

André Lucas da Costa Soares

Orientador(a): Prof. Me. Ismael de Holanda Leal

Co-orientador(a): Patricia Medyna Lauritzen de Lucena Drumond

# **Revisão Sistemática no Contexto de Detecção de Outliers para Monitoramento Ambiental de Cidades Inteligentes**

Picos - PI

15 de junho de 2019

André Lucas da Costa Soares

Orientador(a): Prof. Me. Ismael de Holanda Leal

Co-orientador(a): Patricia Medyna Lauritzen de Lucena Drumond

## **Revisão Sistemática no Contexto de Detecção de Outliers para Monitoramento Ambiental de Cidades Inteligentes**

Trabalho de Conclusão de Curso submetido ao curso de Sistemas de Informação da Universidade Federal do Piauí (UFPI), como requisito parcial para obtenção do título de Bacharel em Sistemas de Informação. Orientador (a): Prof. Me. Ismael de Holanda Leal

Universidade Federal do Piauí  
Campus Senador Helvídio Nunes de Barros  
Bacharelado em Sistemas de Informação

Picos - PI  
15 de junho de 2019

**FICHA CATALOGRÁFICA**  
**Serviço de Processamento Técnico da Universidade Federal do Piauí**  
**Biblioteca José Albano de Macêdo**

**S676r** Soares, André Lucas da Costa.  
Revisão sistemática no contexto de detecção de outliers para monitoramento ambiental de cidades inteligentes. / André Lucas da Costa Soares. – Picos, PI, 2019.  
41 f.  
CD-ROM: 4 ¾ pol.

Trabalho de Conclusão de Curso (Bacharelado em Sistemas de Informação) – Universidade Federal do Piauí, Picos, 2019.

Orientador(A): Prof. Me. Ismael de Holanda Leal.

1. Monitoramento Ambiental. 2. Cidades Inteligentes. 3. Sensores Tecnológicos - Outliers. I. Título.

**CDD 005**

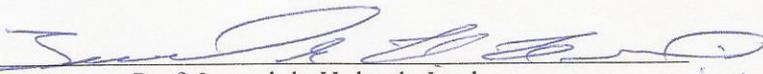
TÍTULO  
REVISÃO SISTEMÁTICA NO CONTEXTO DE DETECÇÃO DE OUTLIERS PARA  
MONITORAMENTO AMBIENTAL DE CIDADES INTELIGENTES

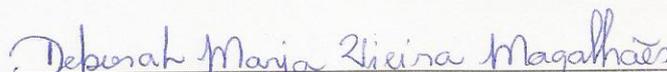
ANDRÉ LUCAS DA COSTA SOARES

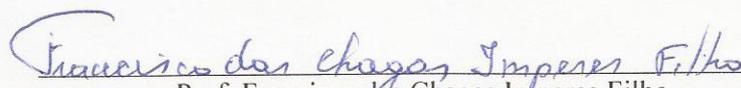
Monografia APROVADA como exigência parcial para obtenção do grau de  
Bacharel em Sistemas de Informação.

Data de Aprovação:

Picos - PI, 12 de JUNHO de 20 19

  
Prof. Ismael de Holanda Leal

  
Profa. Deborah Maria Vieira Magalhães

  
Prof. Francisco das Chagas Imperes Filho

# Agradecimentos

A Deus, que me iluminou e deu a força necessária para concluir essa etapa importante da minha vida.

Ao meu orientador, Ismael de Holanda Leal, e, em especial à minha co-orientadora, Patricia Medyna Lauritzen de Lucena Drumond, pela sua paciência, compreensão, por sempre estar disposta a me ajudar e ter acreditado em mim. Serei eternamente grato por tudo que fez por mim durante toda minha graduação.

Aos meus pais, Antonia Iara da Costa e Francisco das Chagas Soares, à minha irmã, Camila da Costa Soares, e toda a minha família, pelo apoio, carinho, amor e incentivo. Eu amo vocês.

À minha namorada, Fabiana Rocha Sousa, pelas palavras de incentivo, pelo amor e dedicação, por sempre estar comigo e me apoiar em todas minhas decisões. Eu te amo!

Aos meus amigos, em especial aos que formam o Majuca: Davi Luís de Oliveira, Guilherme Vyncius do Vale Miranda, João Marcos de Sousa Rosal, Paulo Henrique Gonçalves Rocha, Tácio de Sousa Moreira e Wildyson Dantas dos Santos, que compartilharam comigo momentos de alegria, tristeza, dúvida, medo e descontração. Sem vocês eu não teria chegado até aqui, obrigado.

*Venho provar que é possível, é  
só dizer glória, amém no lugar  
do seu 'tá' difícil.*

---

*Fabício FBC*

# Resumo

A rápida e crescente urbanização provoca uma série de problemas que trazem prejuízos tanto à vida humana como ao meio ambiente. Nos últimos anos, o conceito de Cidades Inteligentes, do inglês *Smart Cities*, vem emergindo como uma estratégia para mitigar os problemas causados pelo processo de urbanização. Com a utilização de vários sensores de baixo custo, ambientes urbanos são monitorados e dados sobre tais ambientes são coletados. A análise dos dados coletados é uma tarefa crucial para se obter informações relevantes que possam apoiar os governantes no desenvolvimento de políticas públicas. Entretanto, a análise dos dados pode ser afetada pela ocorrência de observações anômalas, os *outliers*. Diversas técnicas para detectar *outliers* tem sido propostas na literatura nos mais variados campos de pesquisa. Dessa forma, este trabalho apresenta uma Revisão Sistemática da Literatura, o qual teve como objetivo realizar o levantamento e apresentação dos métodos empregados na literatura para detectar *outliers* em dados ambientais coletados em centros urbanos. Ao executar a revisão sistemática, foram encontrados oito estudos, que atenderam aos critérios de seleção e às Questões de Pesquisa (QPs). A partir da leitura dos artigos e extração dos dados objeto da pesquisa, foram encontrados nove métodos distintos de detecção de *outliers*, alguns deles formados pela combinação de duas técnicas. Além disso, percebeu-se a preocupação dos pesquisadores com a questão da poluição atmosférica, sendo abordada em cinco dos oito artigos selecionados.

**Palavras-chave:** Revisão Sistemática. Detecção de *Outliers*. Detecção de Anomalias. Cidades Inteligentes. Monitoramento Ambiental.

# Abstract

Rapid and growing urbanization causes a series of problems that harm both human life and the environment. In recent years, the concept of Smart Cities has emerged as a strategy to mitigate the problems caused by the urbanization process. With the use of various low-cost sensors, urban environments are monitored and data on such environments are collected. The analysis of the data collected is a crucial task to obtain relevant information that can support the government in the development of public policies. However, the analysis of the data can be affected by the occurrence of anomalous observations, the outliers. Several techniques for detecting outliers have been proposed in the literature in a variety of research fields. Thus, this paper presents a Systematic Review of Literature, which aimed to carry out the survey and presentation of methods used in the literature to detect outliers in environmental data collected in urban centers. In performing the systematic review, eight studies were found, which met the selection criteria and the Research Questions (QPs). From the reading of the articles and extraction of the data object of the research, were found nine distinct methods of detection of outliers, some of them formed by the combination of two techniques. In addition, the researchers concern with the issue of air pollution was noticed, being addressed in five of the eight articles selected.

**Keywords:** Systematic Review. Outliers Detection. Anomalies Detection. Smart Cities. Environmental Monitoring.

# Lista de ilustrações

Figura 1 – Busca pelo termo “ <i>Smart Cities</i> ” (período: 01/01/2004 a 23/01/2019) .	15
Figura 2 – Elementos da IoT . . . . .	17
Figura 3 – Arquitetura de três camadas para a IoT . . . . .	18
Figura 4 – Exemplos das categorias de <i>outliers</i> . (a) <i>Outlier</i> pontual; (b) <i>outlier</i> coletivo; (c) <i>outlier</i> contextual . . . . .	20
Figura 5 – Quantidade de estudos encontrados por base de dados . . . . .	26

# Lista de tabelas

Tabela 1 – Definições de CI . . . . .	16
Tabela 2 – Categorias e descrições de <i>outliers</i> . . . . .	20
Tabela 3 – Palavras-chave . . . . .	23
Tabela 4 – Versões da <i>string</i> de busca e seus respectivos resultados do teste piloto	24
Tabela 5 – <i>Strings</i> de busca utilizadas . . . . .	25
Tabela 6 – Critérios de seleção de estudos . . . . .	25
Tabela 7 – Aplicação dos filtros sobre os artigos . . . . .	27
Tabela 8 – Artigos selecionados . . . . .	27
Tabela 9 – Principais informações extraídas dos estudos . . . . .	32
Tabela 10 – Métodos de detecção de <i>outliers</i> utilizados . . . . .	34
Tabela 11 – Contextos abordados pelos artigos e as frequências em que aparecem .	35

# Lista de abreviaturas e siglas

CI	Cidades Inteligentes/Smart City/Smart Cities
IoT	Internet das Coisas/Internet of Things
TIC	Tecnologia da Informação e Comunicação
QP	Questão de Pesquisa
ACM	Association for Computing Machinery
IEEE	Institute of Electrical and Electronics Engineers
GC	Grupo de Controle
TSAD	Time Sliced Anomaly Detection
RED	Realtime Emission Detection
DR	Device Ranking
MD	Malfunction Detection
GPS	Global Positioning System
RC	Robust Covariance
IF	Isolation Forrest
CUSUM	Cumulative Sum
HyCARCE	Hyperellipsoidal Clustering Algorithm for Resource-Constrained Environments
ENOF	Ellipsoidal Neighborhood Outlier Factor
HOVSD	Higher-order Singular Value Decomposition
SW4EU	Smart Water For Europe
AURA	Advanced Uncertain Reasoning Architecture
DMA	District Metered Area
MPCA	Multiway Principal Component Analysis
CMMs	Correlation Matrix Memories

RFID      Radio-Frequency Identification

WSN      Wireless Sensor Networks

# Sumário

<b>1</b>	<b>Introdução</b>	<b>13</b>
1.1	Objetivos	14
1.2	Contribuições do Trabalho	14
1.3	Organização do Trabalho	14
<b>2</b>	<b>Referencial Teórico</b>	<b>15</b>
2.1	Cidades Inteligentes	15
2.2	Internet das Coisas	17
2.3	Detecção de <i>Outliers</i>	19
<b>3</b>	<b>Revisão Sistemática</b>	<b>22</b>
3.1	Planejamento	22
3.1.1	Questões de Pesquisa	22
3.1.2	Bases de Dados	22
3.1.3	Palavras-chave	23
3.1.4	<i>String</i> de Busca	23
3.1.5	CrITÉrios de Seleção	25
3.2	Condução	25
3.2.1	Execução das buscas	26
3.2.2	Aplicação dos Filtros	26
3.2.3	Extração dos Dados	28
<b>4</b>	<b>Resultados</b>	<b>29</b>
4.1	Visão Geral dos Estudos	29
4.2	Respostas das Questões de Pesquisa	33
<b>5</b>	<b>Conclusão</b>	<b>36</b>
	<b>Referências</b>	<b>37</b>

# 1 Introdução

Atualmente, a maior parte das pessoas ao redor do mundo encontra-se nas zonas urbanas. Segundo a ONU, mais da metade da população mundial vive nas cidades, cerca de 54%, e até 2050 estima-se que esse número só aumente, girando em torno de 66% (GAUCHAZH, 2014). Esse aglomerado de pessoas causa diversos problemas, o que diminui a qualidade e a expectativa de vida do ser humano que vive nos centros urbanos. Além dos prejuízos à vida humana, a urbanização desenfreada também gera danos ao meio ambiente (clima, fauna, flora etc.).

