



Instituto Superior de Contabilidade e Administração

Politécnico de Coimbra



**Instituto Superior
de Contabilidade
e Administração**

Politécnico de Coimbra

COIMBRA BUSINESS SCHOOL
ISCAC.pt

Luís Jorge Marques Ladeiro

***Big Data Analytics* na Energia:
Uma Revisão Sistemática da Literatura**

Coimbra, julho de 2022



**Instituto Superior
de Contabilidade
e Administração**

Politécnico de Coimbra

COIMBRA **BUSINESS SCHOOL**
ISCAC.pt

Luís Jorge Marques Ladeiro

***Big Data Analytics* na Energia:
Uma Revisão Sistemática da Literatura**

Dissertação submetida ao Instituto Superior de Contabilidade e Administração de Coimbra para cumprimento dos requisitos necessários à obtenção do grau de **Mestre em Sistemas de Informação de Gestão**, realizada sob a orientação da Professora Doutora Isabel Pedrosa e coorientação do Professor Doutor António Trigo.

Coimbra, julho de 2022

TERMO DE RESPONSABILIDADE

Declaro ser o autor desta dissertação, que constitui um trabalho original e inédito, que nunca foi submetido a outra Instituição de ensino superior para obtenção de um grau académico ou outra habilitação. Atesto ainda que todas as citações estão devidamente identificadas e que tenho consciência de que o plágio constitui uma grave falta de ética, que poderá resultar na anulação da presente dissertação.

AGRADECIMENTOS

Gostaria de deixar neste momento os meus agradecimentos a todos os que me acompanharam nesta jornada académica e que contribuíram para o sucesso da realização da minha dissertação.

Num ponto inicial agradeço à minha família pelo acompanhamento durante todo o meu percurso académico porque sem eles certamente não teria conseguido concluir os meus estudos, especialmente pelos seus contributos no que toca a manter-me focado nos meus objetivos, assim como, as suas palavras motivadoras.

Agradeço também aos meus orientadores, professora Isabel Pedrosa e professor António Trigo pela sua ajuda durante todo o processo de realização da dissertação.

Por fim gostaria de agradecer a todos aqueles com que me cruzei durante a minha vida académica, desde colegas de licenciatura e mestrado, a professores: a conclusão desta fase da minha vida não seria possível sem a vosso apoio.

Luís Ladeiro

RESUMO

A transformação para uma sociedade mais digital tem vindo a produzir uma quantidade de dados incomparável com décadas anteriores. *Big data analytics* é agora um termo muito conhecido ao nível de dados financeiros, saúde, marketing, logística, entre outros. Sendo a energia uma das áreas afetadas pela transformação digital, como tem sido aplicado o conceito de *big data analytics* nesta área? Que aplicações surgiram devido ao volume de dados recolhidos? Que ferramentas são utilizadas? Como tem sido afetada por esta mudança?

Esta dissertação pretende responder à pergunta-chave “Como tem sido aplicado *big data analytics* na área da energia?”. Para o efeito procede-se à realização de uma revisão sistemática da literatura, a partir da qual se elabora uma introdução das temáticas discutidas para leitores que possam estar a iniciar estudos nesta área de investigação, garantindo, ao mesmo tempo, que esta seja útil para leitores que já se encontram inseridos nestas áreas.

Para a realização da revisão foi consultada a base de dados Science Direct, sendo construída uma frase booleana através das palavras-chave selecionadas durante a fase de pesquisa da base de dados sendo estas “*Big Data*”, “*Analytics*”, “*Energy*”, “*Smart Grid*”, “*Smart Meters*”, “*Smart Sensor Network*”, “*Meter Data Analytics*”, “*Energy Management*” e “*Energy Consumption*” tendo sido realizado um estudo bibliométrico aos resultados obtidos. Para a seleção dos estudos foram utilizados três métodos de filtragem: um primeiro centrado nos critérios de seleção; o segundo através da leitura dos títulos das publicações e dos resumos; e o terceiro através da realização de um teste de relevância. A exploração da informação permitiu concluir que *big data analytics* na energia tem sido aplicada em áreas como o estudo do consumidor, soluções para análise de dados recolhidos por *smart meters*, desenvolvimento de plataformas de análise das *smart grids*, segurança e gestão inteligente da energia que se refletem na forma como a energia é canalizada, aproveitada e utilizada através da associação de técnicas, processos e sistemas planeados para análise de grandes volumes de dados, resultando numa tomada de decisões mais consciente pelos fornecedores e consumidores de energia.

Palavras-chave: *Big Data Analytics*, Energia, *Data Mining*, *Smart Energy Management*, Revisão Sistemática de Literatura.

ABSTRACT

The transformation towards a more digital society introduced a greater quantity of collected data. Big Data Analytics is now a common term surrounding data in finance, health, and other fields. With energy being one of the fields affected by this digital transformation, how have the Big Data Analytics concepts been applied to this field? What applications have surged due to the collected amount of data? What tools are used? How has it been affected?

This master's thesis wants to answer the main question "How has big data analytics been applied to energy?". For that effect a systematic review is produced, in which an introduction to the discussed themes is produced for readers who are beginning studies in this field of research, ensuring at the same time that it is also useful for readers who already are in these fields.

To produce this review the Science Direct database was consulted, using a Boolean sentence produced using keywords found during the research phase of the database, with the keywords "*Big Data*", "*Analytics*", "*Energy*", "*Smart Grid*", "*Smart Meters*", "*Smart Sensor Network*", "*Meter Data Analytics*", "*Energy Management*" e "*Energy Consumption*", a bibliometric study was also applied to the obtained results. For the selection of studies, three methods were used for the filtering. The first was based on the inclusion and exclusion criteria, the second based on the reading of titles and abstracts, and the third based on a relevancy test. Exploring the information allowed to conclude that big data analytics in energy has been applied to fields such as the study of the consumer, solutions for data analysis gathered from smart meters, development of platforms for analysis of smart grids, security and intelligent energy management, that reflect upon the way energy is gathered, harnessed and used with techniques, processes and systems planned for analysis of big volumes of data resulting in a more conscious decision making by providers and consumers.

Keywords: Big Data Analytics, Energy, Data Mining, Smart Energy Management, Systematic Literature Review

ÍNDICE GERAL

| | | |
|---------|---|----|
| 1 | INTRODUÇÃO..... | 1 |
| 1.1 | Motivação | 3 |
| 1.2 | Objetivos..... | 3 |
| 1.3 | Metodologia..... | 3 |
| 1.4 | Estrutura da Dissertação | 4 |
| 2 | ENQUADRAMENTO TEÓRICO | 6 |
| 2.1 | <i>Big Data Analytics</i> | 6 |
| 2.2 | <i>Internet of Things</i> e a recolha de <i>big data</i> | 9 |
| 2.3 | Armazenamento e processamento de <i>big data</i> | 10 |
| 2.4 | Visualização de <i>big data</i> | 11 |
| 2.5 | <i>Big Data Analytics</i> na energia | 13 |
| 2.5.1 | <i>Data Mining</i> na energia | 17 |
| 2.5.2 | <i>Energy Internet</i> | 19 |
| 2.5.3 | Desafios e oportunidades de <i>big data analytics</i> no setor..... | 20 |
| 3 | ENQUADRAMENTO METODOLÓGICO | 23 |
| 3.1 | Tipologia do estudo | 23 |
| 3.2 | O protocolo aplicado..... | 24 |
| 3.2.1 | Estratégia de pesquisa..... | 24 |
| 3.2.2 | Base de dados utilizada..... | 27 |
| 3.2.3 | Processo de seleção | 27 |
| 3.2.3.1 | Exploração bibliométrica | 29 |
| 3.2.3.2 | Filtro de resultados | 31 |

| | | |
|-------|--|----|
| 3.2.4 | Recolha de dados | 37 |
| 4 | TRABALHO EMPÍRICO | 38 |
| 4.1 | Sumário dos estudos selecionados | 38 |
| 4.2 | Análise dos resultados obtidos..... | 39 |
| 4.2.1 | Estudo do consumidor de energia..... | 39 |
| 4.2.2 | <i>Smart meters e big data analytics</i> | 42 |
| 4.2.3 | <i>Smart Grids</i> | 47 |
| 4.2.4 | Segurança da energia | 56 |
| 4.2.5 | Gestão inteligente da energia..... | 60 |
| 4.3 | Discussão dos conteúdos explorados | 65 |
| 5 | CONCLUSÃO..... | 74 |
| 5.1 | Contributos..... | 75 |
| 5.2 | Limitações..... | 77 |
| 5.3 | Trabalho Futuro | 77 |
| | REFERÊNCIAS BIBLIOGRÁFICAS | 79 |
| | APÊNDICES | 90 |
| | APÊNDICE 1. CLUSTERS OBTIDOS NA EXPLORAÇÃO BIBLIOMÉTRICA. | 91 |
| | ANEXOS | 94 |
| | ANEXO 1. <i>CHECKLIST</i> PRISMA | 95 |
| | ANEXO 2. <i>CHECKLIST</i> PRISMA (PORTUGUÊS)..... | 98 |

