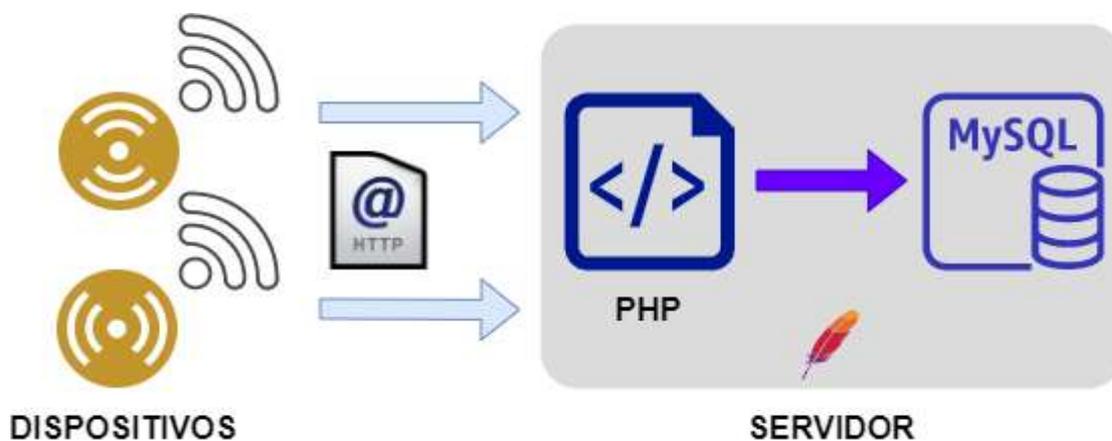
<sup>13</sup> Coleção de softwares responsáveis por preparar e implementar aplicações *web*.

Dessa forma, tanto a parte de interface do sistema, quanto o processamento e controle de funcionalidades são feitas exclusivamente pelo software estatístico, isso permite que o desenvolvimento seja realizado com o foco em uma linguagem apenas.

## 5.2 Dados Coletados

O sistema possui um servidor de banco de dados *MySQL* que é alimentado por dados coletados a partir de variáveis ambientais capturadas pelos sensores já descritos. A entrada de dados no banco do sistema é realizada após o envio de informações das medidas obtidas pelos sensores através de requisições *REST* (*Representational State Transfer*), essas, por sua vez, são realizadas pelo microcontrolador utilizado na aplicação desenvolvida.

**Figura 10 - Arquitetura para coleta de dados**



Fonte: Próprio autor (2020)

A Figura 10 representa a estrutura para a coleta de dados realizada pelos dispositivos IoT.

Ao longo do desenvolvimento, algumas tabelas foram elaboradas considerando cada grupo de sensores do sistema, nomeados de estações. Assim, as tabelas foram enumeradas e descritas de acordo com suas características correspondentes.

**Tabela 2 – Descrição da tabela ESTACAO\_001**

<b>Campo</b>	<b>Tipo</b>
ID	int
DATA_HORA_REGISTRO	timestamp
TEMPERATURA	double
UMIDADE	double
LUMINOSIDADE	double

Fonte: Próprio autor (2020)

A Tabela 2 exhibe um exemplo de descrição de um dos esquemas do sistema, mais especificamente a que está associada à estação 001.

### **5.2.1 Pré-Processamento**

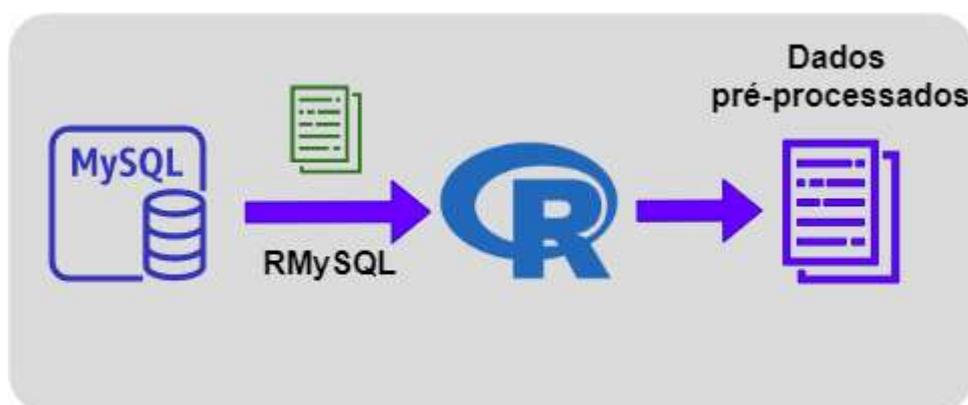
Os dados capturados pelos sensores são gerados com uma frequência de dez minutos, tendo em vista que as medidas monitoradas, em um ambiente naturalmente aberto e livre de interferências externas, tendem a não variar de forma significativa em intervalos de tempo menores que o referido. Nesse período, os valores de temperatura, umidade e luminosidade medidos são adicionados ao banco de dados relacional do sistema, onde posteriormente serão tratados de modo a servir às funcionalidades da aplicação.

O pré-processamento dos dados é necessariamente realizado no sistema, principalmente pelo fato de que esses são provenientes de fontes reais, automáticas e sujeitas a falhas no tempo por diversos motivos, as falhas podem ser causadas por erros de calibração do sensor, que pode resultar em uma medição diferente da situação real verificada, falta de energia, problemas na voltagem do dispositivo, danos físicos, problemas de conexão com a rede conectada ao sistema, que podem provocar a não aferição do dado no período determinado. Tais fatores descritos acabam por demandar que o sistema – responsável por executar análise de dados nesse grupo de valores obtidos – faça operações de limpeza, que tratam de ajustar os valores não devidamente capturados pelos sensores, no conjunto obtido.