Um dos principais problemas enfrentados nas grandes cidades é a poluição. Em 2015, no Brasil, a poluição foi responsável por 7,49% das mortes. Ao redor do mundo, no mesmo período, os números totalizaram cerca de 9 milhões de óbitos. No entanto, a maior parte das mortes ocorreram em países de renda baixa e média, onde a poluição do ar representou dois terços desses óbitos (SILVER, 2017). Além da poluição, pode-se destacar também outros problemas, como: congestionamentos, problemas no abastecimento de água potável, chuva ácida, acúmulo de lixo e ilhas de calor.

Diante desse cenário, o conceito de Cidades Inteligentes (CI), juntamente com de Internet das Coisas, do inglês *Internet of Things* (IoT), vêm surgindo como uma forma de propor melhor qualidade de vida nos centros urbanos. O objetivo das CI é auxiliar a administração da cidade na otimização do uso dos recursos públicos e no aumento da qualidade dos serviços oferecidos aos cidadãos (ZANELLA et al., 2014). A IoT provê a infraestrutura necessária para que o objetivo das CI seja alcançado. Por meio de redes de sensores, geralmente de baixo custo, os ambientes urbanos são monitorados e dados são coletados para posterior, ou até mesmo em tempo real, análise e tomadas de decisão.

Contudo, os dados coletados podem conter *outliers* (anomalias), que são observações que se diferenciam muito das outras (HAWKINS, 1980). Mas por que ocorrem *outliers*? “As principais causas estão relacionadas a erros de medição, erros de execução e variabilidade inerente aos elementos da população” (BELFIORE, 2015). Os *outliers*, se não identificados e tratados corretamente, impactam diretamente na análise dos dados e, por consequência, na tomada de decisão. Assim, a detecção de *outliers* vem sendo um tema amplamente discutido na literatura. Com isso, diversos métodos para detectar *outliers* tem sido propostos e empregados nas mais variadas áreas de pesquisa.

Considerando o que foi exposto, o presente trabalho apresenta uma revisão sistemática da literatura no contexto de detecção de *outliers* para o monitoramento ambiental de CI. O principal objetivo dessa revisão é realizar o levantamento e apresentação dos métodos utilizados na literatura para detectar *outliers* em dados ambientais urbanos. Neste trabalho pretende-se também fornecer uma visão geral sobre cada um dos artigos selecionados pela revisão sistemática, destacando informações significativas dos mesmos, como: obje-

tivo do artigo, forma de aquisição, origem e contexto ambiental dos dados, estudo de caso realizado, entre outras.

## 1.1 Objetivos

O objetivo geral deste trabalho é responder a seguinte Questão de Pesquisa: Quais métodos são utilizados na literatura para detectar *outliers* em dados ambientais no contexto de Cidades Inteligentes?

Este trabalho tem como objetivos específicos:

1. Descrever detalhadamente o protocolo de revisão executado;
2. Auxiliar pesquisadores, servindo como ponto de partida para pesquisas sobre detecção de *outliers*; e
3. Fornecer uma visão geral sobre os estudos selecionados pela revisão sistemática.

## 1.2 Contribuições do Trabalho

Acredita-se que as principais contribuições proporcionadas por este trabalho são as seguintes:

1. Levantamento e apresentação dos métodos utilizados na literatura para detectar *outliers* em dados ambientais urbanos, que servirá como embasamento teórico para outras pesquisas na área;
2. Uma visão geral sobre o contexto de detecção de *outliers* para o monitoramento ambiental de CI, destacando possíveis desafios e lacunas encontradas na área; e
3. Apresentação de informações relevantes sobre cada um dos artigos selecionados pela revisão sistemática.

## 1.3 Organização do Trabalho

Este trabalho está organizado da seguinte forma. O Capítulo 2 apresenta o referencial teórico. O Capítulo 3 detalha todas as etapas da revisão sistemática da literatura. O Capítulo 4 fornece uma visão geral dos artigos selecionados pela revisão sistemática, além de apresentar as respostas das Questões de Pesquisa elaboradas. E, por fim, o Capítulo 5 apresenta as conclusões alcançadas por este trabalho, bem como, os trabalhos futuros pretendidos.

## 2 Referencial Teórico

Para um melhor entendimento deste trabalho, alguns tópicos importantes são apresentados, conceituados e caracterizados neste capítulo. A Seção 2.1 trata do conceito de Cidades Inteligentes. A Seção 2.2 apresenta o paradigma de Internet das Coisas. A Seção 2.3 aborda a Detecção de *Outliers*.

### 2.1 Cidades Inteligentes

Desde os anos 90, o conceito de CI tem sido estudado (DERITTI; FREIRE, 2018). Entretanto, foi nos últimos anos que este conceito ganhou notoriedade, firmando o elo entre a tecnologia e as cidades. Essa afirmação pode ser corroborada pelo crescimento na quantidade de publicações sobre tal contexto (AUNE; GONÇALVES, 2017). Toda essa relevância das CI pode ser observada também na prática. Segundo Garcia-Font, Garrigues e Rifà-Pous (2018), grandes cidades ao redor do mundo tem instalado componentes tecnológicos em suas ruas, visando coletar informações urbanas para possibilitar que as autoridades ofereçam melhores serviços aos cidadãos. Há também CI construídas do zero, como as cidades de Masdar, nos Emirados Árabes Unidos e Songdo, na Coreia do Sul (LOPES, 2017). A Figura 2.1 apresenta uma consulta à ferramenta *Google Trends*<sup>1</sup>, onde pode-se verificar o crescimento da procura pelo termo “*Smart Cities*” ao redor do mundo.

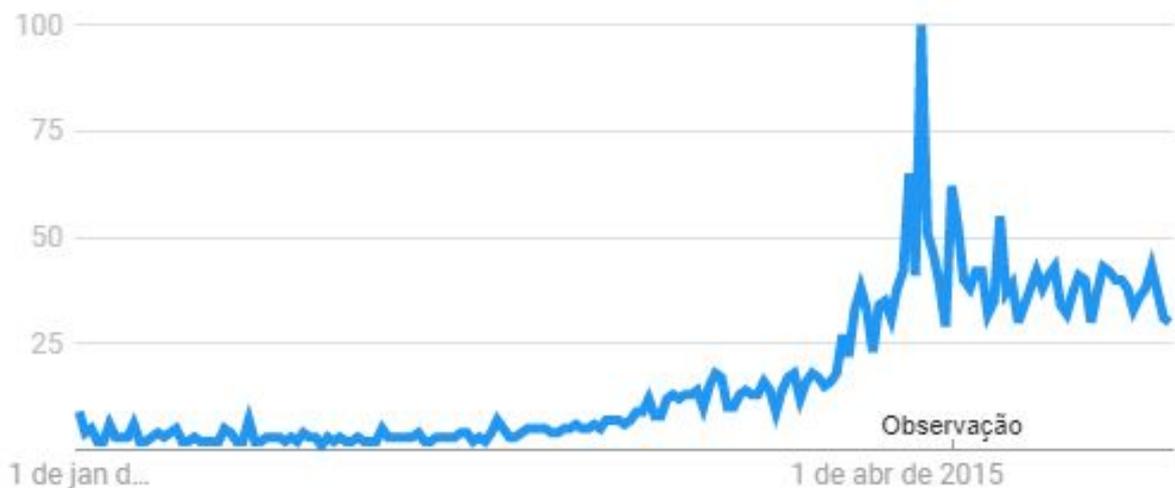


Figura 1: Busca pelo termo “*Smart Cities*” (período: 01/01/2004 a 23/01/2019)

Fonte: Google Trends

<sup>1</sup> <https://trends.google.com/trends/?geo=US>

Segundo [Cunha et al. \(2016\)](#), o fenômeno das CI está atrelado a duas megatendências que motivam a transformação da sociedade moderna: um movimento de urbanização, visto que mais da metade da população mundial vive nos centros urbanos; e a revolução digital, pois o desenvolvimento das Tecnologias da Informação e Comunicação (TICs) permitiu a hiperconectividade tanto entre as pessoas como entre as máquinas, dando chance a chamada sociedade colaborativa. De acordo com [Guedes et al. \(2014\)](#), a urbanização desenfreada originou grandes desafios à gestão das cidades e seus recursos, fazendo com que o conceito de CI surgisse. Para [Andrade e Galvão \(2016\)](#), o conceito de CI vem se provando essencial para aumentar a eficiência das grandes cidades, propiciando melhor qualidade de vida e gestão de recursos naturais, como a água.

Apesar de toda a relevância do tema, ainda não há consenso sobre a definição de CI ou dos componentes mínimos que devem estar presentes em uma ([CUNHA et al., 2016](#)). [Prado e Santos \(2014\)](#), afirmam em seu trabalho, que além da divergência de opiniões, a base teórica dos autores influencia nas diferentes visões sobre CI. De acordo com [Russo, Rindone e Panuccio \(2014\)](#), a ideia de CI é utilizada em tópicos multidisciplinares que adotam sua própria linguagem, por isso, não há definição que inclua todos os aspectos de CI. A Tabela 1 traz algumas definições de CI bem difundidas na literatura.