ÍNDICE DE FIGURAS

| | |
|---|----|
| Figura 1.1. Estrutura da Dissertação..... | 5 |
| Figura 2.1. As três camadas de Big Data..... | 7 |
| Figura 2.2. Dashboard de visualização de consumo energético | 13 |
| Figura 2.3. Processo de big data analytics..... | 16 |
| Figura 2.4. Os principais componentes de Smart Grid..... | 16 |
| Figura 3.1. Protocolo utilizado | 24 |
| Figura 3.2. Frase Booleana considerada para a exploração..... | 28 |
| Figura 3.3. Network Visualization do VOSViewer..... | 30 |
| Figura 3.4. Análise temporal do uso das palavras-chave | 31 |
| Figura 3.5. Diagrama de Resultados da Exploração da Base de dados | 33 |
| Figura 4.1. Temas e tecnologias | 65 |

ÍNDICE DE TABELAS

| | |
|--|----|
| Tabela 3.1. Critérios de Inclusão e Exclusão para a Revisão Sistemática da Literatura..... | 26 |
| Tabela 3.2. Teste de Relevância para artigos encontrados | 32 |
| Tabela 3.3. Artigos Filtrados | 34 |
| Tabela 3.4. Publicações selecionadas para a extração de dados..... | 35 |
| Tabela 4.1. Autores e temas dos artigos abordados..... | 66 |
| Tabela 4.2. Ideias de trabalho futuro presentes nos artigos analisados | 68 |

LISTA DE ABREVIATURAS, ACRÓNIMOS E SIGLAS

DM – *Data Mining*

ETL – *Extract, Transform and Load*

HDFS – *Hadoop Distributed File System*

IEN – *Intelligent Energy Network*

IoT – *Internet of Things*

RSL – *Revisão Sistemática da Literatura*

SCADA – *Supervisory Control and Data Acquisition*

TIC – *Tecnologias da Informação e Comunicação*

1 INTRODUÇÃO

A relevância do papel da energia na sociedade é, nos dias de hoje, uma realidade inquestionável. O uso da energia é fundamental para questões de aquecimento, alimentação e iluminação pelo que, hoje, consideramos a energia como sendo um bem essencial para a sobrevivência da sociedade. Ser um bem essencial implica que o seu uso é diversificado e massificado o que, por si, garante a produção de uma grande quantidade de dados relativos à sua produção, preparação e consumo.

O desenvolvimento acelerado de tecnologia de recolha de dados como sensores, tecnologia de transmissão via *wireless*, *cloud computing*, *smartphones* e o uso de aplicações móveis, tem levado a um aumento considerável de capacidade de recolha e processamento de grandes quantidades de dados (Lyu & Liu, 2021). O setor energético, produtor de grandes volumes de dados relativos à produção e consumo de energia tem verificado uma digitalização dos seus sistemas devido à incorporação das tecnologias mencionadas que permitem essa geração de dados e a análise dos mesmos (Zhou, Fu, et al., 2016).

O crescimento do volume de dados gerados pelo setor energético tem criado vários desafios para este setor ao nível do tratamento dos dados recolhidos, pelo que *big data* associada ao setor de energia é, nos dias de hoje, uma temática que não pode ser ignorada (Yang Zhang et al., 2018).

Os dados diferem também em função da forma como são recolhidos. Os dados utilizados podem existir de diversas formas, pelo que as fontes dos dados são extremamente importantes para a descrição das suas potenciais aplicações. Dependendo da fonte de dados, a recolha dos mesmos pode variar, pelo que devem ser, então, considerados os dados recolhidos de fontes exteriores a aplicações energéticas, mas que têm impacto nos dados: dados recolhidos de fontes terceiras à rede, como clima e preços de eletricidade, devem ser também considerados (Azamat Marlen., Askar Maxim., Ikechi A. Ukaegbu., H. S. V. S. Kumar Nunna., 2019).

Segundo Duan & Xiong (2015), atualmente todos os avanços tecnológicos, económicos entre outros provêm de grandes volumes de dados . No entanto, estes dados por si só não

são suficientes para justificarem os avanços verificados nos últimos tempos. Os dados detêm padrões “escondidos” que requerem o uso de técnicas de análise e apresentação para que a informação, extraída e tratada dê lugar a novas soluções e aplicações eficientes. Os autores afirmam ainda que o termo “*Data Analytics*” é normalmente utilizado para descrever todo este processo de análise de dados recolhidos, com o objetivo de com essa análise tornar os dados mais compreensíveis, permitindo assim retirar valor dos mesmos.

A transformação para uma sociedade mais digital criou novas necessidades e aplicações ao nível do tratamento e compreensão de grandes volumes de dados. *Big data analytics* é agora um termo muito conhecido ao nível de dados financeiros, saúde entre outros (Hariri et al., 2019). A energia é uma das áreas afetadas ao nível de transformação digital nos últimos anos, em especial com a integração das tecnologias de informação associadas a esta área que, em conjunto, originam grandes volumes de dados e transmissão dos mesmos (Zhou, Yang, et al., 2016). Tendo esta informação em conta torna-se necessário questionar os resultados produzidos por todos esses dados recolhidos, levando à colocação da questão de como tem sido aplicada *big data analytics* na área.

Como resposta a esta necessidade de entender como tem sido realizada a aplicação de *big data analytics* no setor energético surge esta Revisão Sistemática da Literatura (RSL), sendo, através da RSL, apresentados os desenvolvimentos existentes no que toca, no setor energético, à aplicação da tecnologia *big data*, técnicas para a análise dos dados recolhidos e o que surgiu das suas aplicações ao nível de potencial e novas oportunidades desenvolvidas na área, assim como, os seus efeitos para áreas conexas.

A RSL compreende a realização de um estudo que não pretende criar novo conhecimento, mas onde se procede à compilação, sintetização e sumarização da informação existente acerca de um tema definido, encontrado em várias publicações científicas, sendo procurada nos estudos a informação relevante à pergunta que dá origem à necessidade de realizar a revisão, com a informação recolhida constitui-se a resposta à questão colocada (Aromataris & Pearson, 2014).

1.1 Motivação

A principal motivação desta dissertação é categorizar a investigação realizada na área de *big data analytics* através de uma revisão sistemática da literatura que retrate a realidade atual da investigação científica sobre o tema *big data analytics* na energia, apresentando o mesmo de uma forma a ser útil para quem se encontra dentro da área de análise de dados de energia, assim como para pessoas que não detenham conhecimento prévio acerca da área de energia ou de análise de dados e que consigam, através da exploração desta dissertação, ser introduzidos ao tema.

O tema de análise de dados não é novo: no que toca às áreas que são hoje consideradas tendências ao nível de aplicação de *big data analytics* verificamos a saúde, os recursos naturais e os seus processos, o setor público, comércio, negócios e sistemas económicos as redes sociais e processos computacionais (Kambatla et al., 2014). Porém, sente-se uma necessidade de sistematização desta temática quando relacionada com a energia.

1.2 Objetivos

O objetivo central desta RSL passa por desenvolver o estado da arte respondendo à pergunta-chave “Como tem sido aplicada *big data analytics* na área da energia?”, garantindo um entendimento do tema que possa servir de base a estudos futuros.

Pretende-se ainda com este trabalho garantir a identificação e exposição de aplicações de *big data analytics* utilizadas no setor da energia e áreas conexas a este, bem como, o seu papel na previsão do consumo de energia, comportamento dos consumidores, voltagens utilizadas, importância da *Internet of Things* (IoT), entre outros.

Durante o desenvolvimento da RSL, pretende-se dar a conhecer as alterações na área da energia decorrentes da introdução de *big data analytics*, a evolução da pesquisa realizada na área, assim como, o impacto a nível de desafios e oportunidades desenvolvidas fruto do estudo de grandes volumes de dados da energia.

1.3 Metodologia

De modo a dar início à realização do trabalho é necessário em primeiro lugar apresentar o processo desenvolvido e utilizado para a realização do mesmo. Utilizando as

orientações do *Preferred Reporting Items for Systematic Reviews and Meta-Analyses* (PRISMA) (PRISMA, 2020) e considerando o objetivo de realização da RSL sobre o tema de *big data analytics* na energia foram adotados os seguintes passos para a sua realização:

1. Desenvolvimento da estratégia de pesquisa da informação que compõe a revisão sistemática a realizar.
2. Realização de uma pesquisa à base de dados selecionada descobrindo assim um conjunto de artigos científicos e temáticas desenvolvidas que mais valor produzem para a revisão.
3. Avaliação e filtro de artigos tendo em conta pressupostos selecionados juntamente com critérios de inclusão e exclusão dos mesmos, para uma garantia da agregação dos artigos que mais valor produzem para a revisão a efetuar.
4. Recolha de dados dos artigos selecionados, sintetizando e apresentando os mesmos num formato que satisfaça o cumprimento dos objetivos e possibilitem, a qualquer leitor, uma leitura informada acerca do tema.
5. Apresentação dos resultados alcançados e da informação que compõe a revisão. Realização de um comentário conclusivo acerca da situação atual de estudo das áreas, apresentando os aspetos centrais existentes nas mesmas e áreas menos exploradas no que corresponde à discussão dos resultados obtidos de modo a potencializar futuras pesquisas que relacionem os dois temas em estudo.

Atendendo à natureza do método e à sua complexidade, este será explicado em capítulo autónomo (Capítulo 3).