As informações relativas aos dois grupos de sensores são organizadas em tabelas dentro de um esquema do banco de dados utilizado no desenvolvimento da aplicação. Assim, esse conjunto de valores é manipulado pelo sistema através do

*software* estatístico R, por meio da biblioteca *RMySQL*, sendo possível tratar os valores de forma antecipada, agrupando dados ausentes e corrigindo de forma adequada os resultados que se mostrem claramente incorretos, obtidos de modo a evitar distorções nos modelos de análise a serem aplicados, como mostra a Figura 11.

**Figura 11 - Estrutura de pré-processamento dos dados**



Fonte: Próprio autor (2020)

Para abordar o problema específico relativo à ausência de algum dado durante uma medição, pelos motivos já descritos no início deste capítulo, foi adotada uma técnica de preenchimento utilizando o valor médio, que consiste em analisar os valores faltantes e substituí-los pela média dos valores imediatamente antes e depois do valor ausente.

### 5.3 Linguagens de Programação

A aplicação desenvolvida neste estudo faz uso de algumas linguagens de programação, cada uma responsável por alguma funcionalidade ou configuração do sistema. A linguagem C++ é utilizada através da plataforma Arduino para adequar os microcontroladores de forma a comunicar com os sensores acoplados, obtendo os valores e enviando-os via requisições *HTTP*.

Para a conexão entre o envio de informações dos sensores e o servidor onde a aplicação reside, fez-se uso da linguagem *PHP* que é responsável por identificar as medições expostas pelos sensores e transportar esses dados diretamente ao banco de dados *MySQL* do sistema. Ainda foi utilizada para criação de algumas outras funcionalidades como login e redirecionamento de páginas *Web*.

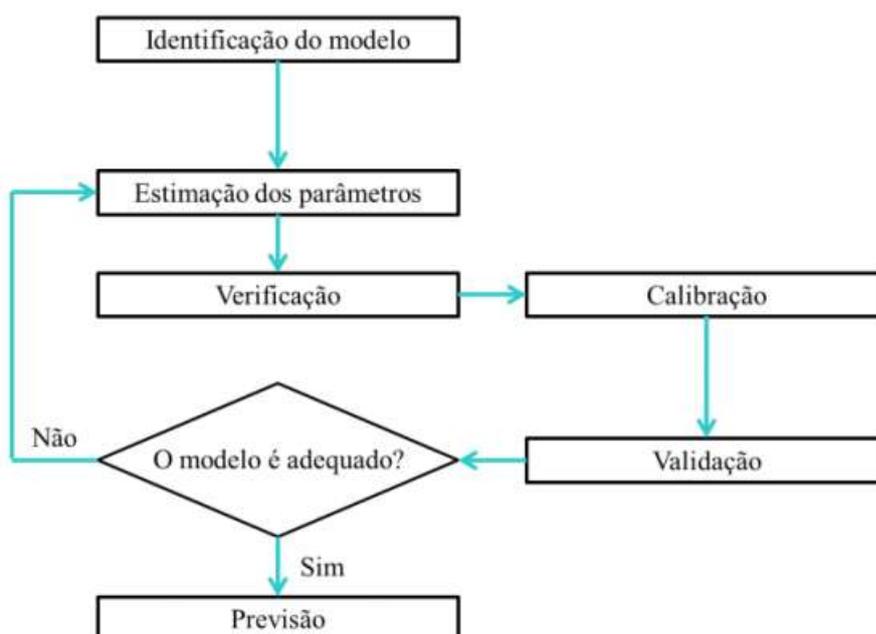
Além das linguagens já descritas, todo o restante do sistema foi composto sob o uso da linguagem R, com o auxílio de pacotes e bibliotecas até mesmo para renderizar a interface da aplicação.

### 5.3.1 Aplicação do Modelo de Previsão e Implementação

A aplicação do modelo estatístico sobre dados trabalhados no sistema é feita através do software *IDE RStudio*. Os algoritmos utilizados foram o *ARIMA* e o *Holt-Winters*, por serem algoritmos bastante comuns quando o padrão da análise se comporta tal qual séries temporais, aplicando-se neste caso. A linguagem R possui bibliotecas, tais como a *forecast*, que foi utilizada para executar esses algoritmos através de métodos com parâmetros customizáveis, que possibilitam a substituição dos critérios de execução desejados diretamente no código.

Sendo tais modelos aplicados diretamente utilizando o *framework Shiny*, a implementação desses na aplicação se dá de forma automática, já que o *server side*<sup>14</sup> e *user side*<sup>15</sup> são definidos diretamente no desenvolvimento do código da aplicação *Web* do sistema.

**Figura 12 - Fluxograma da metodologia Box-Jenkins**



Fonte: Pinto & Lima & Zanetti (2015)

<sup>14</sup> Parte da aplicação onde se desenvolvem as operações executadas pelo servidor em uma rede cliente-servidor.

<sup>15</sup> Parte da aplicação controlada por operações realizadas pelo cliente em uma rede cliente-servidor.

A metodologia Box-Jenkins, apresentada na Figura 12, serviu de base para a implementação do algoritmo *ARIMA* de séries temporais. Certas condições desse método de previsão foram realizadas de forma automática, tendo em vista a arquitetura do sistema que exige tal procedimento. Assim, o modelo de previsão escolhido dentre os utilizados se baseia naquele que possui um menor índice de *RMSE* (*Root-mean-square deviation*), que é um dos parâmetros utilizados para medir a qualidade do modelo.