Tabela 1: Definições de CI

Referência	Definição
<a href="#">Odendaal (2003)</a>	Uma CI é aquela que capitaliza sobre as oportunidades apresentadas pela TIC na promoção de sua prosperidade e influência.
<a href="#">Giffinger et al. (2007)</a>	Uma CI é uma cidade que está em franco desenvolvimento nestas seis características: economia inteligente; pessoas inteligentes; governança inteligente; mobilidade inteligente; ambiente inteligente e; vida inteligente, construídos com uma combinação de doações e autogerenciamento, com cidadãos independentes e conscientes.
<a href="#">Caragliu, Bo e Nijkamp (2009)</a>	Uma cidade pode ser considerada uma CI quando os investimentos em capital humano e social e a tradicional e moderna infraestrutura de TIC serve como impulsionadora de um crescimento econômico sustentável e de uma elevada qualidade de vida, com uma gestão prudente dos recursos naturais através da governança participativa.
<a href="#">Komninos, Schaffers e Pallot (2011)</a>	O conceito de CI está ligado às noções de competitividade global, sustentabilidade, capacitação e qualidade de vida, apoiados por redes de banda larga e tecnologias modernas.
<a href="#">Nam e Pardo (2011)</a>	Uma CI integra tecnologias, sistemas, infraestruturas, serviços e recursos em uma rede orgânica que é suficientemente complexa para desenvolver propriedades emergentes inesperadas.
<a href="#">Batty et al. (2012)</a>	Uma CI é uma cidade em que as TICs são mescladas com infraestruturas tradicionais, coordenando e integrando o uso de novas tecnologias digitais. CI também são instrumentos para melhorar a competitividade de tal forma que a comunidade e a qualidade de vida são reforçadas.

Fonte: [Rizzon et al. \(2017\)](#) adaptado de [Papa et al. \(2015\)](#)

De acordo com a Tabela 1, nota-se que é de fundamental importância, para o conceito de CI, a utilização das TICs. As CI teriam muita dificuldade em serem criadas sem as TICs. Por outro lado, as TICs são apenas um ferramental para facilitar o desenvolvimento das CI (ANDERLE; JUNIOR, 2013).

## 2.2 Internet das Coisas

Devido ao advento das casas, cidades e dispositivos inteligentes, a Internet das Coisas emerge como uma área de grande potencial e crescimento (KHAN; SALAH, 2018). De acordo com Qian et al. (2012), a IoT baseia-se em duas ideias principais. A primeira delas trata a internet como o núcleo e a base da IoT. Já a outra entende que as “coisas” tornam-se usuários terminais, que podem interagir uns com os outros. Peng e Shen (2012) definem a IoT como uma rede que permite que qualquer objeto conectado à Internet possa se comunicar e trocar informações.

Segundo Al-Fuqaha et al. (2015), seis elementos compõem a IoT. A Figura 2 exibe os seis elementos. E, em seguida, uma breve descrição de cada elemento é apresentada.



Figura 2: Elementos da IoT

Fonte: Al-Fuqaha et al. (2015)

- Identificação: a identificação é um dos elementos mais importantes. Os métodos de identificação fornecem uma identidade clara para cada objeto dentro da rede. Os métodos de endereçamento auxiliam na identificação, atribuindo endereços de rede aos objetos.
- Sensoriamento: no sensoriamento, os dados coletados pelos objetos da rede são enviados para um *Data Warehouse*, banco de dados ou nuvem.
- Comunicação: as tecnologias de comunicação conectam objetos heterogêneos para fornecer serviços inteligentes específicos. Uma tecnologia de comunicação bem antiga e conhecida é a Identificação por Radiofrequência, do inglês *Radio-Frequency Identification* (RFID).
- Computação: as unidades de processamento, como microprocessadores, e aplicações de software representam o “cérebro” e a capacidade computacional da IoT.

- Serviços: geralmente, a IoT provê quatro tipos de serviços: serviços relacionados à identidade, serviços de agregação de informações, serviços de colaboração e serviços ubíquos.
- Semântica: a semântica refere-se à capacidade de extrair conhecimento de maneira inteligente por diferentes máquinas para fornecer os serviços necessários.

A Figura 3 apresenta a arquitetura de três camadas para a IoT, adotada por vários trabalhos na literatura (BARBOSA, 2017).

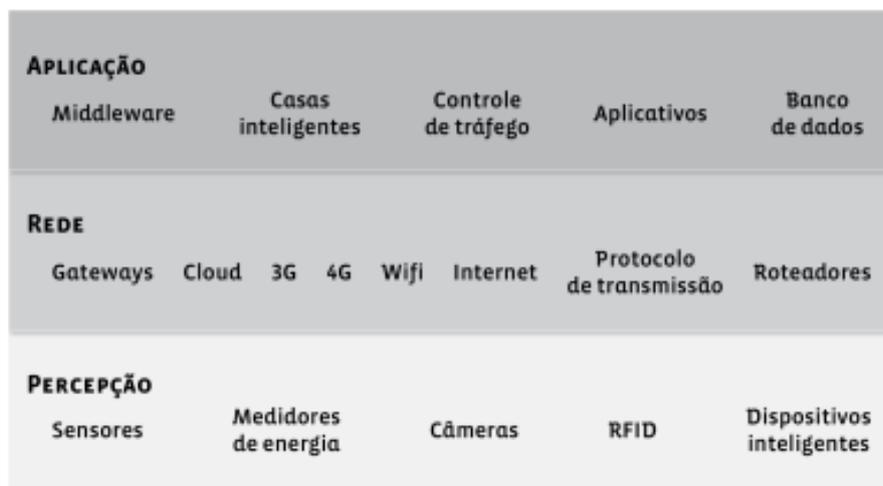


Figura 3: Arquitetura de três camadas para a IoT

Fonte: BARBOSA (2017)

- Camada de percepção: essa camada é a porta de entrada dos dados. Por meio dos dispositivos encontrados nessa camada (sensores, câmeras etc), os dados do ambiente podem ser coletados.
- Camada de rede: camada responsável por transmitir os dados obtidos pela camada de percepção para a camada de aplicação.
- Camada de aplicação: camada responsável por armazenar, processar e analisar os dados obtidos pela camada de percepção. Essa camada deve prover serviços aos clientes.

A IoT tem um enorme potencial para o desenvolvimento de novas aplicações inteligentes em vários campos de atuação (BORGIA, 2014). Diante da gama de possibilidades que a IoT oferece, é natural associá-la ao contexto de CI (TOMAS, 2014). Para Borgia (2014), a IoT é essencial para transformar uma cidade tradicional em uma CI. Para que ocorra essa transformação, a IoT deve ser aliada ao uso das TICs. No contexto das CI,

são vários os cenários de aplicação das TICs e IoT: mobilidade urbana, gerenciamento de energia, monitoramento ambiental, segurança, entre outros.

Tomas (2014), em seu trabalho, apresentou um caso real da eficiência da aplicação das TICs baseadas em IoT. Foi realizado um experimento na cidade de Londres. A partir do experimento, houve a redução de 30% das emissões de carbono, apenas com a mudança de comportamento dos cidadãos. Tal mudança de comportamento foi possível pelas informações do nível de carbono fornecidas, em tempo real, por uma Rede de Sensores Sem Fio, do inglês *Wireless Sensor Network* (WSN).

## 2.3 Detecção de *Outliers*

O cuidado com os *outliers* é antigo e remonta às primeiras tentativas de se analisar um conjunto de dados. Por volta de 1850, foram relatados os esforços iniciais para o desenvolvimento de métodos estatísticos que lidassem com *outliers*. (BECKMAN; COOK, 1983). Apesar disso, os *outliers* ainda não possuem uma definição rigorosa, sendo normalmente definidos como observações que desviam significativamente da maior parte dos dados (SU; TSAI, 2011). De acordo com Silva et al. (2004), um valor é dito *outlier* sempre em relação a um padrão, dessa forma, os *outliers* sempre ocorrem em pequenas partes do conjunto de dados, do contrário, seria impossível determinar um padrão.

Segundo Barros (2014), antes de tomar qualquer decisão em relação aos *outliers*, é interessante que se conheça os fatores que provocam o surgimento dessas observações. Em vários casos, as razões da existência dos *outliers* indicam a forma de tratá-los. Os *outliers* são gerados, principalmente, por (BARNETT; LEWIS, 1984 apud SILVA et al., 2004):

- Erros de medição: esse tipo de erro ocorre na coleta dos dados. E pode ser gerado por erros humanos ou de máquinas, como sensores .
- Erros de execução: ocorrem quando os dados são obtidos mediante amostragem de mais de uma população.
- Variabilidade inerente aos elementos da população: nesse caso, os *outliers* são elementos pertencentes à população.

Para Chandola, Banerjee e Kumar (2009), os *outliers* podem ser classificados em três categorias, baseando-se em sua composição e relação com o restante dos dados. Essa classificação, muito difundida na literatura, é apresentada na Tabela 2. A Figura 4 traz exemplos de cada uma das categorias, onde os *outliers* estão destacados em vermelho.

Tabela 2: Categorias e descrições de *outliers*

Categoria	Descrição
<i>Outlier</i> pontual	É a categoria mais simples. Um elemento é considerado <i>outlier</i> em relação ao restante dos dados. Exemplo: Figura 4 (a).
<i>Outlier</i> coletivo	Um grupo de elementos relacionados é considerado <i>outlier</i> em relação ao restante dos dados. Apenas um elemento não pode ser considerado <i>outlier</i> , mas sim a ocorrência em conjunto de vários elementos. Exemplo: Figura 4 (b).
<i>Outlier</i> contextual	Um elemento é dito <i>outlier</i> em um determinado contexto. Cada elemento é definido usando atributos contextuais e comportamentais. Os atributos contextuais definem a posição e vizinhança de um elemento específico. E os atributos comportamentais definem as características não contextuais do elemento. Exemplo: Figura 4 (c).

Fonte: Adaptado de [Chandola, Banerjee e Kumar \(2009\)](#)

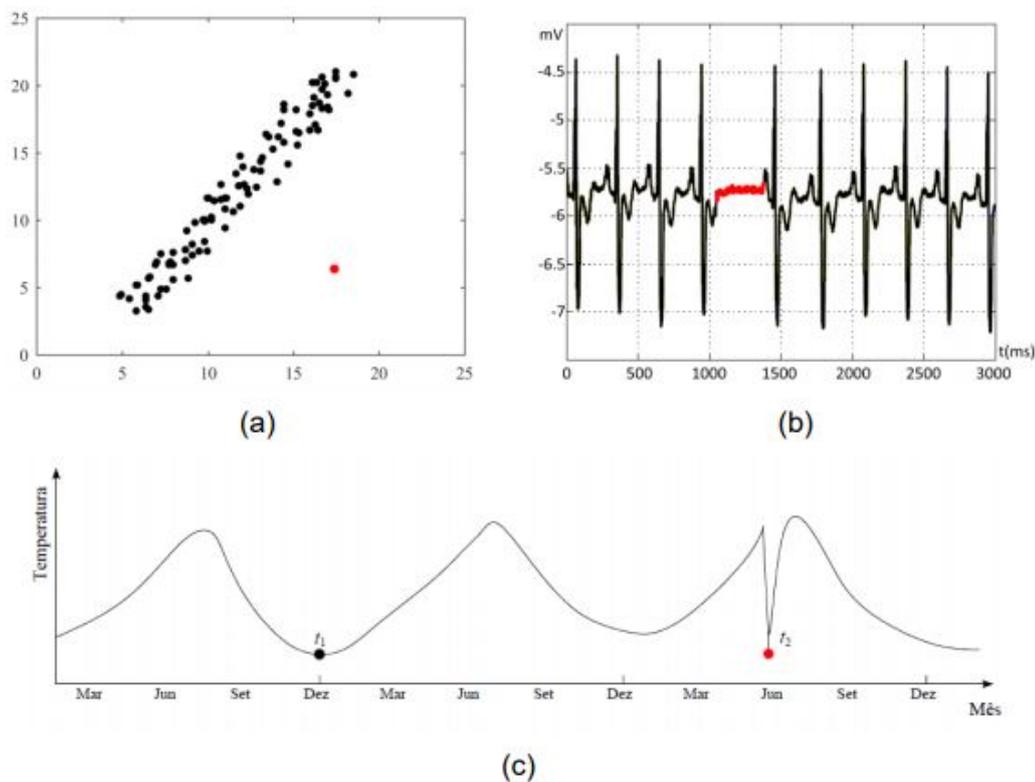


Figura 4: Exemplos das categorias de *outliers*. (a) *Outlier* pontual; (b) *outlier* coletivo; (c) *outlier* contextual