1.4 Estrutura da Dissertação

Esta dissertação encontra-se estruturada em cinco capítulos. O capítulo 1 refere-se à “Introdução” que compreende nela a motivação para o estudo, os objetivos definidos, a justificação e descrição da metodologia utilizada. O capítulo 2 compreende um enquadramento teórico introduzindo o tema de *big data* e *big data analytics*, assim como, as diferentes fases do processo e intervenientes no mesmo fazendo menção a temas como a IoT, armazenamento e processamento de informação recolhida, *Data Mining* (DM) e o papel destes na *big data analytics* na energia. O capítulo 3 detém o enquadramento metodológico, onde é apresentado o protocolo a seguir, a estratégia de pesquisa, critérios

de inclusão e exclusão, bases de dados utilizadas e a recolha de dados realizada. O capítulo 4 passa pela apresentação dos resultados obtidos com a descrição dos estudos recolhidos e consequente análise dos resultados obtidos. Por fim, temos o último capítulo com as conclusões que foram atingidas com o desenvolvimento do trabalho, descrevendo nas mesmas, os contributos e as limitações encontradas durante a realização da dissertação e a partilha de sugestões de trabalhos futuros. A figura 1.1 demonstra o esquema visual desta estrutura.

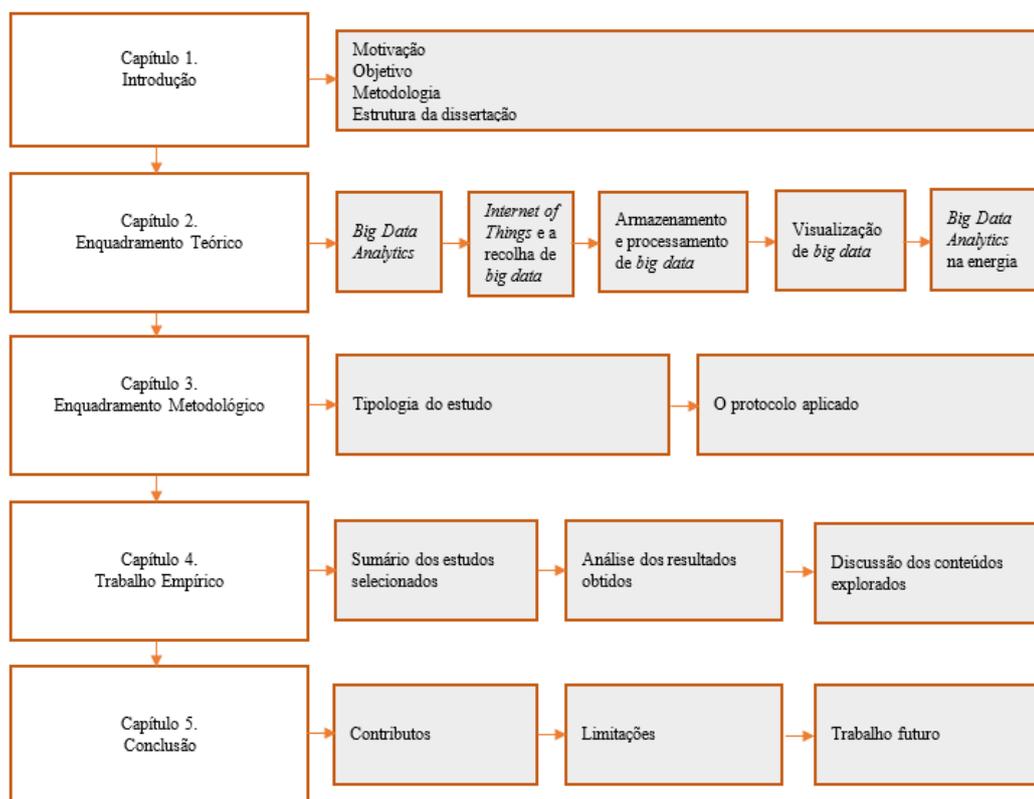


Figura 1.1. Estrutura da Dissertação

A exploração do tema e o desenvolvimento da revisão dependem, primariamente, da informação disponível acerca do mesmo, no entanto, a mesma informação encontra-se disponibilizada para investigadores com conhecimento na área ou pessoas que sejam expostas a esta pela primeira vez. Tendo esse ponto em atenção e tendo a produção desta revisão sido realizada garantindo que produz valor para ambos os potenciais utilizadores foi adotada uma metodologia de produção da RSL que garante o cumprimento deste objetivo. Para tal, antes de se proceder à exploração das aplicações de *big data* em energia é realizado um enquadramento teórico do tema chave e subtemas que o rodeiam.

2 ENQUADRAMENTO TEÓRICO

No seguinte capítulo encontra-se o enquadramento teórico que permite situar o tema garantindo a compreensão das diferentes componentes que compreendem o processo de análise de dados relacionando esse mesmo processo com os dados de energia recolhidos. Assim, abordam-se *big data* e *big data analytics*, o papel da IoT e o seu impacto na recolha de *big data* juntamente com áreas relevantes para o processo de análise de *big data* como o armazenamento e processamento, assim como a sua visualização. Após estabelecer as diversas componentes termina-se este capítulo com o processo de *big data analysis* quando centralizado na energia falando ainda de *DM* de dados energéticos, a junção de IoT e energia, apelidada *energy internet* e concluindo através da indicação de desafios existentes e oportunidades geradas com a introdução das tecnologias de exploração de grandes volumes de dados no setor energético.

2.1 *Big Data Analytics*

Big data, nos dias de hoje, pode compreender diversas definições: numa perspetiva mais redutora, *big data* corresponde ao termo utilizado para descrever uma quantidade de dados astronómica e de complexidade elevada, reconhecida como sendo de difícil processamento aquando do uso de técnicas tradicionais de processamento (Jin et al., 2015). Geralmente, o conceito *Big Data* é também apresentado pelos seus V's. Inicialmente estes compreendiam a Velocidade com que os dados são acedidos o Volume de dados que é cada vez mais elevado através dos diversos avanços nos procedimentos para a recolha dos mesmos e a Variedade dos mesmos visto estes poderem ser representados sobre diversas formatos e recolhidos a partir de diversas fontes (Tsai et al., 2015).

A complexidade e crescimento do termo tem, no entanto, levado a um novo olhar sobre a sua definição tradicional em que a conclusão chegada sobre a mesma foi a de que os 3Vs utilizados para definir *big data* são nos dias de hoje insuficientes para garantir uma visão completa sobre o tema (Rehman et al., 2019).

Pelo que podemos verificar contextos em que *big data* é definida pelos V's já mencionados, sendo adicionando ainda, 4 novos V's: a Veracidade dos dados, ou seja, a

certeza da não existência de dados incorretos que coloquem em causa os objetivos da sua análise; a Variabilidade dos dados, ou seja, dados onde sejam possível verificar diferenças, a Visibilidade, onde os próprios dados devem ser interpretáveis para poderem ser de fácil compreensão e, por fim, que neles exista Valor, seja nos próprios dados ou na interpretação dos mesmos (Mark van Rijmenam, 2014) .

Wu, Guo, Li & Zeng (2016), argumentam ainda que a definição de *big data* se estende ainda além dos seus V's. Sendo esta composta por três camadas distintas, a camada de infraestrutura que compreende as componentes de hardware necessários à formulação, aquisição e recolha dos dados, que incluem os sistemas externos de armazenamento, servidores e *data centers*, a camada de organização, gestão e análise que compreende todo o software necessário para a preparação dos dados recolhidos, de modo a que seja possível extrair valor dos mesmos, o processamento e análise dos dados estruturados e não estruturados, podendo a análise ser executada online ou offline, em tempo real ou não, e, por fim, a camada de serviços que compreende os diversos serviços como as interfaces externas, serviços de integração, consultoria e de segurança, tal como se apresenta na figura 2.1.



Figura 2.1. As três camadas de Big Data

Fonte: J. Wu et al. (2016, p.2)

Big data analytics é conhecida como a ciência de organização de grandes volumes de dados, análise desses dados para descoberta de padrões existentes nos mesmos e extração, preparação e apresentação de informações que sejam importantes para a tomada de

decisões a realizar pela pessoa ou entidade que necessita de investigar os dados recolhidos. O processo de análise de *big data* é normalmente efetuado através do uso de técnicas para separação da informação revelante para o tema em estudo (Sun & Huo, 2021).

O processo de recolha e análise de grandes volumes de dados é motivado por vários fatores e traz consigo muitos benefícios. Nas áreas comerciais e empresariais, o papel de *big data analytics* é reconhecido devido à vantagem do seu uso no que toca ao aumento da eficiência dos negócios, através do seu papel no processo de tomada de decisão. Na área da saúde, a *big data analytics* permite obter resultados a nível das intervenções prognósticas, terapêuticas tendo impacto no estilo de vida e comportamento dos utentes, ao mesmo tempo tornando o processo mais eficiente a nível monetário e garantindo sustentabilidade da infraestrutura. Ao nível da interatividade com clientes, *big data analytics* tem impacto sobre a experiência do cliente; levando os negócios a adaptarem o seu ambiente operacional às expectativas do cliente e proporcionando uma visão e perspetiva diferentes ao empresário sobre as necessidades do seu cliente. Em sistemas mais complexos que exigem moldagem e controlo, *big data analytics* permite uma evolução mais controlada (Kambatla et al., 2014).