A implementação do resultado da execução tanto dos modelos de previsão, quanto das análises descritivas, é realizada com o auxílio do *framework Shiny*, utilizado nesta aplicação por ser a ferramenta *open source* mais comum e completa em sistemas *web* na linguagem R, sendo possível a partir disso o acesso ao sistema por parte do usuário.

**Figura 13 - Arquitetura de implementação de modelos do Sistema**



Fonte: Próprio autor (2020)

A Figura 13 apresenta o esquemático da relação cliente-servidor, assim como dos atores envolvidos nesse processo que compreende a execução de algoritmos específicos e a visualização dos resultados obtidos.

## 6 RESULTADOS

Este capítulo visa discutir os resultados alcançados tanto na pesquisa feita em forma de revisão literária, quanto no desenvolvimento da prova de conceito, descrito no Capítulo 5.

### 6.1 Desafios e Oportunidades

Aplicações que requerem um sistema integrado com análise de dados e IoT enfrentam desafios que em certa medida podem ser considerados intermináveis, já que a evolução da tecnologia no tempo acaba por estender continuamente tais tarefas relacionados a certas condições de destaque, estão entre elas principalmente: segurança, sociedade, novas formas de negócio e a indústria. Essas áreas passam por mudanças contínuas e demandam novas formas de aproximação a todo tempo.

Como dados são a principal matéria em sistemas responsáveis por extrair conhecimento e contribuir na tomada de decisão, qualquer condição direcionada a esses terá impacto direto no sistema como um todo, seja na sua infraestrutura, nos algoritmos de análise ou nos dispositivos embarcados. Diferentemente dos processos de coleta de dados usuais, o contexto IoT indica desafios como a comunicação entre máquinas de diferentes arquiteturas, latências, protocolos de transmissão, fontes heterogêneas, frequências independentes, etc.

Outra questão está na natureza operacional, sabendo-se que certos aspectos do ambiente apresentado como troca de sensores, interrupção de medições e mudanças na topologia de redes são situações não muito incomuns, a necessidade de adequação de algoritmos de mineração de dados relativos a esses processos se apresenta de forma urgente, exigindo uma alta dinamicidade no uso de técnicas de análise que possam garantir uma tomada de decisão coerente dentro da aplicação que está alocada. O processamento de dados na forma tradicional ocorre muitas vezes na nuvem, em servidores dedicados, não é o caso quando o sistema está inserido em um ambiente de Internet das Coisas, onde muitas vezes o processamento se dá na borda, ou seja, nos próprios dispositivos de *hardware*.

Os algoritmos estudados nesta pesquisa não possuem eficiência perfeita, apesar de bem estruturados e com bons resultados, o que os fazem serem amplamente utilizados, como o *KNN*, *K-Means* e *SVM*, há sempre situações especiais em que um deles terá melhores resultados em detrimento a outros, ainda mais em ambientes caóticos que exigem mais da velocidade, acurácia e robustez, como é o caso de sistemas imersos no paradigma IoT. O *KNN*, por exemplo, requer a alocação da parcela de dados dedicada para o treino do modelo, fazendo com que possa ocorrer problemas de escalabilidade, diferentemente do *K-Means*, que apresenta aspectos de rapidez e boa escalabilidade, mais utilizado em processos de clusterização. Já os algoritmos de *SVM* são aplicáveis em várias arquiteturas de sistemas IoT, pois são eficientes em termos de uso de memória e lidam bem com dados em massa. Assim como os algoritmos *ARIMA* e *Holt-Winters*, que possuem ampla aplicabilidade em variados projetos ligados a séries temporais.

Além disso, um dos principais desafios atualmente enfrentados está na área de segurança, tendo em vista que grande parte dos dados coletados sob o contexto utilizado são características pessoais ou que exigem privacidade e comedimento. Assim, há um esforço também voltado para o uso de técnicas como a anonimização, criptografia, identificação temporária, dentre outras, que ajudam a diminuir certos problemas inerentes ao sistema.

Baseado na pesquisa realizada, fica claro uma tendência de que a Internet das Coisas, juntamente com técnicas de análise de dados, tem uma alta capacidade de gerar grandes benefícios em diversos setores como transporte, saúde, domótica, indústria, etc. E que as oportunidades a serem aproveitadas têm o potencial de conduzir a sociedade para um novo e melhor estilo de vida.

## **6.2 Prova de Conceito**

Este tópico descreve os resultados alcançados a partir do desenvolvimento da prova de conceito apresentada.

### 6.2.1 Cenário e Funcionalidades Implementadas

O desenvolvimento do experimento foi realizado com o objetivo de viabilizar o procedimento e a construção de um sistema *Web*, voltado para o contexto da Internet das Coisas.

O cenário da realização dos testes envolveu dois grupos de sensores independentes capturando informações de variáveis ambientais descritas nos capítulos anteriores, tais grupos foram posicionados em um ambiente aberto, expostos a modificações naturais de clima. Os valores considerados para esse estudo foram medidos de forma ininterrupta entre os dias 15/05/2017 e 06/11/2017.