Fonte: [Rodrigues \(2018\)](#) adaptado de [Aggarwal \(2013\)](#)

Os *outliers* podem conter informações significativas, e às vezes críticas, em inúmeros domínios de aplicação ([CHANDOLA; BANERJEE; KUMAR, 2009](#)). Por essa razão, entender, detectar e tratar *outliers* de forma correta é importante e vem tendo cada vez mais atenção no meio científico. De acordo com [Habeeb et al. \(2018\)](#), a detecção de *outliers* é um tema relevante que tem sido estudado em várias áreas de pesquisa e domínios de aplicação. [Rodrigues \(2018\)](#) afirma que a detecção de *outliers* busca identificar as amostras

que destoam acentuadamente dos padrões em um conjunto de dados, sendo assim, uma tarefa crucial para descoberta de conhecimento em mineração de dados.

Como dito anteriormente, a detecção de *outliers* tem sido um problema amplamente explorado, por consequência, inúmeras técnicas, fundamentadas nas mais variadas abordagens, foram propostas na literatura. Pode-se citar algumas abordagens, como: abordagem estatística, baseada em profundidade, baseada no desvio, baseada em distância, baseada em densidade, abordagem de alta dimensionalidade, entre outras. Cada abordagem apresenta seus próprios pressupostos teóricos para solucionar impasses encontrados nos vários cenários possíveis. Dessa maneira, a detecção de *outliers* se faz dependente do contexto do problema e da natureza do conjunto de dados (RODRIGUES, 2018).

## 3 Revisão Sistemática

Neste capítulo, a revisão sistemática da literatura é apresentada detalhadamente. A revisão sistemática pode ser definida “como uma síntese de estudos primários que contém objetivos, materiais e métodos claramente explicitados e que foi conduzida de acordo com uma metodologia clara e reprodutível” (GREENHALGH, 1997). De acordo com Ciliska, Cullum e Marks (2001), utilizando uma metodologia de pesquisa rigorosa, a revisão sistemática busca superar vieses em todas as etapas, distinguindo-se de outros métodos de pesquisa. A revisão sistemática pode ser dividida em três etapas: Planejamento, Condução e Relatório da Revisão (MIRANDA, 2011).

### 3.1 Planejamento

Na etapa de planejamento, o protocolo de revisão é estabelecido e por meio dele é possível que outros pesquisadores repitam a pesquisa e extrapolem os resultados para os anos que se seguem (MORAES T. M.; SOUZA, 2011). No protocolo de revisão são especificadas as Questões de Pesquisa (QPs), as bases de dados que são utilizadas, a *string* de busca e os critérios de seleção de estudos (inclusão e exclusão).

#### 3.1.1 Questões de Pesquisa

Definir uma QP é o passo mais importante na elaboração da revisão sistemática, pois proporciona a direção para a execução das outras etapas do processo (DICKSON, 1999 apud GALVÃO; SAWADA; TREVIZAN, 2004). A principal QP formulada para este trabalho é a seguinte: **QP 1 - Quais métodos são utilizados na literatura para detectar *outliers* em dados ambientais no contexto de CI?** Também foram elaboradas QPs secundárias:

- **QP 2 - Os resultados da detecção de *outliers* são validados?**
- **QP 3 - Quais os contextos do monitoramento ambiental são abordados pelos estudos?**

#### 3.1.2 Bases de Dados

As bases de dados utilizadas nesta revisão sistemática foram as seguintes: *Association for Computing Machinery (ACM) Digital Library*<sup>1</sup>, *Institute of Electrical and Electronics*

<sup>1</sup> <https://dl.acm.org>

*Engineers* (IEEE) *Digital Library*<sup>2</sup>, *Science Direct*<sup>3</sup>, *Web of Science*<sup>4</sup>, *Scopus*<sup>5</sup> e *Engineering Village*<sup>6</sup>. Inicialmente, considerou-se também a base de dados *Springer Link*<sup>7</sup>, entretanto, ao executar as buscas, constatou-se que os artigos retornados não estavam disponíveis gratuitamente no site e nem em outra fonte na *Web*.

### 3.1.3 Palavras-chave

As palavras-chave devem estar diretamente relacionadas com as QPs para que a *string* de busca possa retornar os melhores resultados. Por isso, baseando-se nas QPs, as palavras-chave foram escolhidas e, para verificar a relevância de tais palavras, foi executada uma revisão informal da literatura, na qual buscou-se estudos relacionados ao tema deste trabalho. Através dessa revisão informal, pôde-se verificar que os termos escolhidos são comumente utilizados no título, resumo ou nas palavras-chave dos estudos encontrados. A Tabela 3 apresenta as palavras-chave definidas para este trabalho.

Tabela 3: Palavras-chave

<b>Palavras-chave</b>
<i>Outlier Detection</i> (Detecção de Outliers)
<i>Anomaly Detection</i> (Detecção de Anomalias)
<i>Urban Data</i> (Dados Urbanos)
<i>Smart Cities</i> (Cidades Inteligentes)

É importante ressaltar que termos relacionados ao contexto ambiental não foram escolhidos como palavras-chave, pois a literatura utiliza uma grande variedade de termos para se referir ao tema e, na maioria das vezes, esses termos são bem específicos. Essa variedade de termos dificulta a definição de palavras-chave que englobem todo o contexto ambiental. Esse fato foi constatado após a realização da revisão informal, citada anteriormente, e também por testes iniciais com a *string* de busca.

### 3.1.4 *String* de Busca

Segundo Munzlinger, Narcizo e Queiroz (2012), a *string* de busca “trata-se de um conjunto composto por palavras-chave, utilizado para realizar pesquisas de publicações em bibliotecas digitais”. Ao formular uma *string* de busca é recomendável agrupar termos relacionados a um mesmo tópico com o operador lógico *OR*, formando grupos de “sinônimos”, e depois associar esses grupos através do operador *AND* (FALBO, 2018).

<sup>2</sup> <https://www.ieee.org>

<sup>3</sup> <https://www.sciencedirect.com>

<sup>4</sup> <https://webofknowledge.com>

<sup>5</sup> <https://www.scopus.com>

<sup>6</sup> <https://www.engineeringvillage.com/>

<sup>7</sup> <https://link.springer.com/>

Tomando como base essa boa prática, a seguinte *string* de busca foi formulada: ((“*outlier detection*” OR “*anomaly detection*”) AND (“*urban data*” OR “*smart cities*”)).

Alguns testes foram realizados para se chegar na *string* de busca final e garantir sua qualidade. Por meio da revisão informal, mencionada na Subseção 3.1.3, um Grupo de Controle (GC), formado por cinco estudos primários, foi estabelecido. Segundo Falbo (2018), utilizar um GC para avaliar a *string* de busca é um meio eficiente e muito utilizado. Após definir o GC, o procedimento conhecido como teste piloto foi executado. O teste piloto consiste na submissão da *string* de busca nas bases de dados que contém os estudos do GC, a fim de verificar se os mesmos são retornados. Considerou-se para o teste piloto as bases de dados IEEE e *Web of Science*, pois foram as primeiras bases escolhidas para esta revisão sistemática.

A Tabela 4 exibe cada versão testada da *string* de busca e seu respectivo resultado do teste piloto. Nas versões I, II e III da *string* de busca foram utilizados termos relacionados ao monitoramento ambiental, visando ratificar o entendimento sobre a dificuldade de definir termos genéricos para esse tema. Diante disso, decidiu-se generalizar a *string* de busca na versão IV, retirando os termos referentes ao contexto ambiental. Dessa forma, qualquer trabalho que abordasse a detecção de *outliers* em dados urbanos seria retornado como resultado.

Tabela 4: Versões da *string* de busca e seus respectivos resultados do teste piloto

Versão	String de Busca	Quantidade de Estudos Retornados (GC)
I	((“outlier detection” OR “anomaly detection”) AND (“urban data” OR “smart cities”) AND (“environmental data”))	1
II	((“outlier detection” OR “anomaly detection”) AND (“urban data” OR “smart cities”) AND (“environmental data” OR “environment”))	2
III	((“outlier detection” OR “anomaly detection”) AND (“urban data” OR “smart cities”) AND (“environmental data” OR “environment” OR “environmental monitoring”))	2
IV	((“outlier detection” OR “anomaly detection”) AND (“urban data” OR “smart cities”))	5

A *string* de busca deve ser adaptada corretamente para cada base de dados. Isso se deve ao fato de que, geralmente, as bases de dados trabalham de modo diferente, possuindo sintaxe, limitações e funcionalidades próprias. Na Tabela 5, estão as *strings* de busca utilizadas em cada base de dados.

Tabela 5: *Strings* de busca utilizadas

Base(s) de Dados	String de Busca
ACM	((acmdlTitle:(“outlier detection”) OR recordAbstract:(“outlier detection”) OR keywords.author.keyword:(“outlier detection”)) OR ((acmdlTitle: (“anomaly detection”) OR recordAbstract:(“anomaly detection”) OR keywords.author.keyword:(“anomaly detection”))) AND (((acmdlTitle: (“smart cities”) OR recordAbstract:(“smart cities”) OR keywords.author.keyword:(“smart cities”)) OR ((acmdlTitle: (“urban data”) OR record Abstract:(“urban data”) OR keywords.author.keyword:(“urban data”))))
IEEE, Science Direct, Web of Science, Scopus e <i>Engineering Village</i>	((“outlier detection” OR “anomaly detection”) AND (“urban data” OR “smart cities”))

Para esta revisão sistemática, a *string* de busca foi modificada apenas para a base de dados ACM. A sintaxe empregada encontra-se no *site* da Universidade Federal do Espírito Santo (UFES)<sup>8</sup>.