Segundo Syed et al. (2020), no que toca aos tipos de dados recolhidos, o processo de *big data analysis* lida com diferentes estruturas de dados: *structured data*, ou seja, dados cujo tipo de dado, estrutura, formato e padrões se encontram claramente definidos. Como exemplos na área de energia podem apontar-se os dados provenientes dos parâmetros definidos nos contadores que permitem a leitura do consumo de energia. *Semi-structured data*, ou seja, dados semiestruturados com padrões perceptíveis em formato textual, como por exemplo dados referentes a qualidade da energia e a potência utilizada no consumo e *quasi-structured data*, dados em formatos irregulares que podem ser transformados e explorados com tempo e as ferramentas corretas. A maior diferença entre *semi-structured data* e *quasi-structured data* é o facto de o primeiro poder ser mais facilmente estruturado e formatado enquanto o segundo depende de um processo muito mais cuidado para obtenção de conhecimento contido neles, alguns exemplos deste tipo de dados são pesquisas em motores de busca ou *web-scraping*. Por fim, existe *unstructured data* que

se refere a dados que não têm modelos ou esquemas pré-definidos como os dados meteorológicos.

Segundo Hariri, Fredericks & Bowers (2019), várias técnicas avançadas de exploração destes dados são utilizadas durante o processo de tratamento dos mesmos. Técnicas de *DM*, *machine learning*, *natural language processing* e *computational intelligence* são implementadas juntamente com estratégias de parametrização, *divide-and-conquer*, *incremental learning* e *sampling* de modo a converter os grandes problemas em problemas mais pequenos permitindo que estes sejam analisados com clareza levando a uma formulação mais simples e eficiente de decisões, redução de custos e permitindo o processamento mais rápido.

2.2 Internet of Things e a recolha de big data

Os avanços tecnológicos verificados nas últimas décadas levaram ao desenvolvimento de ligações entre as diferentes tecnologias por nós utilizadas. O número de dispositivos conectados à Internet leva a um crescimento do volume de dados recolhidos de várias fontes, por sua vez criando *big data*. Atualmente, a IoT é apontada como sendo o principal condutor da revolução tecnológica que vivemos, devido ao seu contributo no que toca à recolha de dados para diferentes indústrias (Lv & Singh, 2021).

Esta ideia é apoiada por Sestino, Prete, Piper & Guido (2020), que consideram que a proliferação e disseminação de diferentes objetos, equipados de sensores com acesso a redes *wireless* e inovação de capacidades computacionais deram lugar a grandes oportunidades de desenvolvimento para diferentes negócios quanto a integrarem novos mercados e obterem *insights* dos seus clientes, melhorando assim o seu planeamento e implementação, concluindo que *big data* e a sua consequente análise representam uma revolução nos processos de tomada de decisões a diferentes níveis. De acordo com os autores, estas tecnologias podem ainda afetar a competitividade das empresas através da transformação dos seus produtos e serviços em oportunidades de negócio digitais.

Segundo Ahmed et al., (2017), os dados gerados pelos dispositivos associados a IoT podem ser utilizados para encontrar potenciais tendências, assim como, investigar o impacto de certos eventos ou decisões. Ao nível de recolha de dados, os dispositivos IoT

tornam-se uma fonte relevante de recolha de dados, com grandes volumes associados, variedade e velocidade o que torna esses dispositivos interessantes para a investigação de *big data* (Ge et al., 2018).

2.3 Armazenamento e processamento de *big data*

O armazenamento e o tratamento de dados são dois dos elementos na preparação de dados para obtenção de conhecimentos capazes de influenciar positivamente a tomada de decisões. A grande quantidade de dados e características específicas dos dados recolhidos leva à necessidade de encontrar respostas aos desafios variados no que toca ao tratamento dos dados.

Um desses desafios passa pelo facto de a escala de dados a monitorizar ser incrivelmente elevada, aumentada pela inclusão de diversos tipos de dados desde *structured data* a *non-structured data*, tornando assim necessário o uso de soluções aptas a tratar dados com estas as características como o Apache Hadoop que permite o armazenamento distribuído de dados. Um segundo ponto importante a referir passa pelo facto de a velocidade de processamento dos dados necessitar de ser rápida para agilizar as diferentes fases do processamento de dados, desde a limpeza dos dados não relevantes, criação de novos campos, à conversão dos diferentes formatos de dados de fontes distintas. A extração e o reconhecimento de padrões dos dados são outros aspetos a ter em conta visto que podem causar problemas no processo de análise dos dados podendo sentir-se efeitos sobre a tomada de decisão caso estes processos não sejam realizados a uma velocidade eficiente. (Li et al., 2019).

Segundo Jiang, K. Wang, Y. Wang, Gao & Zhang (2016), o processo de armazenamento de dados envolve a fusão de dados, a integração dos mesmos, a sua gestão e transformação, sendo este processo normalmente denominado de *Extract Transform, and Load* (ETL). Em geral, os dados referentes à energia quando recolhida necessita de escalabilidade, auto-organização, algoritmos de encaminhamento e disseminação. Como a informação recolhida contém o perfil do consumo e dados privados, torna-se necessário garantir a confidencialidade e segurança dos mesmos, pelo que são normalmente implementadas tecnologias de cifração e decifração de dados para o armazenamento e acesso aos mesmos. O processo de ETL é normalmente apoiado por diferentes

ferramentas para facilitar as três fases que o compreendem: a extração, transformação e carregamento de dados nas *data warehouses* (Bansal, 2014).

Ao nível do armazenamento dos diferentes tipos de dados, segundo Siddiqa, Karim & Gani (2017), a crescente tendência de mudança para inclusão de *non-structured data* tem tornado as bases de dados relacionais tradicionais obsoletas para armazenamento. Esta inadequação tem motivado o desenvolvimento de novos mecanismos de armazenamento de *big data*, mais escaláveis, eficientes e de confiança capazes de colmatar o aumento do volume de dados a processar e melhorar a facilidade de acesso aos mesmos. Como exemplos mais conhecidos de sistemas utilizados temos o Google File System, o Hadoop Distributed File System (HDFS), BigTable da Google e, por fim, o MongoDB.

Uma forma de colmatar o impacto do grande volume de dados e tornar o processo de análise mais rápido passa por, durante o processo de recolha de dados, garantir a integração e limpeza dos dados para assim conseguir evitar a redundância de dados e comprimir os mesmos: a este processo dá-se o nome de *data preprocessing*, sendo esta a primeira fase de processamento dos dados (Ghorbanian et al., 2019).

Os valores de dados armazenados nos dias de hoje podem facilmente ultrapassar os *terabytes* ou chegar mesmo a atingir os *petabytes* (1024 *terabytes*) ou *exabytes* (1024 *petabytes*) (S. Wu et al., 2011). Sendo atualmente previsto que o volume de dados criados, recolhidos, copiados e consumidos em todo o mundo para 2025 seja de 181 *zettabytes* (1 *zettabyte* = 1024 *exabytes*) (Statista, 2022). Dada a imensa quantidade de dados estes não podem ser todos processados durante a fase de pré-processamento, pelo que vários autores destacam a importância do uso de tecnologias mais eficientes para o processamento dos dados recolhidos. No que toca a ferramentas de processamento, segundo os autores Makrani, Tabatabaei, Rafatirad & Homayoun (2018), o Hadoop MapReduce é considerada a mais popular *framework* para *big data* visto suportar escalabilidade no armazenamento e recursos de processamento de *big data*.

2.4 Visualização de *big data*

Ao nível da visualização de *big data*, esta prova ser mais complexa que a visualização de dados tradicionais devido à complexidade acrescida dos V's de *big data* (Yaqoob et al., 2016).

Kumar & Goyal (2016), argumentam que, dada a complexidade aumentada, a geração de relatórios prova ser insuficiente para a interpretação dos dados recolhidos, importantes para a tomada de decisão, no entanto admitem que, com a criação de materiais em formato gráfico, torna-se possível facilitar esse processo. Segundo os autores, a representação dos dados é importante em situações em que a quantidade de informação a ser considerada é demasiado grande em tamanho e muito difícil de ser utilizada no processo de tomada de decisão, pelo que ferramentas de visualização de dados são mais uteis quando associadas a *big data analytics*, devido ao volume dos dados e ao facto de, por si só, os relatórios serem menos atrativos. Segundo os autores, a associação dos mesmos com um componente visual permite a apresentação dos resultados de uma forma mais amigável fomentando as capacidades de tomada de decisões dos intervenientes.

Segundo Chou & Truong (2019), interesse na área de gestão inteligente de consumo de energia através de *DM* e *time-series analysis* têm aumentado. Os autores apresentam no seu artigo um eficiente sistema de informação para gestão de energia que monitoriza o consumo de energia das habitações, este sistema interage com os seus utilizadores finais através de *dashboards* e *emails*. A visualização desenvolvida pelos autores permite aos utilizadores visualizar em tempo real o seu consumo e determinar se o consumo de energia está em linha com o esperado ou acima dos valores estabelecidos como normais. A figura 2.2 mostra a versão *web* do *dashboard* desenvolvido pelos autores para visualização do consumo de energia.