**Figura 14 - Tabela de uma estação de medição**

ID	TIMESTAMP	TEMPERATURA	UMIDADE	LUMINOSIDADE
83	2017-08-16 05:55:48	25.6	56.5	5
84	2017-08-16 06:05:49	25.6	56.6	5
85	2017-08-16 06:15:49	25.6	56.5	12
86	2017-08-16 06:25:49	25.6	56.5	11
87	2017-08-16 06:35:50	25.6	56.4	16
88	2017-08-16 06:45:50	25.6	56.5	29
89	2017-08-16 06:56:05	25.7	56.5	40
90	2017-08-16 07:06:06	25.8	56.4	41
91	2017-08-16 07:16:06	26	56.1	41
92	2017-08-16 07:26:06	26.2	55.6	39
93	2017-08-16 07:36:07	26.4	55.8	32
94	2017-08-16 07:46:07	26.6	55.3	37
95	2017-08-16 07:56:07	26.8	55.2	33
96	2017-08-16 08:06:08	27	54.9	29

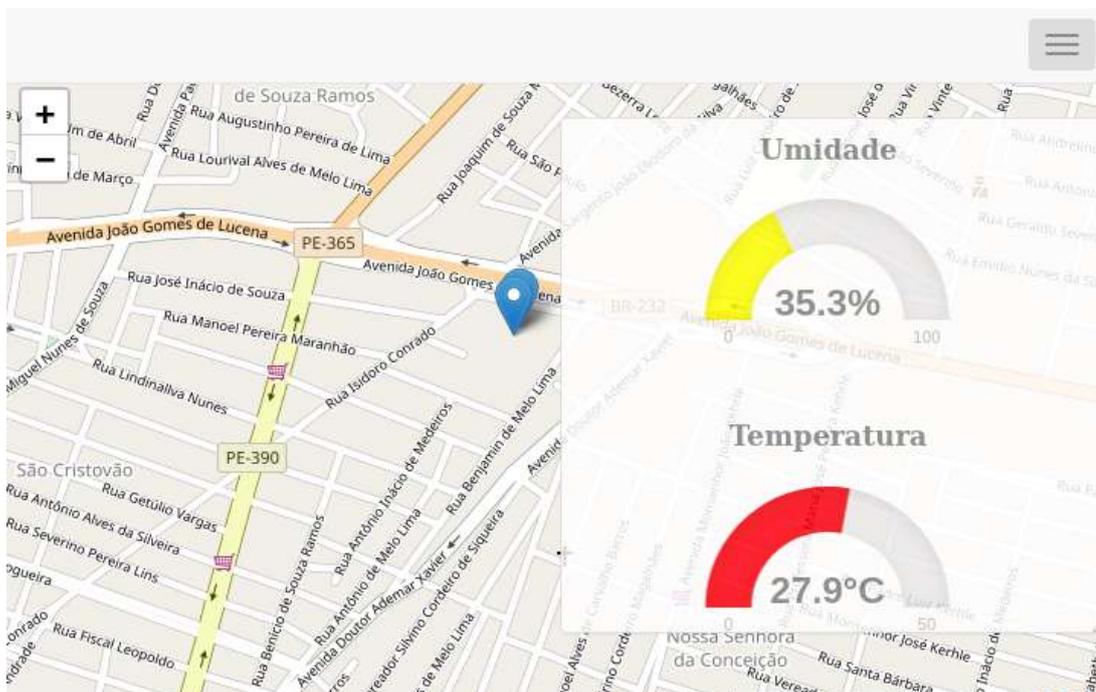
Fonte: Próprio autor (2020)

A Figura 14 exibe alguns dos valores capturados, já inseridos na base dados, em uma das estações de sensores do sistema.

A página *web* de apresentação do sistema é compilada através de uma das funcionalidades da aplicação em questão, que trata da apresentação dos últimos dados capturados pelos sensores, a visualização dos dados acontece dinamicamente na tela após curtos intervalos de tempo.

A Figura 15 exibe a interface inicial do protótipo construído, na qual através de gráficos do tipo *Gauge* é possível perceber os níveis das variáveis medidas, esses gráficos também são atualizados de forma a garantir a percepção do último valor medido pelos sensores, facilitando a percepção das variações.

**Figura 15 - Dashboard para visualização de aferições**



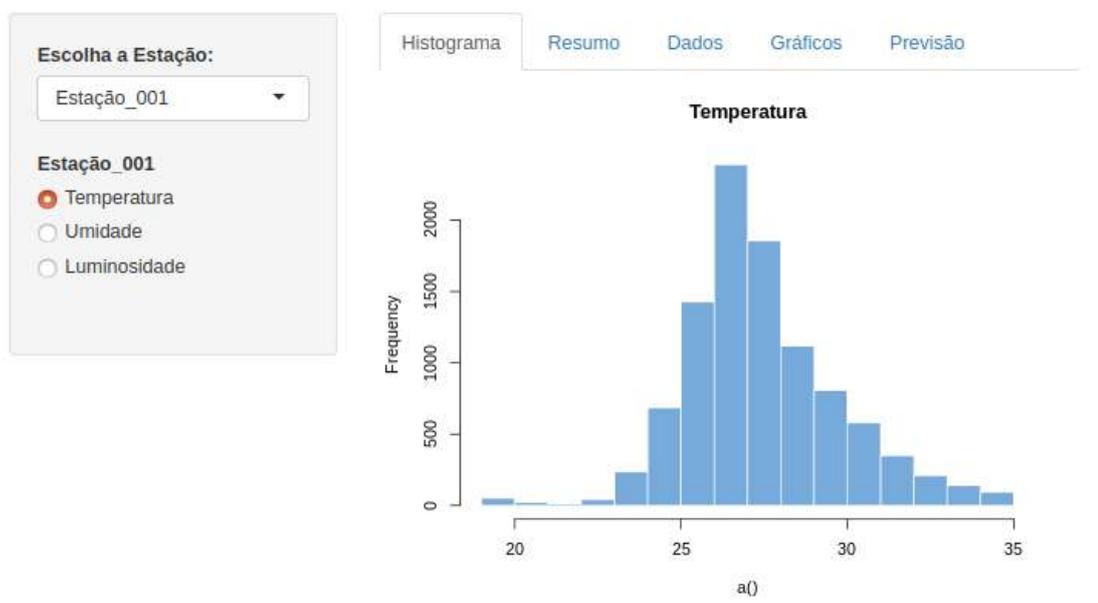
Fonte: Próprio autor (2020)