### 3.1.5 Critérios de Seleção

Os critérios de seleção são divididos em duas categorias: inclusão e exclusão. A Tabela 6 apresenta os critérios de seleção estabelecidos nesta revisão sistemática.

Tabela 6: Critérios de seleção de estudos

Inclusão	Exclusão
1- Utiliza alguma técnica para detectar <i>outliers</i>	1- Artigo não redigido em inglês
2- Área de Ciência da Computação	2- Título e resumo fora do escopo desta revisão
3- Trabalha com dados ambientais reais urbanos	3- Todo o texto fora do escopo desta revisão
4- Artigo redigido em inglês	4- Capítulo de livro ou livro completo
5- Artigo completo disponível na <i>Web</i>	5- Teses, dissertações e relatórios técnicos

Os critérios de seleção são elaborados com base nas QPs e devem ser rigorosamente aplicados (FALBO, 2018). Por meio desses critérios, somente os trabalhos que estejam estritamente relacionados com o contexto da revisão sistemática são considerados relevantes.

## 3.2 Condução

Na etapa de condução ocorre a execução do protocolo de revisão. Os estudos primários são identificados e, posteriormente, selecionados mediante a aplicação dos critérios de inclusão e exclusão. Após realização da seleção dos estudos, os dados necessários para responder às QPs são extraídos e sintetizados.

<sup>8</sup> [inf.ufes.br/~vitorsouza/wp-content/uploads/teaching-br-metpesq-20182-especificacao-atividade2-exemplo-buscas.txt](http://inf.ufes.br/~vitorsouza/wp-content/uploads/teaching-br-metpesq-20182-especificacao-atividade2-exemplo-buscas.txt)

### 3.2.1 Execução das buscas

A submissão da *string* de busca nas fontes de pesquisa, com intuito de identificar os estudos primários, ocorreu no período de 10/01/2019 a 17/01/2019, exceto na *Engineering Village*, na qual realizou-se posteriormente, no dia 09/03/2019. Após a submissão da *string* de busca, os arquivos encontrados que estavam disponíveis na *Web* foram baixados e devidamente organizados em pastas<sup>9</sup>. Todo o processo de busca e *download* dos trabalhos foi realizado de forma manual, sem apoio de qualquer *software*.

Ao finalizar as buscas e os *downloads*, foram obtidos 293 artigos no total, divididos da forma como é mostrado na Figura 5: 6 na ACM (2%), 27 no IEEE (9%), 167 na *Science Direct* (57%), 40 na *Scopus* (14%), 23 na *Web of Science* (8%) e 30 na *Engineering Village* (10%).

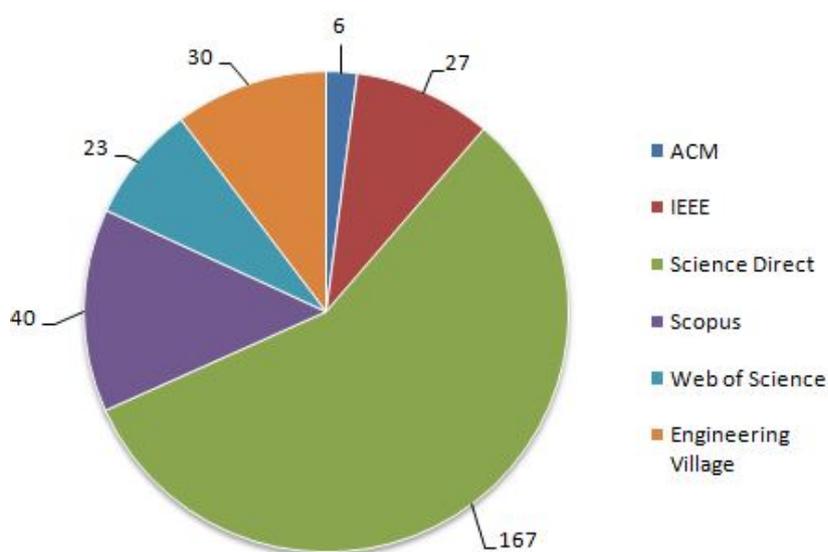


Figura 5: Quantidade de estudos encontrados por base de dados

Como pode-se perceber na Figura 5, a *Science Direct* destacou-se como a base de dados que mais retornou estudos nesta revisão sistemática. Enquanto a ACM apresentou a menor quantidade de estudos retornados.

### 3.2.2 Aplicação dos Filtros

Neste passo da etapa de Condução ocorreu a seleção dos estudos primários. Para isso, foram definidos três filtros. O primeiro tratou dos artigos que estavam duplicados entre as bases de dados. Foram identificados 71 estudos duplicados (24%), assim, 222 artigos passaram para o segundo filtro (76%).

<sup>9</sup> Os arquivos baixados podem ser encontrados em: [https://drive.google.com/open?id=16hx\\_a-u88lCRjg7R9zk2swTBMyh7\\_Pyw](https://drive.google.com/open?id=16hx_a-u88lCRjg7R9zk2swTBMyh7_Pyw)

O segundo filtro consistiu na aplicação dos critérios de seleção mediante a leitura do título e resumo dos estudos. Em alguns casos surgiram dúvidas em relação à decisão de aceitar ou não determinado trabalho, nesses casos, a introdução e conclusão dos trabalhos também foram lidas para apoiar a decisão. Foram aceitos apenas 15 (7%) do total de 222 artigos.

No terceiro e último filtro, os artigos foram lidos de forma completa e detalhada para que os critérios de seleção fossem aplicados. Após a aplicação do terceiro filtro, selecionou-se oito (53%) dos 15 artigos restantes. Dessa forma, apenas 3% dos 293 estudos encontrados foram selecionados nesta revisão.

Na Tabela 7 é exibido o passo a passo do processo de aplicação dos filtros. A Tabela 8 apresenta os artigos selecionados nesta revisão, destacando os autores, o título e a base de dados onde cada artigo foi encontrado.

Tabela 7: Aplicação dos filtros sobre os artigos

Base de Dados	Artigos Encontrados	1º Filtro	2º Filtro	3º Filtro
ACM	6	6	0	0
IEEE	27	27	3	2
<i>Science Direct</i>	167	167	9	3
<i>Scopus</i>	40	13	1	1
<i>Web of Science</i>	23	7	2	2
<i>Engineering Village</i>	30	2	0	0
<b>Crítérios</b>	Artigos disponíveis e redigidos em inglês	Artigos Duplicados	Crítérios de seleção aplicados ao título e resumo	Crítérios de seleção aplicados ao texto completo

Tabela 8: Artigos selecionados

Autores/Ano	Título do Trabalho	Base de Dados
Mounce et al. (2015)	<i>Cloud based machine learning approaches for leakage and management in smart water networks</i>	<i>Science Direct</i>
Trilles et al. (2016)	<i>A domain-independent methodology to analyze IoT data streams in real-time. A proof of concept implementation for anomaly detection from environmental data</i>	<i>Web of Science</i>
Cicirelli et al. (2017)	<i>An edge-based platform for dynamic Smart City applications</i>	<i>Science Direct</i>
Santos et al. (2017)	<i>Fog Computing: Enabling the Management and Orchestration of Smart City Applications in 5G Networks</i>	<i>Web of Science</i>
Chen et al. (2018)	<i>ADF: An Anomaly Detection Framework for Large-Scale PM2.5 Sensing Systems</i>	IEEE
Rathore et al. (2018)	<i>Real-Time Urban Microclimate Analysis Using Internet of Things</i>	<i>Scopus</i>
Santos et al. (2018)	<i>Anomaly detection for Smart City applications over 5G Low Power Wide Area Networks</i>	IEEE
Souza, Aquino e Gomes (2019)	<i>A method to detect data outliers from smart urban spaces via tensor analysis</i>	<i>Science Direct</i>

Após a seleção dos artigos, percebeu-se que a *Web of Science* obteve o melhor aproveitamento entre todas as bases de dados, levando em conta a quantidade de artigos selecionados e a quantidade inicial retornada pela base. Dos 23 estudos primários retornados inicialmente, dois (cerca de 9%) foram selecionados por esta revisão sistemática. O pior aproveitamento ficou por conta da *Science Direct*, apenas três (cerca de 2%) dos 167 estudos retornados em tal base foram selecionados.

### 3.2.3 Extração dos Dados

Os artigos selecionados foram lidos novamente e dados relevantes foram extraídos dos mesmos. Esses dados serviram para responder às QPs elaboradas e fornecer informações gerais sobre os artigos. Após a leitura dos artigos, foram extraídos os seguintes dados:

- Método/Técnica para detecção de *outliers*;
- Objetivo do estudo;
- Cenário de avaliação;
- Forma de validação dos resultados da detecção de *outliers*;
- Forma de aquisição dos dados;
- Origem dos dados; e
- Contexto ambiental dos dados.

## 4 Resultados

Inicialmente, ao submeter a *string* de busca nas bases de dados, encontrou-se 293 artigos redigidos em inglês e disponíveis por completo na *Web*. Após armazenar adequadamente esses artigos, aplicou-se os três filtros definidos (Subseção 3.2.2) e, assim, selecionou-se oito artigos que atenderam às QPs. Os artigos selecionados estão distribuídos entre os anos de 2015 a 2019, o que confirma a relevância do assunto, e pelas bases de dados IEEE (dois artigos), *Science Direct* (três artigos), *Scopus* (um artigo) e *Web of Science* (dois artigos).

Este capítulo está dividido da seguinte forma. A Seção 4.1 fornece uma visão geral dos estudos primários selecionados, destacando os objetivos, o cenário de avaliação/estudo de caso, a forma de aquisição, origem e o contexto dos dados utilizados. A Seção 4.2 discute e responde cada uma das QPs formuladas.

### 4.1 Visão Geral dos Estudos

Mounce et al. (2015) apresentaram um *site* de demonstração do projeto *Smart Water For Europe*<sup>1</sup> (SW4EU). O *site* de demonstração apresentado foi o *Thames Water*<sup>2</sup> do Reino Unido, que foca em gerenciar vazamentos de água, além de explorar a otimização de energia e interação com o cliente. O local de demonstração do *Thames Water* está situado em Reading, na Inglaterra, e conta com cerca de 1.000 medidores de água *Advanced Metering Infrastructure* (AMI) inteligentes. Em seu trabalho, Mounce et al. (2015) propuseram um estudo de caso para analisar dados sobre o fluxo da água e utilizaram um sistema baseado na *Advanced Uncertain Reasoning Architecture* (AURA), chamado de *AURA-Alert*, para detectar anomalias. Os dados utilizados no estudo foram coletados entre janeiro de 2013 a junho de 2014 em cinco *District Metered Areas* (DMAs), que são áreas discretas da rede de distribuição de água que podem ser isoladas para medição do fluxo.