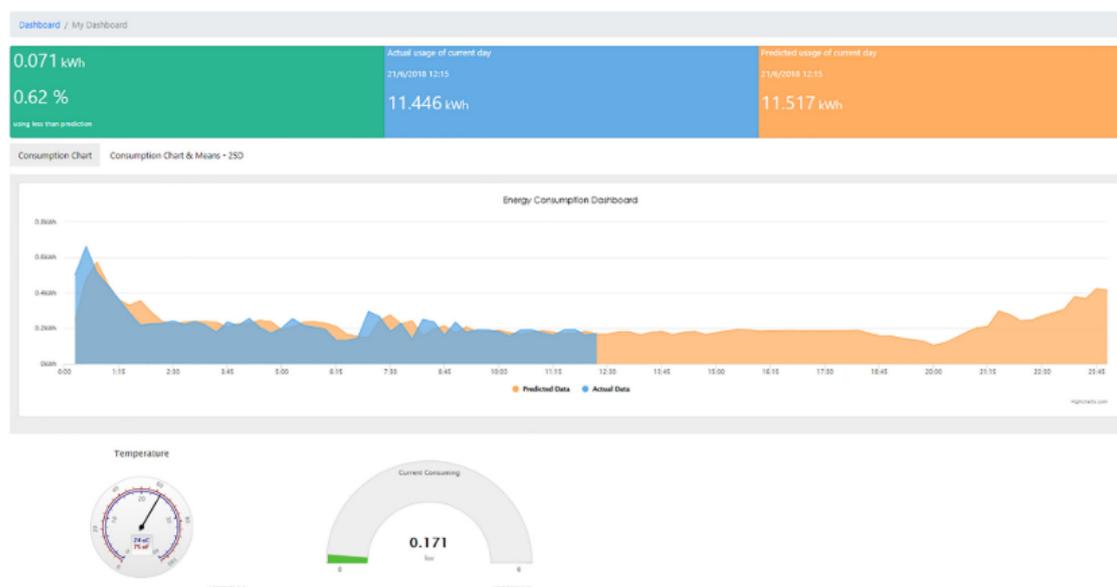


Figura 2.2. Dashboard de visualização de consumo energético

Fonte: Chou & Truong (2019, p. 8)

2.5 Big Data Analytics na energia

Os avanços tecnológicos e o processo de constante agregação de dados, já mencionados anteriormente, têm tido efeitos em variados setores. O setor energético é um dos setores que tem sofrido um movimento de transformação para tecnologias ditas “*smart*” em especial nas suas redes elétricas (El Khaouat & Benhlima, 2017). O motivo central que leva esta transformação e introdução de tecnologias de informação e comunicação (TIC) passa pelo facto de conseguir garantir através das mesmas uma rede mais segura, confiável, resiliente e flexível (mais próxima dos V’s associados a *big data*), equipada de sensores e outros dispositivos que garantam processos mais precisos de recolha, monitorização, controlo e partilha dos dados integrados na rede (associando-se assim à IoT) (Junaidi & Shaaban, 2018).

No que toca à recolha de dados de energia, segundo Motlagh, Mohammadrezaei, Hunt & Zakeri (2020), os sensores são os “atores” principais sendo utilizados para recolha e transmissão de dados em tempo real. Estes sensores são aplicados em diversas indústrias, inclusive, na energia. Na energia são, muitas vezes, utilizados em diferentes fases do processo desde a produção, transmissão, distribuição e consumo. Com estes é possível serem desenvolvidas estratégias de gestão energética inteligente, ao entregarem

informação em tempo real para a otimização energética e facilitar assim a gestão de energia utilizada.

Jaradat, Jarrah, Bousselham, Jararweh & Al-Alyyoub (2015) concluem que as tecnologias inteligentes e a sua integração no setor energético se traduz num custo acrescido a nível de armazenamento e processamento de grandes volumes de dados. Ao nível da informação, referem-se às necessidades de potência energética reportadas pelos utilizadores, problemas existentes nas linhas de energia, o estado atual dos componentes integrados na rede, os prazos de recolha de consumos energéticos, as condições meteorológicas, os dados recolhidos pelos contadores, o histórico de falhas energéticas, entre outros. Tendo consciência da magnitude de dados e a sua variedade, os autores argumentam que torna-se necessário às empresas do setor deterem uma infraestrutura capaz, para assim conseguir retirar o devido potencial das tecnologias, onde se inclui o *software* para tratamento dos dados e o *hardware* capaz de armazenar e processar eficientemente o volume de dados recolhido.

Ao nível da energia, *big data analytics* tem um papel importante no desenvolvimento de *intelligent energy networks* (IENs). As IENs são redes energéticas inteligentes, normalmente definidas como sendo redes que permitem uma otimização energética com a troca bilateral de informação entre os produtores e consumidores através da integração das TICs nas redes energéticas convencionais. Estas redes tem tido um desenvolvimento rápido em anos recentes para enfrentar as necessidades de entrega de energia robusta, flexível, amiga do ambiente e de uma forma económica. O desenvolvimento origina um aumento de dados recolhidos pelas redes energéticas, que necessitam de técnicas de análises de dados mais robustas e eficazes (Ma et al., 2017).

Segundo Jaradat, Jarrah, Bousselham, Jararweh & Al-Alyyoub (2015), os contadores-inteligentes recolhem várias leituras ao longo do dia, o que obriga a gerir o desafio de quantidade massiva de dados gerada pelos contadores e outros equipamentos na rede, o qual é enfrentado pelas empresas de fornecimento energético através do uso de serviços de infraestruturas na *cloud*. As empresas que comercializam estes serviços incluem nos seus serviços, espaço para armazenamento de dados, gestão de recursos energéticos, assim como, centrais elétricas virtuais. A opção pelos serviços *cloud* para

armazenamento, pode desencadear nas empresas do setor energético, dificuldades a nível de segurança e latência no acesso à *cloud*, no entanto, esta tecnologia pode beneficiar o desenvolvimento dos sistemas *Supervisory Control and Data Acquisition (SCADA)*. O sistema SCADA é utilizado na energia para a monitorização e controlo da rede energética em tempo real ao recolher e disponibilizar os dados dos vários sensores distribuídos pela rede. A gestão da informação recolhida através do sistema SCADA auxilia no processo de controlo da rede e aumento de eficiência energética.

No que toca ao valor que a análise de dados tem para o setor energético, a capacidade de garantir uma produção e monitorização da energia é fundamental. Nos tempos que vivemos existe uma crescente necessidade de garantir respostas aos problemas que surgem no setor como controlo das potências ou até mesmo otimização de preços de comercialização de energia no mercado. Estes desafios dependem da necessidade de garantir um controlo apertado às redes, inclusive devido ao papel que estas têm na integração das energias renováveis. Desafios como os já apontados também exigem um maior controlo devido à volatilidade verificada nos tempos de hoje, fruto de fatores difíceis de controlar como as condições climatéricas, estas características justificam a necessidade de análise de dados que resulte em informação ao nível da produção, distribuição, gestão e consumo de energia, assim como, à informação meteorológica que pode afetar a energia desde a sua produção à sua distribuição (Corizzo et al., 2019).

A figura 2.3 permite observar um exemplo de um processo de *big data analytics* onde se verificam as diferentes fases e componentes do processo de análise e a sua relação.

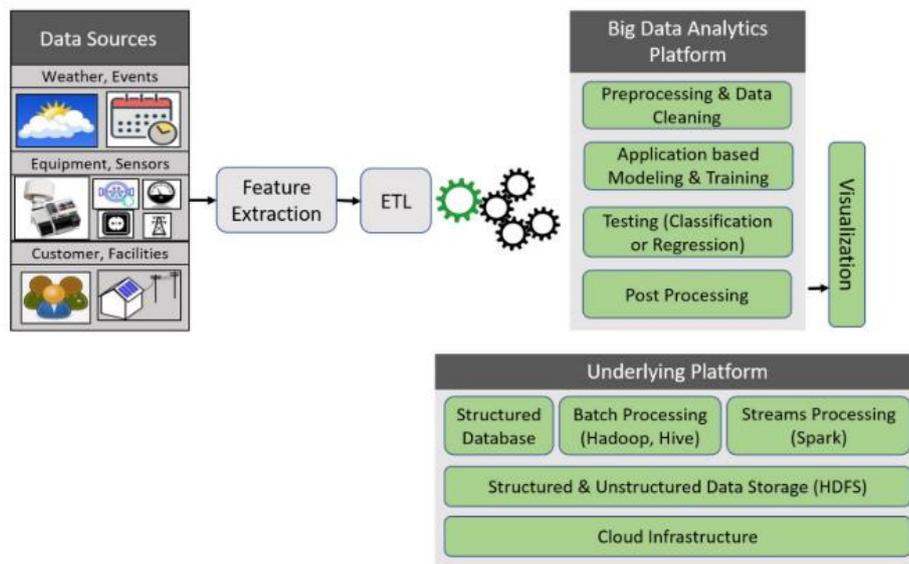


Figura 2.3. Processo de big data analytics

Fonte: Syed et al. (2021, p. 6)

De acordo com Diamantoulakis, Kapinas & Karagiannidis (2015), uma *smart-grid* é o sistema de geração e gestão energética do futuro capaz de colmatar as crescentes necessidades energéticas de uma forma sustentável para o ambiente e para a economia através da aplicação de avançadas tecnologias de informação e comunicação digitais.

A figura 2.4 permite observar os diferentes componentes participantes de uma *smart grid*.