O sistema desenvolvido possui algumas funcionalidades importantes para o cumprimento do objetivo de sua construção, dentre elas pode-se destacar a visualização dos dados capturados em tempo real, representações gráficas em forma de histogramas e gráficos de linha aplicados aos valores analisados, sistema de login, *download* dos dados em formato *csv*<sup>16</sup> e análise de previsões dos dados de temperatura com base na aplicação dos modelos de predição para séries temporais.

### **6.2.2 Análise Exploratória dos Dados**

Através da página dedicada à exploração dos dados é possível verificar gráficos estatísticos como histogramas, gráficos de linha e *boxplots* que são exibidos para as variáveis escolhidas através de seleções de *radio button's* e *select menu's* que são elementos de controle gráfico que permitem ao usuário escolher opções predefinidas da interface da aplicação.

<sup>16</sup> Arquivo de texto que serve para armazenar dados tabulares usando ponto e vírgula como separador dos valores.

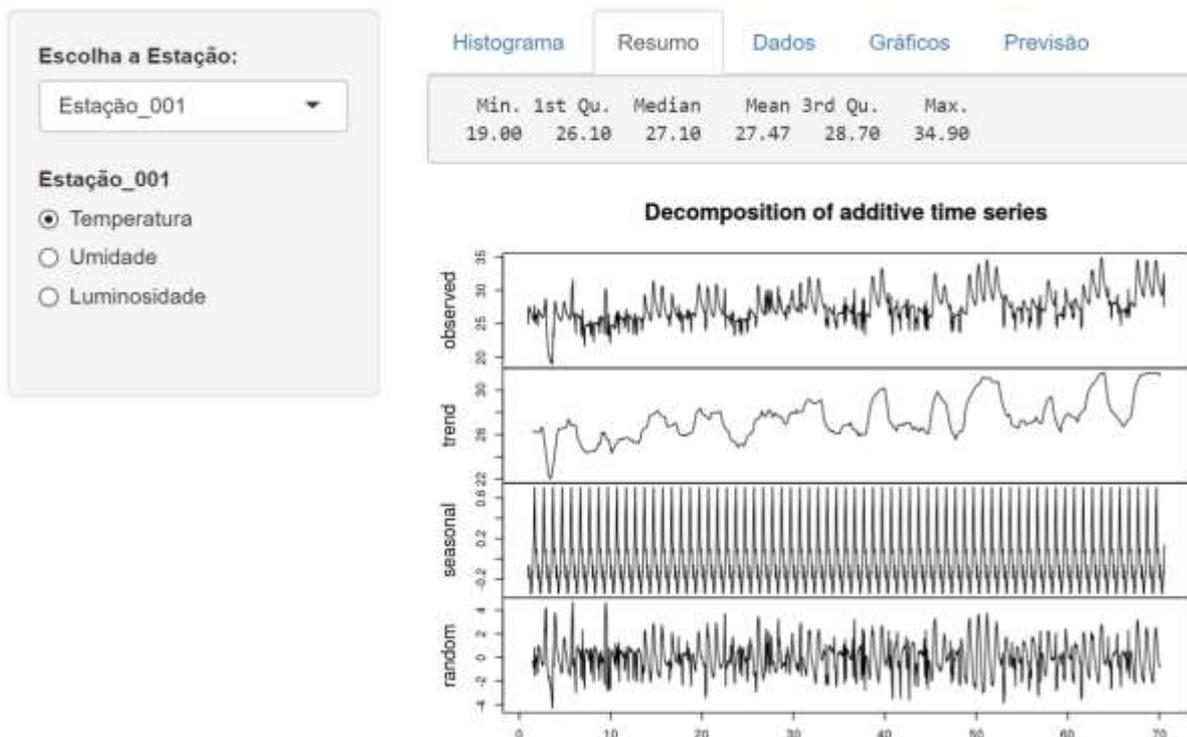
**Figura 16 – Histograma dos valores de temperatura**

Fonte: Próprio autor (2020)

Consultas informativas sobre o comportamento dos valores capturados pelos sensores podem ser realizadas como mostrado na Figura 16, através de um histograma que analisa a frequência das medições de temperatura durante o período de aferições dos dados.

A visualização descritiva dos dados também pode ser exibida de outras formas, a decomposição de elementos da série temporal pode ser percebida na Figura 17 e um gráfico de linha na Figura 18.