Trilles et al. (2016) introduziram uma metodologia independente de domínio para analisar o fluxo de dados em redes de sensores. O teste da metodologia deu-se por meio da implementação de uma prova de conceito, que teve como objetivo detectar anomalias em dados relacionados à qualidade do ar. Os dados utilizados para o teste foram coletados através de uma rede de sensores implantada pelo governo de Valencia, na Espanha<sup>3</sup>. Essa rede de sensores mediu o nível de vários poluentes atmosféricos e também as condições meteorológicas em áreas urbanas, rurais e industriais. O algoritmo *Cumulative Sum*

<sup>1</sup> <https://sw4eu.com/>

<sup>2</sup> <https://www.thameswater.co.uk/>

<sup>3</sup> <http://www.agroambient.gva.es/va/web/calidad-ambiental/datos-on-line>

(CUSUM) foi o escolhido para detectar anomalias nos dados. De acordo com [Trilles et al. \(2016\)](#), a escolha do CUSUM baseou-se no sucesso alcançado pela técnica em alguns trabalhos relacionados. Para ajustar os parâmetros do CUSUM, foi utilizado um *dataset* coletado pela rede de sensores do governo de Valencia, entre o período de 01 de janeiro a 31 de dezembro de 2013. Após o ajuste de parâmetros, foram definidos dois limites: inferior e superior, utilizados para detectar anomalias.

[Cicirelli et al. \(2017\)](#) apresentaram a plataforma distribuída *iSapiens*, baseada no paradigma *Edge Computing*, para o gerenciamento de uma rede de nós de computação espalhados em áreas urbanas. Um conjunto de diretrizes de *design* também foi proposto pelos autores. A abordagem (plataforma *iSapiens* e o conjunto de diretrizes) foi validada pela implementação de um estudo de caso na cidade de Consenza, na Itália. No estudo de caso, sensores foram instalados em alguns locais de Consenza, visando avaliar a qualidade do clima e os níveis de poluição atmosférica e sonora. Foram desenvolvidos agentes para detectar anomalias em cada sensor relacionado à poluição sonora. Para detectar as anomalias, foram exploradas as medidas estatísticas da média e o desvio padrão global dos sensores, o que caracterizou uma distribuição gama. Agentes de agregação de medidas calcularam a média e o desvio padrão utilizados pelos agentes de detecção de anomalias.

[Santos et al. \(2017\)](#) propuseram uma *framework* de *Fog Computing* para gerenciamento e orquestração de aplicativos de CI em redes 5G. O cenário de avaliação da *framework* foi baseado em um caso de uso de monitoramento da qualidade do ar dentro do escopo da *City of Things* de Antuérpia, na Bélgica. Para coletar os dados, foram implementados sensores em dois carros do serviço postal belga. Além da localização do *Global Positioning System* (GPS), os sensores enviaram dados sobre PM1, PM2.5, PM10, temperatura e umidade. Dois *datasets* foram coletados durante o período de 09 de maio a 29 junho de 2017. Com o objetivo de detectar anomalias nas medições de PM10, temperatura e umidade. [Santos et al. \(2017\)](#) aplicaram os algoritmos de aprendizado de máquina não supervisionados: *Robust Covariance* (RC) e *Isolation Forrest* (IF), implementados em *Python* pela biblioteca *Scikit-Learn*. Os resultados obtidos pelas técnicas RC e IF foram comparados.

[Chen et al. \(2018\)](#) propuseram a *Anomaly Detection Framework* (ADF) para garantir a qualidade dos dados em sistemas de sensoriamento ambiental. A ADF é baseada em um módulo central chamado *Time-Sliced Anomaly Detection* (TSAD), responsável por detectar as anomalias. Além do TSAD, a ADF ainda conta com outros três módulos, são eles: (II) *Realtime Emission Detection* (RED); (III) *Device Ranking* (DR); e (IV) *Malfunction Detection* (MD). Para os autores, a ADF pode identificar efetivamente *outliers* em dados brutos e também inferir eventos anômalos. No trabalho de [Chen et al. \(2018\)](#) foram utilizados dois *datasets*, o primeiro, coletado em uma escola de Taiwan, serviu para ajuste de parâmetros do TSAD e o segundo, coletado em cinco cidades de Taiwan, foi usado para a realização dos experimentos de detecção de anomalias. Ambos *datasets* foram obtidos por

meio do projeto *AirBox*<sup>4</sup>, que tem como objetivo monitorar o poluente atmosférico conhecido como material particulado ou *particulate matter* (PM), com partículas que chegam até 2.5 micrômetros de tamanho (PM2.5).

Rathore et al. (2018) apresentaram uma *framework* de geovisualização multivariada e interativa para monitoramento em tempo real e análise de padrões complexos em dados de microclima urbano. A *framework* constituiu-se de dois métodos analíticos: (I) Estimção Espaço-Temporal e (II) Detecção de Padrões, que buscou identificar padrões incomuns nos dados. A Detecção de Padrões utilizou os algoritmos *Hyperellipsoidal Clustering Algorithm for Resource-Constrained Environments* (HyCARCE) e *Ellipsoidal Neighborhood Outlier Factor* (ENOF) como método de detecção de anomalias. O método de detecção de anomalias foi aplicado sobre dados de temperatura e umidade, coletados por meio da implantação de sensores em locais estratégicos da cidade de Melbourne, na Austrália<sup>5</sup>, durante o período de 21 de dezembro de 2014 a 11 de janeiro de 2015. Os eventos interessantes revelados pelo método de detecção de anomalias forneceram informações úteis para a equipe florestal de Melbourne.

Santos et al. (2018) introduziram uma solução de detecção de anomalias para aplicações em CI, com base nas vantagens do paradigma *Fog Computing*. O cenário de avaliação da solução foi o mesmo utilizado em Santos et al. (2017), um caso de uso de monitoramento da qualidade do ar. Foram utilizados dois *datasets* sobre os poluentes: PM1, PM2.5 e PM10. Esses *datasets* foram coletados no trabalho anterior dos autores (SANTOS et al., 2017). Para detectar *outliers* nos *datasets*, Santos et al. (2018) escolheram a técnica *Robust Covariance*. Também foi utilizado o algoritmo de *clustering Birch*, com intenção de encontrar padrões nos *datasets*. Tanto o RC como o *Birch* foram escritos em *Python* usando a biblioteca *Scikit-Learn*. Assim como no trabalho anterior de Santos et al. (2017), os resultados alcançados pelas técnicas foram comparados.

Souza, Aquino e Gomes (2019) propuseram um novo método para detectar *outliers* em dados coletados por sensores urbanos inteligentes. O método proposto baseou-se na combinação das técnicas de decomposição de tensores, que utilizou o *Higher-order Singular Value Decomposition* (HOSVD), e classificação de dados, que fez uso do *K-Means*. Na avaliação do método de detecção de *outliers*, Souza, Aquino e Gomes (2019) usaram um *dataset* com medições de temperatura, umidade, brilho, ruído, monóxido de carbono e dióxido de nitrogênio de quatro cidades: Elda e Rois (Espanha), Nuremberg (Alemanha) e Taillinn (Estônia). O *dataset* foi coletado entre 01 de julho a 15 de julho de 2017 e obtido através da plataforma *Smart Citizen*<sup>6</sup>. Foi realizada também uma comparação entre o método de detecção proposto (HOSVD + *K-means*) e o *Multiway Principal Component Analysis* (MPCA), combinado ao *K-means*.

A Tabela 9 sumariza as principais informações descritas nesta seção.

<sup>4</sup> <https://airbox.edimaxcloud.com/>

<sup>5</sup> [http://issnip.unimelb.edu.au/research\\_program/Internet\\_of\\_Things/iot\\_deployment](http://issnip.unimelb.edu.au/research_program/Internet_of_Things/iot_deployment)

<sup>6</sup> <https://smartcitizen.me/>

Tabela 9: Principais informações extraídas dos estudos

<b>Autores</b>	<b>Objetivo</b>	<b>Técnica de Detecção de Outliers</b>	<b>Estudo de Caso</b>	<b>Origem dos Dados</b>
<a href="#">Mounce et al. (2015)</a>	Apresentar o site de demonstração <i>Thames Water</i> do projeto SW4EU	AURA-Alert	Analisar dados sobre o fluxo de água e identificar anomalias	Inglaterra
<a href="#">Trilles et al. (2016)</a>	Apresentar uma metodologia para analisar fluxo de dados em redes de sensores	CUSUM	Detectar anomalias em dados relacionados à poluição do ar	Espanha
<a href="#">Cicirelli et al. (2017)</a>	Apresentar uma <i>framework</i> para gerenciamento de uma rede de sensores em CI	Medidas Estatísticas (média e desvio padrão)	Detectar anomalias em dados de poluição sonora	Itália
<a href="#">Santos et al. (2017)</a>	Propor uma <i>framework</i> para gerenciamento de aplicativos de CI em redes 5G	RC e IF	Detectar anomalias em dados de poluição do ar e clima	Bélgica
<a href="#">Chen et al. (2018)</a>	Apresentar uma <i>framework</i> que visa garantir a qualidade de dados de sensoriamento ambiental	TSAD (método próprio)	Detectar anomalias em dados de poluição do ar	Taiwan
<a href="#">Rathore et al. (2018)</a>	Apresentar uma <i>framework</i> de geovisualização para monitoramento e análise de padrões em dados de microclima urbano	HyCARCE e ENOF	Detectar anomalias em dados climáticos	Austrália
<a href="#">Santos et al. (2018)</a>	Introduzir uma solução de detecção de anomalias para aplicações em CI	RC e <i>Birch</i>	Detectar anomalias em dados de poluição do ar	Bélgica
<a href="#">Souza, Aquino e Gomes (2019)</a>	Propor um novo método de detecção de <i>outliers</i>	HOSVD e <i>K-means</i>	Detectar <i>outliers</i> em dados climáticos e de poluição sonora/atmosférica	Espanha, Alemanha e Estônia

Com as informações apresentadas nesta seção, pretendeu-se fornecer uma visão geral sobre os estudos primários selecionados pela revisão sistemática executada. Para criar essa visão geral, foram descritas informações relevantes dos artigos, como: objetivo do artigo, técnica de detecção de *outliers* utilizada, estudo de caso, forma de aquisição, origem e contexto ambiental dos dados, entre outras.

## 4.2 Respostas das Questões de Pesquisa

**QP 1 - Quais métodos são utilizados na literatura para detectar *outliers* em dados ambientais no contexto de CI?**

Chen et al. (2018), Cicirelli et al. (2017) e Trilles et al. (2016) exploraram métodos de detecção de *outliers* baseados em estatística. Chen et al. (2018) desenvolveram o módulo TSAD para detectar três tipos de anomalias: espaciais, temporais e espaço-temporais, definidas estatisticamente. Para detectar as anomalias, o TSAD definiu valores limite, obtidos através de ajustes dos parâmetros do módulo. Segundo os autores, os parâmetros do TSAD devem ser atualizados periodicamente para que o esquema proposto seja eficaz.