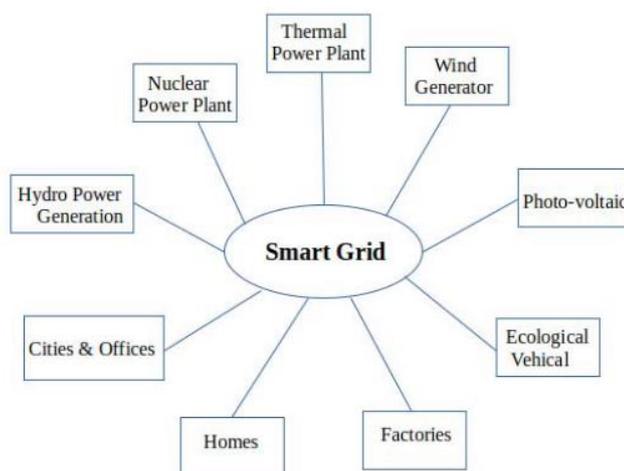


Figura 2.4. Os principais componentes de Smart Grid

Fonte: Shobol et al. (2019, p. 1)

Ge, Bangui & Buhnova (2018), descrevem as *smart grids* como sendo um sistema que oferece vários serviços valiosos ao garantir uma melhor integração da energia renovável nos sistemas energéticos, assim como, oferecer um maior controlo do consumo da energia aos operadores e consumidores com a ideia de otimizar o consumo energético pelas massas. Para tal, as *smart grids* cooperam com tecnologias de IoT na criação de serviços inteligentes associados às tecnologias de *big data* para, com as mesmas garantir, a construção de conhecimento que resulte em estratégias de aumento da eficiência na gestão de energia.

O contributo de Ku, Park & Choi (2017) serve como um exemplo de *big data analytics* na energia: os autores apresentam o conceito da *microgrid* como sendo a aplicação de uma *smart grid* de dimensão reduzida, integrando na mesma, as tecnologias da informação tendo em conta as características da região que a rede cobre. No seu exemplo, referem-se a uma plataforma de gestão energética por eles produzida de modo a maximizar a eficiência da *microgrid* onde para controlar e monitorizar os recursos distribuídos e o gasto energético realizam-se a *big data analytics*, para a produção e realização da otimização dos dados energéticos recolhidos e desenvolvimento de análises, relatórios e *dashboards*, para controlo da energia e descoberta de tendências do seu uso. Com a informação acerca das necessidades energéticas dos diferentes consumidores, planeiam uma estratégia eficiente de distribuição da energia.

2.5.1 Data Mining na energia

A necessidade de desenvolvimento de novas tecnologias, capazes de estar alinhadas com o aumento do volume de dados produzido e que sejam capazes de utilizar, de forma inteligente, a informação gerada a partir deles, tem-se tornado cada vez mais importante. Os grandes volumes de dados existentes em bases de dados levam a uma necessidade de desenvolvimento de técnicas que possibilitam novas formas de os trabalhar. DM é o processo de extrair informações, não triviais, desconhecidas e com potencial para serem úteis, para a realização de previsões e correlações entre diferentes parâmetros (Mukherjee et al., 2015).

O processo de *DM* é normalmente separado em *DM* descritivo e preditivo. *DM* descritivo refere-se à produção de informação e conhecimento não trivial enquanto *DM* preditivo

produz modelos e funções de sistemas. O *DM* descritivo compreende atividades de *clustering*, *summarization*, *dependecy* e *modeling*, enquanto que o *DM* preditivo inclui a *classification*, *regression*, *change* e *deviation detection* (Sun & Huo, 2021).

DM aplicado a *big data* de energia é, normalmente, utilizado para o apoio à realização de previsões de energia necessária e capaz de corresponder à oferta e procura existente em qualquer momento. No entanto, a crescente evolução tecnológica na área de motores elétricos, especialmente nos veículos elétricos, tem também levado à aplicação de técnicas como *deep learning*, *transfer learning*, *randomized learning*, *granular computing*, *multisource data fusion* de modo a levar à exploração do estado e carga necessária das baterias que compõe o carro elétrico, juntamente com os efeitos de uso crescente das mesmas na autonomia do carro e possíveis variações consoante o ambiente onde este é utilizado. O uso destas técnicas também é verificado nos processos de simulação e previsão dos preços do mercado energético (W. Liu et al., 2021).

O uso de técnicas de *DM* aplicadas a *big data* para questões de análise dos dados recolhidos via sensores não se verifica apenas por estes motivos. Um exemplo do uso das mesmas passa pela possibilidade de estimar, através de *DM*, o potencial de placas fotovoltaicas aplicadas em habitações. Os dados recolhidos e analisados contribuem com informação importante para o objetivo de descarbonização estabelecido no acordo de paris sobre as alterações climáticas (Walch et al., 2020).

Concluindo, os processos de *DM* detêm um papel fundamental após a sua introdução associada ao tratamento de dados de energia, em especial ao tratamento de dados de energia renovável. Os sistemas de energias renováveis são conhecidos por gerarem uma massiva quantidade de dados, provenientes de diferentes fontes como, biomassa, solar, eólica e hídrica, sendo estes dados recolhidos em diferentes períodos temporais, através de diferentes sensores e dispositivos. Tendo em conta estas características, juntamente com as restantes características de *big data*, verifica-se a importância da aplicação de técnicas como *decision trees*, *neural networks*, *k-nearest neighbor*, *apriori*, *fp growth* e *k-means clustering* capazes de fazerem frente aos desafios de volume, variedade, velocidade e veracidade existentes. O uso destas técnicas pode levar a uma utilização

mais correta da energia e a uma redução de custos, assim como, possibilita às empresas a possibilidade de controlar a produção e consumo energéticos (Fawzy et al., 2016).

2.5.2 Energy Internet

De acordo com, Wang, Li, Feng & Tian (2017) os dados que alimentam o processo de análise podem surgir em qualquer altura e qualquer lugar em toda a rede energética. Estes podem surgir através da atividade de monitorização energética realizada pelos fornecedores de energia, ser fruto do uso da energia nas mais variadas operações e ainda existirem sobre diferentes formas como dados climáticos, dados provenientes de transportes que funcionam através de energia elétrica ou dados de negócio. Os mesmos preveem, ainda, que as quantidades de dados continuem a expandir, apelidando a esta nova era dos dados energéticos como “*Energy Internet*” ou “*Internet of Energy*”.

Segundo Zhou, Yang & Shao (2016), a *Energy Internet* pode ser definida pelo desenvolvimento de novas formas de sistemas energéticos que integram o fluxo de energia, informação e negócio da mesma. Para este setor, as TIC emergentes continuam a entrar em todo o processo de produção, transmissão, distribuição e consumo de energia. Na perspetiva de negócio, a *Energy Internet* verifica-se em três tipos de fluxos distintos: 1) o fluxo de energia que compreende as transações da mesma nas suas diferentes formas 2) o fluxo de informação que flui na diferente infraestrutura construída para sua comunicação, gestão, produção e consumo e, por fim, 3) o fluxo do processo de negócio que compreende desde a produção de energia à sua aplicação.

Ao nível das tecnologias chave envolvidas na *Energy Internet* estas são similares às já existentes para a exploração de *big data* relacionada a outros temas. Alguns exemplos da tecnologia associada são os *energy routers* que têm como função aumentar a confiabilidade, eficiência e segurança do sistema energético, os sistemas de armazenamento, as fontes de energia renovável, a interface de *plug-and-play* existente e, por fim, a sua integração em diferentes contextos como nos veículos elétricos (Kafle et al., 2015).

A introdução da temática da *Energy Internet* tem originado diversos projetos de aplicação de *big data analytics* ao setor energético, a diferentes níveis, sejam estes desenvolvidos por negócios privados ou por entidades públicas.

Um exemplo verifica-se através do desenvolvimento de ações inovadoras nos programas europeus, nomeadamente no programa *Horizon 2020 Framework Programme*, no tópico DT-ICT-11-2019, cujo objetivo a cumprir passa pela inovação das redes energéticas existentes através da criação de sistemas heterogéneos interligados, a um aumento da propagação de diferentes equipamentos capazes de auxiliar nos processos de gestão de geração de energia e consumo da mesma. Neste programa, é lançado o desafio de desenvolvimento e aplicação de uma arquitetura capaz de gerir, processar e analisar em tempo real dados recolhidos pelos equipamentos, assim como, transformar esta num processo de análise de dados para garantir melhor regulação de operações na rede e inovações nos serviços energéticos (European Commission, 2018).

Ao nível privado começa a verificar-se uma exploração da *Energy Internet*, especialmente através do uso das *smart grids*, com o desenvolvimento de sistemas baseados puramente na web onde se verifica a introdução de diferentes componentes garantindo interações *machine-to-machine* e *human-to-machine* (Bui et al., 2012).

2.5.3 Desafios e oportunidades de *big data analytics* no setor

Big data analytics tem o potencial de criar várias oportunidades para as redes de energia, em especial que aumentem os ganhos das organizações a nível técnico, social e económico. No entanto, também traz consigo uma série de desafios especialmente para uma área como o setor energético, considerada complexa quer devido ao número de intervenientes quer quanto às diferentes tecnologias utilizadas (Ghorbanian et al., 2019).

Alguns dos desafios enfrentados já mencionados são o facto de se prever que o volume dos dados gerados venha a aumentar para valores muito mais elevados o que implica um desafio quanto ao armazenamento, *DM*, processamento e indexação dos dados recolhida (Dang-Ha et al., 2015).

A necessidade de decisões em tempo real é também um desafio existente devido à quantidade de dados a processar juntamente com a quantidade de dados históricos. Essa

quantidade e variedade de dados, somada à necessidade de respostas em tempo real, poderá criar dificuldades no desenvolvimento de algoritmos que conseguiram fornecer uma representação em tempo real do conhecimento obtido através desses dados (Hariri et al., 2019).