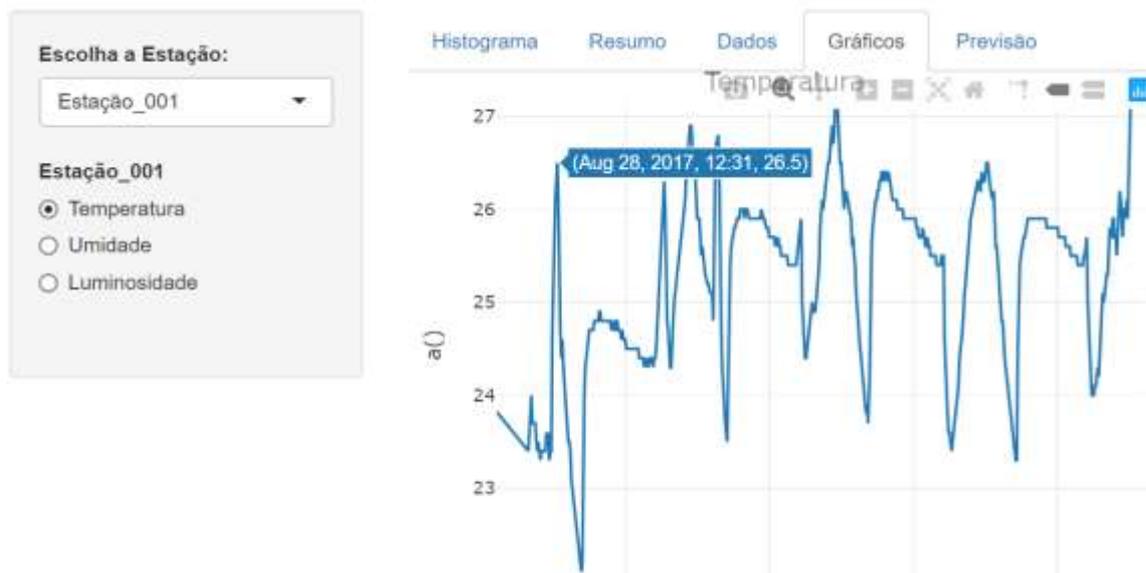
**Figura 17 - Resumo de elementos da Série Temporal**



Fonte: Próprio autor (2020)

A Figura 17 apresenta uma funcionalidade do sistema que permite a exibição de características importantes da série temporal em questão, na qual o parâmetro selecionado é a temperatura, assim é possível perceber o comportamento da tendência, sazonalidade e aleatoriedade, além de percebê-la por completo em seu valor normalmente observado.

Além disso, nota-se na Figura 18, a representação desses dados observados em forma de gráfico de linha, onde pode-se aferir o dia, hora e valor que foi medido naquele instante, também é possível fazer ajustes no gráfico com o intuito de estudar melhor os dados exibidos.

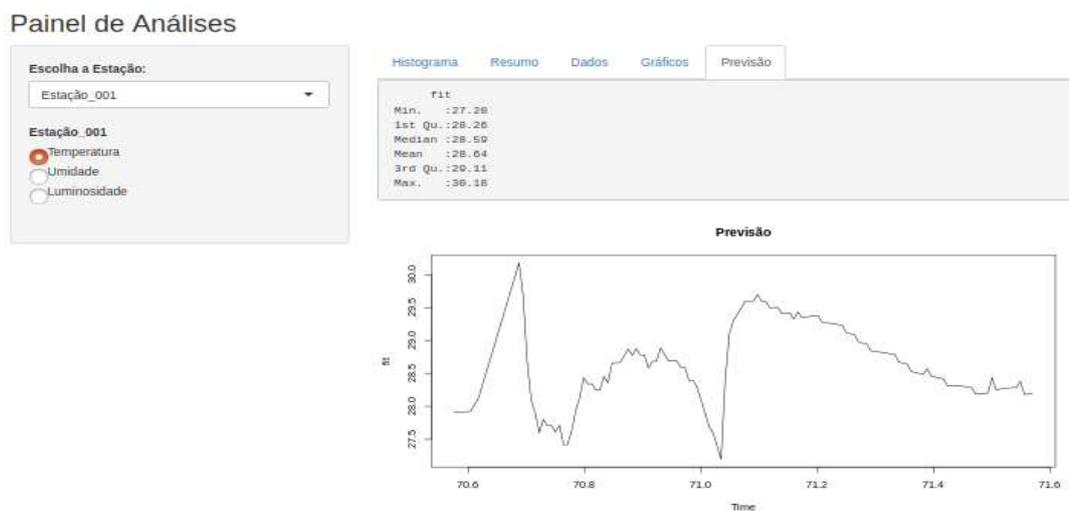
**Figura 18 - Visualização dos dados em forma de gráfico de linha**

Fonte: Próprio autor (2020)

Os dados observados ainda podem ser adquiridos em formato *csv*, de modo a fornecer ao usuário a possibilidade de exportá-los para a máquina utilizada, com o intuito de garantir a melhor observância dos dados da forma que se julgue necessária.

### **6.2.3 Aplicação do Modelo Preditivo**

O resultado da implementação do modelo de previsão é apresentado através de um gráfico de linha como é possível perceber na Figura 19.

**Figura 19 – Gráfico dos valores previstos**

Fonte: Próprio autor (2020)

O gráfico presente na figura exibe os valores previstos de temperatura para os sete dias seguintes, representando a saída da execução de um dos algoritmos de previsão para séries temporais executado após a ação do usuário.

Após a coleta, pré-processamento, escolha de modelo e execução, pode-se obter um resultado visual de suas operações e garantir a sua observância através da interface de usuário do sistema, o que facilita a compreensão dos mecanismos utilizados e contribui para o surgimento de ideias que podem ser aplicadas em diversas áreas seguindo tal arquitetura.

## 7 CONSIDERAÇÕES FINAIS

Neste trabalho foi constatado que o desenvolvimento acelerado de tecnologias, especialmente relativas a sistemas embarcados, nos últimos anos serviu de agente catalisador para o número de dispositivos conectados à Internet. A partir disso, diversas formas de análise dos dados se tornaram necessárias em um contexto que envolve esses sistemas e a estrutura para a Internet das Coisas (IoT), que diferentemente das condições de análise convencionais, traz uma série de novos desafios, dificuldades e riscos a serem enfrentados, por isso a importância apontada nesta pesquisa de se estudar aspectos ligados a esse novo paradigma que surge com alta capacidade de revolução nos meios sociais e econômicos em todo o mundo.