Cicirelli et al. (2017) usaram as medidas estatísticas de média e desvio padrão para caracterizar uma distribuição Gama e produzir um agente que detectasse dois tipos de anomalias: de área (em tempo real) e históricas (durante um período passado). Trilles et al. (2016) empregaram o algoritmo CUSUM. Apesar de terem escolhido o CUSUM para identificar anomalias, os autores afirmaram que tal algoritmo apresentou limitações quando aplicado a dados ambientais.

Rathore et al. (2018) e Souza, Aquino e Gomes (2019) propuseram métodos de detecção de *outliers* compostos por uma combinação de técnicas. Rathore et al. (2018) utilizou o algoritmo HyCARCE para gerar *clusters* hiperelipsoidais, em seguida, aplicou o algoritmo ENOF para identificar os *clusters* anômalos. O método proposto teve como propósito encontrar anomalias locais (em um único sensor) e globais (em vários sensores). Souza, Aquino e Gomes (2019) combinaram a técnica HOSVD e o algoritmo *K-means*. A técnica HOSVD explorou a natureza multidimensional dos dados e o *K-means* extraiu padrões dos dados coletados.

Santos et al. (2017) fizeram uso de duas técnicas de detecção de *outliers* baseadas em aprendizado de máquina não supervisionado, foram elas: RC e IF. Em seu outro artigo, Santos et al. (2018) utilizou novamente a técnica RC só que, dessa vez, em conjunto com o algoritmo de *clustering Birch*. Os resultados das técnicas RC e *Birch* foram comparados para encontrar padrões ou eventos incomuns nos *datasets*.

Mounce et al. (2015) empregaram o método AURA-Alert. O AURA-Alert, utilizando *Correlation Matrix Memories* (CMMs), modelou o comportamento “normal” do sistema monitorado, assim, qualquer evento fora desse padrão foi identificado.

Com foco na questão ambiental, diversos métodos de detecção de *outliers* foram propostos e utilizados na literatura. A Tabela 10 apresenta os métodos de detecção de *outliers* utilizados e os respectivos artigos em que foram encontrados.

Tabela 10: Métodos de detecção de *outliers* utilizados

Trabalho	Método(s) Utilizado(s)
Mounce et al. (2015)	AURA-Alert
Trilles et al. (2016)	CUSUM
Cicirelli et al. (2017)	Medidas Estatísticas (média e desvio padrão)
Santos et al. (2017)	RC e IF
Chen et al. (2018)	TSAD (método próprio)
Rathore et al. (2018)	HyCARCE e ENOF
Santos et al. (2018)	RC e <i>Birch</i>
Souza, Aquino e Gomes (2019)	HOSVD e K-means

### QP 2 - Os resultados da detecção de *outliers* são validados?

A partir das informações extraídas, pode-se afirmar que não houve validações formais dos resultados em nenhum dos artigos. Talvez a ausência de validação possa ser explicada pelo fato de que a maioria dos artigos não tinham como foco principal a detecção de *outliers*. Outros pontos que também devem ser destacados são o pouco conhecimento dos autores sobre o domínio dos dados e a falta de consulta a especialistas.

Alguns trabalhos realizaram comparações entre os resultados de distintas técnicas de detecção de *outliers*, o que de certa forma tornou tais resultados mais confiáveis. Como foi o caso dos trabalhos de Santos et al. (2017) e Santos et al. (2018), que aplicaram duas técnicas aos mesmos *datasets* e compararam os resultados obtidos para, em seguida, realizarem uma análise global sobre os *outliers* identificados, utilizando a localização de GPS dos mesmos.

No trabalho de Cicirelli et al. (2017), foram apresentadas estatísticas relacionadas aos dados coletados pelos sensores de ruídos durante a primeira semana de junho de 2016. Cicirelli et al. (2017) constataram que as anomalias identificadas pelo o agente de anomalias, na data de 3 de junho de 2016, corresponderam a realidade. A validação deu-se através da verificação de que no referido dia ocorreu um protesto nas proximidades do sensor de ruído. Nenhuma validação foi realizada em relação às outras anomalias detectadas durante a semana.

### QP 3 - Quais os contextos do monitoramento ambiental são abordados pelos estudos?

A Tabela 11 apresenta os contextos ambientais tratados pelos artigos e a frequência em que cada um aparece.

Tabela 11: Contextos abordados pelos artigos e as frequências em que aparecem

<b>Contexto Ambiental</b>	<b>Frequência</b>	<b>Trabalhos</b>
Poluição Atmosférica	5	Trilles et al. (2016), Santos et al. (2017), Chen et al. (2018), Santos et al. (2018) e Souza, Aquino e Gomes (2019)
Dados Climáticos	3	Santos et al. (2017), Rathore et al. (2018) e Souza, Aquino e Gomes (2019)
Poluição Sonora	2	Cicirelli et al. (2017) e Souza, Aquino e Gomes (2019)
Vazamento de Água Potável	1	Mounce et al. (2015)

Os estudos selecionados abordaram quatro contextos dentro do monitoramento ambiental, foram eles: poluição atmosférica e sonora, dados referentes ao clima e vazamento de água potável. A poluição atmosférica foi o tema mais recorrente, sendo abordado em cinco trabalhos: Chen et al. (2018), Santos et al. (2017), Santos et al. (2018), Trilles et al. (2016) e Souza, Aquino e Gomes (2019).

## 5 Conclusão

Este trabalho teve como objetivo principal o levantamento e apresentação dos métodos utilizados na literatura para detectar *outliers* em dados ambientais urbanos. Para atingir esse objetivo, uma revisão sistemática da literatura foi planejada e executada. Após a execução da revisão sistemática, foram identificados 293 estudos primários. Dos 293 estudos, oito foram selecionados, pois satisfizeram os critérios de seleção e às questões de pesquisa. Os oito estudos primários foram lidos integralmente, buscando informações necessárias para responder às questões de pesquisa. Também foi fornecida uma visão geral sobre cada um dos estudos, enfatizando informações consideradas importantes para o contexto deste trabalho.

Com a realização da revisão sistemática, percebeu-se que uma grande variedade de métodos de detecção de *outliers*, baseados em distintas abordagens, foram empregados nas pesquisas. Notou-se que os pesquisadores não validaram os resultados obtidos pelas técnicas de detecção de *outliers* utilizadas. Assim, a questão de validação dos resultados apresenta-se como um dos desafios a serem explorados para tal contexto. Constatou-se também que a maioria dos autores possuíam pouco conhecimento sobre o domínio dos dados utilizados. Em razão disso, técnicas baseadas em clusterização tiveram bastante destaque, sendo usadas em conjunto com outras técnicas de detecção de *outliers*. Além do que já foi citado, observou-se ainda que a literatura existente sobre o tema de detecção de *outliers* em dados ambientais urbanos tendeu muito à questão da poluição atmosférica.

Como trabalhos futuros, pretende-se ampliar a revisão sistemática para outras bases de dados, bem como, utilizar outras técnicas, como *snowballing*, somadas a busca automatizada, visando alcançar uma revisão mais abrangente e completa possível. Pretende-se ainda efetuar comparações entre os métodos de detecção de *outliers* encontrados, com intuito de comprovar a acurácia de tais métodos. Além disso, planeja-se elaborar artigos científicos com base neste trabalho para a publicação em grandes eventos nacionais e internacionais relacionados ao tópico de Cidades Inteligentes.

# Referências

- AGGARWAL, C. C. Outlier analysis. *Springer Publishing Company, Incorporated*, 2013. Citado na página 20.
- AL-FUQAHA, A. et al. Internet of things: A survey on enabling technologies, protocols, and applications. *IEEE communications surveys & tutorials*, IEEE, v. 17, n. 4, p. 2347–2376, 2015. Citado na página 17.
- ANDERLE, D. F.; JUNIOR, V. F. A utilização da tecnologia da informação nas smart cities – um estudo bibliométrico. *Instituto Federal Catarinense, sd*, 2013. Citado na página 17.
- ANDRADE, J. N.; GALVÃO, D. C. O conceito de smart cities aliado à mobilidade urbana. *Revista Hum@nae*, v. 10, n. 1, p. 5, 2016. Citado na página 16.
- AUNE, A.; GONÇALVES, S. R. *Human Smart Cities – O cenário brasileiro e a importância da abordagem joined-up na definição de Cidade Inteligente*. Dissertação (Mestrado) — PUC-Rio, 2017. Citado na página 15.
- BARBOSA, G. B. N. *Sistema de segurança para IoT baseado em agrupamento de smart cards gerenciados por FPGA*. Dissertação (Mestrado) — Universidade Federal de Pernambuco, 2017. Citado na página 18.
- BARNETT, V.; LEWIS, T. *Outliers in statistical data*. [S.l.]: Wiley, 1984. (Wiley series in probability and mathematical statistics: Applied probability and statistics). ISBN 9780471905073. Citado na página 19.
- BARROS, C. de M. *Identificação de pontos influentes em uma amostra da distribuição de Watson*. Dissertação (Mestrado) — Universidade Federal de Pernambuco, 2014. Citado na página 19.
- BATTY, M. et al. Smart cities of the future. *The European Physical Journal Special Topics*, Springer, v. 214, n. 1, p. 481–518, 2012. Citado na página 16.
- BECKMAN, R. J.; COOK, R. D. Outlier..... s. *Technometrics*, Taylor & Francis Group, v. 25, n. 2, p. 119–149, 1983. Citado na página 19.
- BELFIORE, P. F. *Estatística aplicada a administração, contabilidade e economia com excel e SPSS*. [S.l.]: Elsevier Editora Ltda., 2015. ISBN 9788535263565. Citado na página 13.
- BORGIA, E. The internet of things vision: Key features, applications and open issues. *Computer Communications*, Elsevier, v. 54, p. 1–31, 2014. Citado na página 18.
- CARAGLIU, A.; BO, C. D.; NIJKAMP, P. Smart cities in europe. *Proceedings of the 3rd Central European Conference on Regional Science, Košice*, 2009. Citado na página 16.
- CHANDOLA, V.; BANERJEE, A.; KUMAR, V. Anomaly detection: A survey. *ACM computing surveys (CSUR)*, ACM, v. 41, n. 3, p. 15, 2009. Citado 2 vezes nas páginas 19 e 20.