A estes desafios soma-se ainda o associado à incerteza dos dados gerados, visto a existência de possíveis falhas na compreensão dos dados, seja por estes serem incompletos, seja por falta de conhecimento acerca do processo operacional. Existem, também, algumas preocupações quanto à segurança dos dados gerados visto que os dados recolhidos pela tecnologia implementada envolvem dados dos consumidores privados pelo que são levantadas algumas questões ao nível da privacidade, integridade e autenticação (Bhattarai et al., 2019). A estas preocupações acrescentam-se ainda a possibilidade de alguém intercetar e obter as informações de consumo energético de diferentes habitações, a manipulação dos dados e corrupção dos mesmos. Dado os equipamentos poderem ser acedidos e intervencionados remotamente, podem sofrer também ataques à distância ou serem comprometidos por software desenvolvido para ataque aos mesmos. As redes baseadas em IoT em grande escala encontram-se vulneráveis pelo que também podem sofrer ataques de *denial of service*, ou seja, o desenvolvimento de medidas de defesa é essencial para garantir o pleno uso destas redes (Chin et al., 2017).

No que respeita a oportunidades criadas com a introdução da *big data* no setor, verifica-se através das aplicações de *big data* a oportunidade de participação conjunta dos pequenos consumidores, produtores de energia, operadores do sistema de distribuição e restantes participantes nos sistemas de informação. Através da inclusão de grandes bases de dados necessárias para processar os dados recolhidos, criam-se também oportunidades para melhor controlo, monitorização e gestão das redes elétricas (Bhattarai et al., 2019).

Outra oportunidade desenvolvida pela associação destas tecnologias às empresas do setor passa pelo facto de, através da análise preditiva, ser possível ser mais eficiente em relação à gestão das cargas de energia como no caso das energias renováveis, conseguindo, assim, enviar sinais aos seus clientes e adaptar as cargas às possíveis irregularidades energéticas que podem afetar as condições da rede. Para que tal seja possível é necessária a

constituição de um sistema eficiente em torno da tecnologia, permitindo, assim, uma resposta em tempo real. O uso de tecnologias mais inteligentes, necessárias para a abertura do processo de *big data analytics*, irá também facilitar o estudo, modelização e otimização dos dados criando mais valor às organizações (Schuelke-Leech et al., 2015).

A criação e uso da integração dos processos inteligentes necessários para a geração e análise de dados de *big data*, que fomentam os sistemas de informação trazem consigo uma oportunidade de melhoria dos sistemas de deteção e correção de anomalias, os quais podem ser intervencionados através das plataformas de gestão das organizações. O desenvolvimento de um sistema de cibersegurança capaz de detetar estas anomalias tem-se tornado fundamental no processo de garantia de privacidade e segurança dos dados energéticos. Nos últimos tempos têm sido desenvolvidos algoritmos capazes de responder a estas necessidades de proteção originadas com a integração de *big data* na energia. Os algoritmos mais conhecidos são categorizados em dois grupos: 1) algoritmos generalistas orientados para a descoberta de anomalias em *big data* de energia, sendo estes aplicados normalmente para evitar roubo de energia, falhas na mesma ou entradas no sistema por fontes externas; e 2) algoritmos para deteção de anomalias baseadas no furto de energia e esquemas de intrusão física para combater ataques que tentam introduzir dados falsos no sistema (Hu & Vasilakos, 2016).

3 ENQUADRAMENTO METODOLÓGICO

O seguinte capítulo pretende expor o percurso percorrido durante a realização da revisão de literatura. Neste capítulo encontra-se referido a tipologia do estudo, o protocolo aplicado para a sua correta realização, a estratégia de pesquisa, a base de dados considerada para exploração, o processo de seleção dos artigos e por fim o padrão utilizado para a recolha de dados.

3.1 Tipologia do estudo

No sentido de dar resposta à questão colocada e tendo em atenção a ideia de concentrar as aplicações de *big data* em energia, optou-se pela realização de uma RSL tendo em conta o objetivo principal de recolher diferentes aplicações de *big data analytics*, demonstrando, assim, como tem interagido com este setor.

Uma RSL difere de uma revisão de literatura tradicional já que, ao contrário da revisão sistemática, a tradicional foca-se no conhecimento e experiências dos autores e não numa apresentação exaustiva do tópico em discussão e, adicionalmente, a tradicional segue uma metodologia difícil de replicar enquanto a metodologia de RSL assenta num protocolo executável (Aromataris & Pearson, 2014).

Dado o objetivo de realização de pesquisa de informação com valor para o tema em discussão, a realização da recolha e consequente análise da informação foi efetuada tendo numa perspetiva de obter um número de dados, com grande veracidade, cujo conhecimento fosse apresentado de forma a ser rapidamente absorvido pelas partes que o consultem. Ao mesmo tempo espera-se que sirva como um contributo à exploração do uso de tecnologia e técnicas de *big data analytics* aplicadas à energia e seja valorizado por partes já conhecedoras do tema que consigam utilizar a informação descrita no mesmo e novos leitores interessados na área que com este conseguem ser introduzidos aos conceitos chave de *big data analytics* e a sua interação com a área energética.

Para o controlo e realização desta RSL foi adotada uma adaptação da *checklist* PRISMA de 2020. A *checklist* PRISMA compreende um conjunto de itens para a realização de revisão sistemática e meta-análises de apoio à realização das RSLs. Nos anexos 1 e 2 encontram-se a *checklista* PRISMA utilizada durante esta RSL.

3.2 O protocolo aplicado

Para a realização de qualquer revisão sistemática da literatura recomenda-se a criação de um protocolo que descreva o método planeado para a execução da revisão (Siddaway et al., 2018).

Estes protocolos são normalmente compostos por uma questão-chave, questões específicas para apoio na resposta à questão-chave, a estratégia a utilizar para a realização da revisão, os critérios de seleção de bases de dados a considerar, os métodos utilizados para seleção de publicações que acrescentem valor à revisão, os critérios de inclusão e exclusão a considerar, os processos de filtração, avaliação e escolha de artigos e por fim a recolha de dados e apresentação de resultados (Siddaway et al., 2018).

Tendo em consideração o processo de construção do protocolo a figura 3.1 pretende delimitar as diferentes fases do protocolo de realização desta RSL.

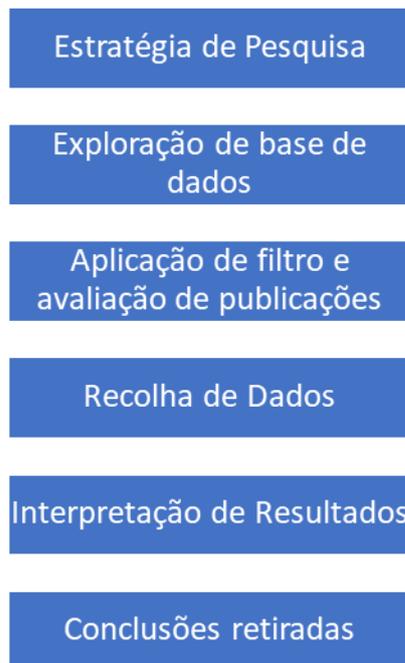


Figura 3.1. Protocolo utilizado

3.2.1 Estratégia de pesquisa

A estratégia de pesquisa e consideração dos artigos a utilizar para a realização da RSL foi idealizada tendo em atenção o objetivo de garantir uma compreensão do tema aos interessados, quer tenham ou não conhecimento prévio sobre o tema.

Tendo em conta esse objetivo, torna-se fundamental o desenvolvimento de uma estratégia de idealização, pesquisa e apresentação de resultados cujo processo de realização seja compreensível desde o ponto inicial de realização da RSL. Pelo que um entendimento da estratégia implementada é fundamental para a compreensão dos resultados obtidos.

Para tal, a RSL inicia-se com o estabelecimento dos critérios utilizados sendo este o ponto inicial para a realização duma pesquisa à literatura existente.

Para o processo de pesquisa e filtro de publicações a considerar foi utilizada uma variedade de palavras-chave relacionadas com o tema, escolha essa efetuada através de uma breve pesquisa geral da literatura (pesquisa essa cujos efeitos se repercutem no desenvolvimento do enquadramento teórico já apresentado no capítulo 2 desta dissertação), no sentido de obter uma primeira visão sobre o impacto entre os dois temas explorados e assim descobrir palavras-chave alinhadas com o objetivo de aplicar estas palavras e, com as mesmas, filtrar o número de publicações existentes na base de dados selecionada, simplificando, assim, o processo de avaliação da situação atual no que toca à literatura que explora o confronto entre os dois temas: energia e *big data analytics*.

Em relação à seleção dos pressupostos a ter em conta durante a pesquisa dos artigos presentes nesta revisão, opta-se por fazer uma escolha no que toca aos critérios a considerar, sejam estes para a inclusão ou exclusão de artigos. Quanto mais específicos forem estes critérios, mais limitado se torna o processo de descoberta de heterogeneidade entre as publicações existentes, pelo que, inicialmente, foram considerados critérios relativamente abrangentes, sendo posteriormente feita uma avaliação que permita a visão mais centralizada no valor das publicações para a RSL (Harris et al., 2014).

Os seguintes critérios de inclusão e exclusão de artigos a considerar para a realização da RSL podem ser observados na tabela 3.1.