O estudo realizado descreveu alguns métodos importantes e comuns para análise de dados, onde foi possível relacionar suas restrições e oportunidades existentes em um ambiente partícipe das condições apresentadas, isso se tornou possível a partir de uma pesquisa baseada no contexto descrito e no desenvolvimento de uma prova de conceito que abrange de forma geral o tema do trabalho.

A percepção dos modelos analíticos existentes foi apresentada de modo descritivo sobre alguns dos principais algoritmos de mineração de dados e com perspectivas discricionárias no que se refere às aplicações a que tais sistemas estão inseridos.

Durante o trabalho foi possível identificar limitações, barreiras e compreensões importantes sobre a realização da análise de dados coletados por sistemas embarcados em um sistema característico da infraestrutura IoT, além de aplicar também técnicas e algoritmos de forma empírica, o que implicou em uma experiência prática e essencial para um melhor entendimento do escopo do tema.

Os dados apresentados neste estudo foram coletados no Instituto Federal de Educação, Ciência e Tecnologia de Pernambuco (IFPE), assim como a montagem do *hardware*, incluindo sua alimentação, organização, calibragem e conexão de sensores, além da montagem do servidor *Linux* utilizado, onde são executadas todas as rotinas referentes ao pacote *Shiny*, pertencente ao software R e responsável principalmente pela implantação de uma forma geral de partes cruciais

do sistema, como a execução dos modelos de análise de dados, possibilitando também sua própria visualização para o usuário final.

Algumas dificuldades foram encontradas no presente estudo, dentre elas, o tempo reduzido para a coleta de dados, a baixa capacidade do servidor em executar algoritmos que exigem uma quantidade maior de recursos computacionais, a impossibilidade de explorar uma quantidade massiva de dados devido à falta de recursos e matéria prima, além da quantidade razoavelmente pequena de estudos que contemplassem diretamente sistemas embarcados, IoT e análise de dados de forma geral em um projeto único.

As contribuições principais deste trabalho se dão na pesquisa apresentada, na disseminação do conhecimento necessário e na apresentação de um caso de uso *open source* que capacita a busca por soluções em diversos problemas das áreas mais distintas, sobre um escopo direcionado à análise de dados em sistemas IoT, permitindo que os espaços de discussão desse tema sejam amplificados, isso porque tal tarefa se mostra de grande importância, tendo em vista o mundo extremamente conectado em que se vive na contemporaneidade.

## 7.1 Trabalhos Futuros

Diante do trabalho apresentado foi possível tecer algumas recomendações a partir das limitações enfrentadas e percepções adquiridas durante o seu processo de construção.

Sugere-se a realização da pesquisa com uma abordagem mais direcionada ao aprendizado de máquina automatizado (AutoML), que trata da automatização dos processos de análise, fazendo com que a interação humana nos processos de modelagem seja menor. Tal condição é capaz de reduzir fortemente o tempo despendido na criação de modelos preditivos, reduzir erros analíticos comuns e dessa forma pular etapas na construção de sistemas IoT, garantindo maior confiabilidade e agilidade ao processo.

Outro ponto importante que pode ser explorado em trabalhos futuros parte da própria infraestrutura da prova de conceito apresentada, faz-se interessante na criação de um sistema *Web* como o apresentado, a utilização de um banco de dados *NoSQL*, tendo em vista que processos que permaneçam ativos por muito tempo

podem necessitar de um poder de escalabilidade enorme dentro da base de dados coletada, tal percepção vai ao encontro de uma das características de sistemas de gerenciamento de bancos desse tipo, como o *MongoDB*. Além disso, sugere-se o estudo e a aplicação de outra linguagem de programação também comum no ambiente da análise de dados, trata-se da linguagem *Python*. Estudos de infraestrutura nos componentes de sistemas embarcados com módulos de conectividade como *RFID*, *Bluetooth* e *ZigBee* são importantes para mesclar diferentes necessidades, já que aplicações diferentes possuem formas de conexão mais ou menos eficazes, a depender de sua estrutura e de seus fins.

Além disso, durante a análise de vários trabalhos que contribuíram com pesquisas envolvendo Internet das Coisas e análise de dados, foi possível perceber o empenho e a evolução de diversos aspectos ligados à melhoria de desempenho de algoritmos, aprimoramento de sistemas embarcados, surgimento de ideias importantes tanto para empresas comerciais como para a sociedade. Porém, algo que parece se apresentar com grande urgência hoje nesse sentido é a questão ética, devido a problemas de privacidade e segurança, tal demanda não pode ficar em segundo plano, por isso sugere-se também um estudo que avalie as condições e problemas éticos relacionados ao avanço dessas tecnologias em específico, enfatizando processos como anonimização de dados e aplicação da recente LGPD (Lei Geral de Proteção de Dados).

Ainda, como reflexo das limitações enfrentadas neste trabalho, recomenda-se a criação de uma plataforma *Web* capaz de tratar da análise de dados como apresentado neste estudo, mas que possa também trabalhar com dados massivos, explorando mais o termo *Big Data*, através de ferramentas como *Scala* e *Spark*, e fornecendo contribuições importantes sobre esse cenário a partir de recursos que hoje ainda possuem certa dificuldade de serem acessados.

## REFERÊNCIAS

AHMED, Ejaz et al. The role of big data analytics in Internet of Things. **Computer Networks**, Elsevier, v. 129, p. 459–471, dec. 2017. ISSN 13891286.