- CHEN, L.-J. et al. Adf: An anomaly detection framework for large-scale pm2.5 sensing systems. *IEEE Internet of Things Journal*, IEEE, v. 5, n. 2, p. 559–570, 2018. Citado 6 vezes nas páginas 27, 30, 32, 33, 34 e 35.
- CICIRELLI, F. et al. An edge-based platform for dynamic smart city applications. *Future Generation Computer Systems*, Elsevier, v. 76, p. 106–118, 2017. Citado 6 vezes nas páginas 27, 30, 32, 33, 34 e 35.
- CILISKA, D.; CULLUM, N.; MARKS, S. Evaluation of systematic reviews of treatment or prevention interventions. *Evidence-Based Nursing*, Royal College of Nursing, v. 4, n. 4, p. 100–104, 2001. Citado na página 22.
- CUNHA, M. A. et al. *Smart Cities: Transformação digital de cidades*. [S.l.]: Programa Gestão Pública e Cidadania, 2016. Citado na página 16.
- DERITTI, M. D.; FREIRE, F. O surgimento do conceito cim através da interoperabilidade entre bim e gis e sua importância para as smart cities. *Revista Técnico-Científica*, n. 13, p. 2, 2018. Citado na página 15.
- DICKSON, R. Systematic reviews. *Achieving evidence-based practice: A handbook for practitioners*, Harcourt Brace, London, p. 41–60, 1999. Citado na página 22.
- FALBO, R. de A. Mapeamento sistemático. *Retrieved October*, v. 7, 2018. Citado 3 vezes nas páginas 23, 24 e 25.
- GALVÃO, C. M.; SAWADA, N. O.; TREVIZAN, M. A. Revisão sistemática: recurso que proporciona a incorporação das evidências na prática da enfermagem. *Rev Latino-am enfermagem*, SciELO Brasil, v. 12, n. 3, p. 549–556, 2004. Citado na página 22.
- GARCIA-FONT, V.; GARRIGUES, C.; RIFÀ-POUS, H. Difficulties and challenges of anomaly detection in smart cities: A laboratory analysis. *Sensors*, Multidisciplinary Digital Publishing Institute, v. 18, n. 10, p. 3198, 2018. Citado na página 15.
- GAUCHAZH. *Mais de metade da população mundial já vive em áreas urbanas, diz ONU*. 2014. Disponível em: <<https://gauchazh.clicrbs.com.br/educacao-e-emprego/noticia/2014/07/Mais-de-metade-da-populacao-mundial-ja-vive-em-areas-urbanas-diz-ONU-4547970.html>>. Acesso em: 27 jul. 2018. Citado na página 13.
- GIFFINGER, R. et al. Smart cities - ranking of european medium-sized cities. 2007. Citado na página 16.
- GREENHALGH, T. Papers that summarise other papers (systematic reviews and meta-analyses). *BMJ: British Medical Journal*, BMJ Publishing Group, v. 315, n. 7109, p. 672, 1997. Citado na página 22.
- GUEDES, A. et al. *Cidades inteligentes - “smart cities”*. 2014. Disponível em: <<https://paginas.fe.up.pt/>>. Acesso em: 24 jan. 2019. Citado na página 16.
- HABEEB, R. A. A. et al. Real-time big data processing for anomaly detection: A survey. *International Journal of Information Management*, Elsevier, 2018. Citado na página 20.
- HAWKINS, D. Identifications of outliers, monograph on applied probability and statistic. *Reading, London Chapman and Hall*, 1980. Citado na página 13.

- KHAN, M. A.; SALAH, K. Iot security: Review, blockchain solutions, and open challenges. *Future Generation Computer Systems*, Elsevier, v. 82, p. 395–411, 2018. Citado na página 17.
- KOMNINOS, N.; SCHAFFERS, H.; PALLOT, M. Developing a policy roadmap for smart cities and the future internet. In: *45th Hawaii International Conference on System Sciences*. [S.l.: s.n.], 2011. Citado na página 16.
- LOPES, E. E. *Smart Cities: conheça as cidades inteligentes do futuro*. 2017. Disponível em: <<https://portogente.com.br/noticias/meio-ambiente/99018-smart-cities-conheca-as-cidades-inteligentes-do-futuro>>. Acesso em: 24 jan. 2019. Citado na página 15.
- MIRANDA, R. C. G. de. Uma revisão sistemática sobre equipes de desenvolvimento de software: Tipologia, características e critérios de formação. Universidade Federal de Pernambuco, 2011. Citado na página 22.
- MORAES T. M.; SOUZA, A. d. Revisão sistemática sobre a comunicação dentro do processo de desenvolvimento de software. *Universidade Federal de Goiás-GO*, p. 57, 2011. Citado na página 22.
- MOUNCE, S. et al. Cloud based machine learning approaches for leakage assessment and management in smart water networks. *Procedia Engineering*, Elsevier, v. 119, p. 43–52, 2015. Citado 6 vezes nas páginas 27, 29, 32, 33, 34 e 35.
- MUNZLINGER, E.; NARCIZO, F. B.; QUEIROZ, J. E. R. de. Sistematização de revisões bibliográficas em pesquisas da área de ihc. In: BRAZILIAN COMPUTER SOCIETY. *Companion Proceedings of the 11th Brazilian Symposium on Human Factors in Computing Systems*. [S.l.], 2012. p. 51–54. Citado na página 23.
- NAM, T.; PARDO, T. A. Conceptualizing smart city with dimensions of technology, people, and institutions. In: ACM. *Proceedings of the 12th annual international digital government research conference: digital government innovation in challenging times*. [S.l.], 2011. p. 282–291. Citado na página 16.
- ODENDAAL, N. Information and communication technology and local governance: understanding the difference between cities in developed and emerging economies. *Computers, Environment and Urban Systems*, Elsevier, v. 27, n. 6, p. 585–607, 2003. Citado na página 16.
- PAPA, R. et al. Smart and resilient cities. a systemic approach for developing cross-sectoral strategies in the face of climate change. *TeMA Journal of Land Use, Mobility and Environment*, Università degli studi di Napoli Federico II, Dipartimento di Ingegneria . . . , v. 8, n. 1, p. 19–49, 2015. Citado na página 16.
- PENG, S.; SHEN, H. Security technology analysis of iot. In: *Internet of Things*. [S.l.]: Springer, 2012. p. 401–408. Citado na página 17.
- PRADO, K. C. D.; SANTOS, P. E. d. Smart cities: Conceito, iniciativas e o cenário carioca. *Escola Politécnica da Universidade Federal do Rio de Janeiro*. Rio de Janeiro, 2014. Citado na página 16.

- QIAN, Z. et al. M/i adaptation layer network protocol for iot based on 6lowpan. In: *Internet of Things*. [S.l.]: Springer, 2012. p. 208–215. Citado na página 17.
- RATHORE, P. et al. Real-time urban microclimate analysis using internet of things. *IEEE Internet of Things Journal*, IEEE, v. 5, n. 2, p. 500–511, 2018. Citado 6 vezes nas páginas 27, 31, 32, 33, 34 e 35.
- RIZZON, F. et al. Smart city: um conceito em construção. *Revista Metropolitana de Sustentabilidade*, v. 7, n. 3, 2017. Citado na página 16.
- RODRIGUES, R. D. *Deteção de outliers baseada em caminhada determinística do turista*. Dissertação (Mestrado) — Universidade de São Paulo, 2018. Citado 2 vezes nas páginas 20 e 21.
- RUSSO, F.; RINDONE, C.; PANUCCIO, P. The process of smart city definition at an eu level. *WIT Transactions on Ecology and the Environment*, WIT Press, v. 191, p. 979–989, 2014. Citado na página 16.
- SANTOS, J. et al. Anomaly detection for smart city applications over 5g low power wide area networks. In: IEEE. *NOMS 2018-2018 IEEE/IFIP Network Operations and Management Symposium*. [S.l.], 2018. p. 1–9. Citado 6 vezes nas páginas 27, 31, 32, 33, 34 e 35.
- SANTOS, J. et al. Fog computing: Enabling the management and orchestration of smart city applications in 5g networks. *Entropy*, Multidisciplinary Digital Publishing Institute, v. 20, n. 1, p. 4, 2017. Citado 7 vezes nas páginas 27, 30, 31, 32, 33, 34 e 35.
- SILVA, F. R. et al. *Uma abordagem para detecção de outliers em dados categoricos*. Dissertação (Mestrado) — Universidade Estadual de Campinas, 2004. Citado na página 19.
- SILVER, K. *Poluição mata mais de 100 mil pessoas por ano no Brasil, diz relatório*. 2017. Disponível em: <<https://www.bbc.com/portuguese/geral-41692503>>. Acesso em: 10 jan. 2019. Citado na página 13.
- SOUZA, T. I.; AQUINO, A. L.; GOMES, D. G. A method to detect data outliers from smart urban spaces via tensor analysis. *Future Generation Computer Systems*, Elsevier, v. 92, p. 290–301, 2019. Citado 5 vezes nas páginas 27, 31, 33, 34 e 35.
- SU, X.; TSAI, C.-L. Outlier detection. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, Wiley Online Library, v. 1, n. 3, p. 261–268, 2011. Citado na página 19.
- TOMAS, G. H. R. P. *Uma arquitetura para cidades inteligentes baseada na internet das coisas*. Universidade Federal de Pernambuco, 2014. Citado 2 vezes nas páginas 18 e 19.
- TRILLES, S. O. et al. *A domain-independent methodology to analyze IoT data streams in real-time. A proof of concept implementation for anomaly detection from environmental data*. [S.l.]: Taylor & Francis, 2016. Citado 7 vezes nas páginas 27, 29, 30, 32, 33, 34 e 35.
- ZANELLA, A. et al. Internet of things for smart cities. *IEEE Internet of Things journal*, IEEE, v. 1, n. 1, p. 22–32, 2014. Citado na página 13.



TERMO DE AUTORIZAÇÃO PARA PUBLICAÇÃO DIGITAL NA BIBLIOTECA  
"JOSÉ ALBANO DE MACEDO"

Identificação do Tipo de Documento

- ( ) Tese  
( ) Dissertação  
(X) Monografia  
( ) Artigo

Eu, Anchi Lucas de Costa Soares,  
autorizo com base na Lei Federal nº 9.610 de 19 de Fevereiro de 1998 e na Lei nº 10.973 de  
02 de dezembro de 2004, a biblioteca da Universidade Federal do Piauí a divulgar,  
gratuitamente, sem ressarcimento de direitos autorais, o texto integral da publicação  
REVISÃO SISTEMÁTICA NO CONTEXTO DE DETECÇÃO DE DUTUEIS PARA  
MONITORAMENTO AMBIENTAL DE CIDADES INTELIGENTES  
de minha autoria, em formato PDF, para fins de leitura e/ou impressão, pela internet a título  
de divulgação da produção científica gerada pela Universidade.

Picos-PI 15 de JUNHO de 2019.

Anchi Lucas de Costa Soares  
Assinatura

\_\_\_\_\_  
Assinatura