ALAM, Furqan et al. Analysis of Eight Data Mining Algorithms for Smarter Internet of Things (IoT). *Procedia - Procedia Computer Science*, v. 98, p. 437–442, 2016.

BUSTAMANTE, J. **Smart Home Statistics**. Disponível em: <https://ipropertymanagement.com/research/iot-statistics>. Acesso em: 13 de jun. de 2019.

CHEN, F; DENG, P; WAN, J. **Data Mining for the Internet of Things: Literature Review and Challenges**. v. 2015, n. i, 2015.

DIETRICH, David. **Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data** (1ª ed.). Wiley Publishing, 2015.

EVANS, D. **The internet of things: How the next evolution of the internet is changing everything**. v. 1, p. 1–11, 01 2011.

FACELI, K. (2001). **Combinação de métodos de inteligência artificial para fusão de sensores**. Dissertação de Mestrado, São Carlos: Instituto de Ciências Matemáticas e da Computação, Universidade de São Paulo, 2001.

FORBES. **Strong analytics skills 3x more likely to lead to success with internet of things(iot)**. Disponível em: <https://www.forbes.com/sites/louiscolombus/2016/09/21/strong-analytics-skills-3x-more-likely-to-lead-to-success-with-internet-of-things-iot/#d94c09f3563c>. Acesso em 10 jun. 2019.

GE, Mouzhi; BANGUI, Hind; BUHNOVA, Barbora. Big Data for Internet of Things: A Survey. **Future Generation Computer Systems**, Elsevier B.V., v. 87, p. 601–614, 2018. ISSN 0167739X.

IDC – International Data Corporation. **The Growth in Connected IoT Devices**. Disponível em: <https://www.idc.com/getdoc.jsp?containerId=prUS45213219>. Acesso em: 18 de out. de 2020.

IMD. **Smart City Index, 2020**. Disponível em: <https://www.imd.org/smart-city-observatory/smart-city-index/>. Acesso em: 22 de out. de 2020.

ITU – **International Telecommunication Union**. Overview of the Internet of things. Series Y: Global information infrastructure, internet protocol aspects and nextgeneration networks - Frameworks and functional architecture models, p. 22, 2012. ISSN 0018-8646.

JESUS, Tarciso De et al. **Internet das Coisas e sua evolução tecnológica para as Cidades Inteligentes**, 2016.

LIMA, J. R.; CUNHA, D A. **Monitoramento de ambientes especiais aliado ao conceito de internet das coisas (iot)**, 2018.

LANGER, G. A. **Análise de dados aplicada para consciência situacional de pedestres**, 2018.

LI, B. et al. **Aided Diagnosis and Treatment of Coronavirus Disease 2019 (COVID-19)**, vol. 3, 2020.

MAHDAVINEJAD, M. S. et al. Machine learning for internet of things data analysis: a survey. **Digital Communications and Networks**, Elsevier, v. 4, n. 3, p. 161–175, aug 2018. ISSN 2352-8648.

MARJANI, M. et al. **Big IoT data analytics: architecture, opportunities, and open research challenges**. *leeeexplore.lee.org*, p. 5247–5261, 2017.

NAIT, Y. et al. The 7th International Conference on Current and Future Trends of Information and Communication Technologies in Healthcare (ICTH 2017) On the use of IoT and Big Data Technologies for Real-time Monitoring and Data Processing. **Procedia Computer Science**, v. 113, p. 429–434, 2017.

NALLAKARUPPAN, M. K.; KURAMAN, U. S. IoT based Machine Learning Techniques for Climate Predictive Analysis. **International Journal of Recent Technology and Engineering (IJRTE)** ISSN: 2277-3878, v.7 jan. 2019.

PINTO, W.; LIMA, G.; ZANETTI, J. (2015). Estudo de predição e previsão de temperaturas médias diárias da cidade de Cariacica, Espírito Santo, Brasil, utilizando a metodologia de séries temporais. **Revista Gestão Inovação e Tecnologias**. 5. 1881-1895. 10.7198/S2237-0722201500010022.

PIVOTO, D. et al. Scientific development of smart farming technologies and their application in Brazil. **Information Processing in Agriculture**, [s. l.], v. 5, n. 1, p. 21–32, 2018.

QIU, J. et al. A survey of machine learning for big data processing. **EURASIP Journal on Advances in Signal Processing**, v. 2016, n. 1, p. 67, dec 2016. ISSN 1687-6180.

SAMIE, F.; BAUER, L.; HENKEL, J. **IoT Technologies for Embedded Computing: A Survey**, 2016. 10.1145/2968456.2974004.

SCHMITT, V. F. **Uma análise comparativa de técnicas de aprendizagem de máquina para prever a popularidade de postagens do Facebook**, 2013.

SCULLY, P. **IoT Analytics**, 2020. Disponível em: <https://iot-analytics.com/top-10-iot-applications-in-2020/>. Acesso em: 20 de out. de 2020.

TSAI, C. W. et al. Data mining for internet of things: A survey. **IEEE Communications Surveys and Tutorials**, 2014. ISSN 1553877X.

URBANSYSTEMS. **Ranking Connected Smart Cities**, 2020. Disponível em: <http://ranking.connectedsmartcities.com.br>. Acesso em: 22 de out. de 2020.

WITKOWSKI K. Internet of Things, Big Data, Industry 4.0 – Innovative Solutions in Logistics and Supply Chains Management. **Procedia Engineering**, Elsevier, 2017.

ZHOU, Z. et al. Big Data Opportunities and Challenges: Discussions from Data Analytics Perspectives. **IEEE Computational Intelligence Magazine**, v. 42, n. 7, p. 62–74, 2014. ISSN 00431397.