Tabela 3.1. Critérios de Inclusão e Exclusão para a Revisão Sistemática da Literatura

| Critérios de Inclusão e Exclusão para a Revisão Sistemática da Literatura | |
|--|---|
| Critérios de Inclusão | |
| Período de tempo | Os artigos recolhidos devem ser pertencentes ao período de tempo de 2016 a 2022 de modo a garantir que são recolhidas e apresentadas as informações mais recentes no que toca aos desenvolvimentos existentes na matéria de <i>big data analytics</i> aplicado à energia. |
| Língua utilizada | Todos os artigos que fazem parte desta revisão estão escritos em inglês, devido a este ser o idioma mais utilizado na partilha de informação científica, sendo, assim, descartada qualquer pesquisa, análise ou recolha de dados de artigos produzidos com outro idioma que não este. |
| Temas anexos | Exploração de temáticas que influenciem ou sejam influenciadas, direta ou indiretamente, por <i>big data analytics</i> na energia. |
| Critérios de Exclusão | |
| Artigos incompletos ou Amostras | Não foram utilizados quaisquer artigos incompletos ou amostras de artigos visto que, para a correta seleção de artigos a utilizar e avaliação dos dados apresentados pelos mesmos, é necessário proceder-se à análise da totalidade da publicação e não apenas de passagens da mesma. |
| Textos de opinião | A informação apresentada sobre a forma de textos de opinião foi também desconsiderada devido à dificuldade existente de avaliar o valor dos mesmos, assim como, os conteúdos destes e como são apresentados. Estes artigos não foram sujeitos a <i>blind review</i> . |
| Fuga ao tema | Em relação aos artigos a utilizar não foram considerados aqueles que não se focuem em <i>big data</i> , energia ou que não sejam importantes para a contextualização destes dois temas. Neste caso, incluem-se artigos que compreendam matérias com realce no foro energético, mas que não constituem obras científicas que exploram estes dois temas ou o seu potencial, mas sim focados em temas diferentes destes. |

3.2.2 Base de dados utilizada

Ao nível da base de dados a utilizar para a realização da revisão, foi considerada apenas a base de dados Science Direct. Durante o processo de enquadramento teórico foram utilizados artigos de várias bases de dados como a IEEE Xplore, Taylor & Francis, Springer Link e a Science Direct. Durante esta exploração, foi possível obter uma visão sobre a possível abordagem às matérias que fazem parte desta dissertação.

A Science Direct é uma base de dados conhecida mundialmente, sendo as publicações existentes na mesma sujeitas a critérios de revisão bastante rígidos. O processo de avaliação de publicações a incluir na Science Direct é moroso de modo a garantir que esta base de dados inclua apenas artigos com valor científico elevado. A Science Direct é também uma das maiores bibliotecas no que toca ao número de artigos entre as bases de dados exploradas durante o enquadramento, detendo mais de 18 milhões de artigos distribuídos em 2.650 *peer-reviewed journals*, assim como, uma variedade de temas explorados de áreas como ciências, tecnologia, artes e humanidades. É ainda uma base de dados utilizada por uma variedade de utilizadores distintos, servindo desde instituições académicas, organizações governamentais e de investigação e desenvolvimento em variadas indústrias.

3.2.3 Processo de seleção

Numa tentativa de exploração inicial à base de dados, verifica-se que, uma simplificação da pesquisa através de uma simples procura por “*Big Data Analytics Energy*” prova ser insuficiente visto que, na ScienceDirect, o uso destes termos devolve mais de cem mil resultados. Nesta situação, torna necessário optar por um método de pesquisa que seja mais profundo ao nível das palavras-chave que assim permita filtrar o elevado número de artigos existentes que relacionam as duas áreas.

Para tal foram utilizadas para o processo de filtro de artigos existentes nesta base de dados algumas palavras-chave para descoberta daqueles que mais se aproximam do tema, numa tentativa de encontrar os artigos que mais valor produzem para a construção da resposta à questão principal que origina esta RSL.

Tendo esta informação em consideração, assim como, os temas já mencionados no enquadramento teórico, foram consideradas palavras-chave como “*Smart Grid*”, “*Smart Meters*”, “*Smart Sensor Network*”, “*Meter Data Analytics*”, “*Energy Management*” e “*Energy Consumption*”, assim como, *Big Data*, *Analytics* e *Energy* para garantir uma pesquisa por artigos com estes termos.

Com a seleção das várias palavras-chave a utilizar no processo de filtro, torna-se possível a construção da frase booleana. Esta frase consiste na junção de todos os termos através do uso de operadores booleanos, resultando no foco em artigos centralizados nas palavras utilizadas durante a exploração (Bramer et al., 2018).

A nível de operadores a utilizar, foram apenas considerados o operador AND e o operador OR, como forma de traduzir a combinação dos dois temas *big data* e energia.

Tendo esta informação em consideração constitui-se, na figura 3.2, o exemplo de frase booleana criado através da ligação das palavras-chave e operadores selecionados:

“Big Data” AND "Analytics" AND "Energy" AND ("Smart Grid" OR “Smart Meters” OR “Smart Sensor Network” OR “Meter Data Analytics” OR “Energy Management” OR “Energy Consumption”)

Figura 3.2. Frase Booleana considerada para a exploração

A frase booleana constituída foi então aplicada na base de dados já mencionada, para assim garantir um filtro inicial na exploração da mesma e reduzir o número de artigos encontrados, para que a pesquisa se centralize nos artigos que contêm estas palavras no título, resumo ou sejam palavras-chave específicas utilizadas.

A ligação entre a frase booleana constituída com os filtros existentes na base de dados é essencial para a produção de uma pesquisa aprofundada, pelo que o uso destes filtros não pode ser desconsiderado.

Com a definição da estratégia de filtro/consulta, foi realizada em novembro de 2021 a pesquisa à base de dados selecionada tendo a mesma conduzido aos resultados no que respeita ao número de artigos obtidos que cumprem os filtros mencionados até agora.

A aplicação da frase juntamente com o filtro para os mesmos conterem apenas artigos de 2016 a 2022, permitiu obter um total de 73 artigos distribuídos por 25 jornais, e relacionados com 10 áreas distintas.

3.2.3.1 Exploração bibliométrica

No sentido de completar a exploração realizada à base de dados, foi realizada uma revisão bibliométrica aos resultados obtidos com a pesquisa e filtro realizados na base de dados ScienceDirect.

A análise bibliométrica é uma análise popular aplicada com o sentido de permitir uma exploração de grandes volumes de dados científicos. A mesma permite observar os dados sobre uma visão evolutiva e, ao mesmo tempo, expor novas áreas emergentes relacionadas ao tema em estudo e padrões de colaboração entre os autores (Donthu et al., 2021).

Para a realização desta dissertação, optou-se pelo uso da ferramenta VOSViewer para a realização da exploração dos dados. Através desta ferramenta apresentam-se os resultados obtidos com a exploração ao nível da coocorrência de palavras-chave entre os autores das publicações científicas.

Através da ferramenta VOSViewer podemos observar a seguinte visualização, disponível na figura 3.3, correspondente à rede de palavras-chave utilizadas entre as publicações.

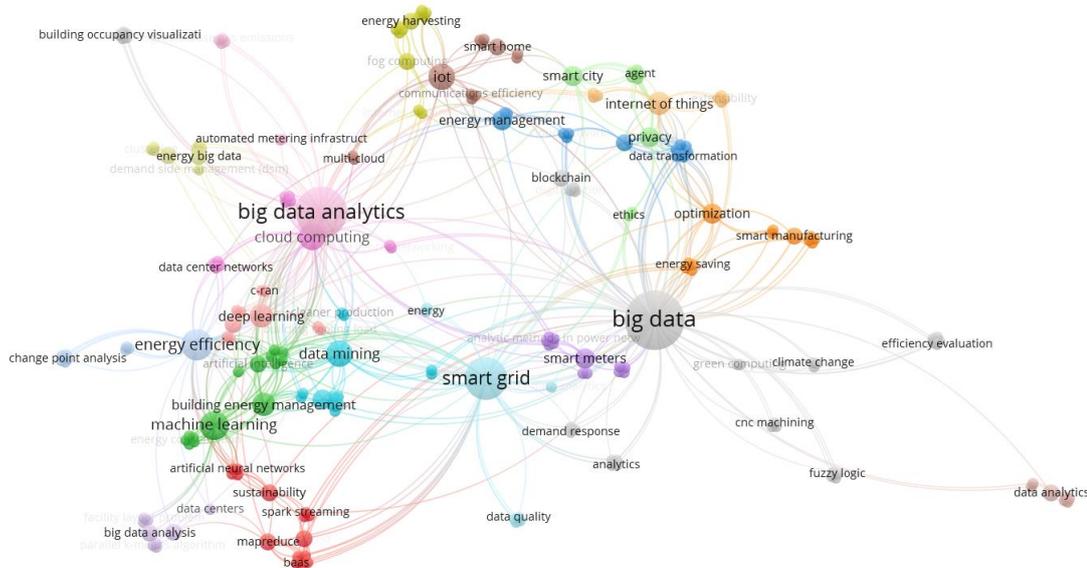


Figura 3.3. Network Visualization do VOSViewer

A figura 3.3 permite-nos observar as diferentes relações entre os diferentes artigos sendo as palavras-chave destes artigos a unidade de análise para formulação da visualização, sendo assim possível observar os diferentes *clusters* de palavras-chave formados pelos 73 artigos. Ao nível dos resultados podemos observar um total de 224 palavras-chave relacionadas entre si, num total de 274 palavras-chave existentes nos artigos, estando estas divididas por 26 clusters distintos (visíveis no apêndice 1). Como expectável, a palavra-chave *big data* é a coocorrência mais comum contendo 25 ocorrências, seguida de *big data analytics* com 19 e *smart grid* com 13.

Ao mesmo tempo, o uso do VOSViewer permite-nos fazer uma análise temporal do uso destas palavras-chave através da sua *Overlay Visualization*.

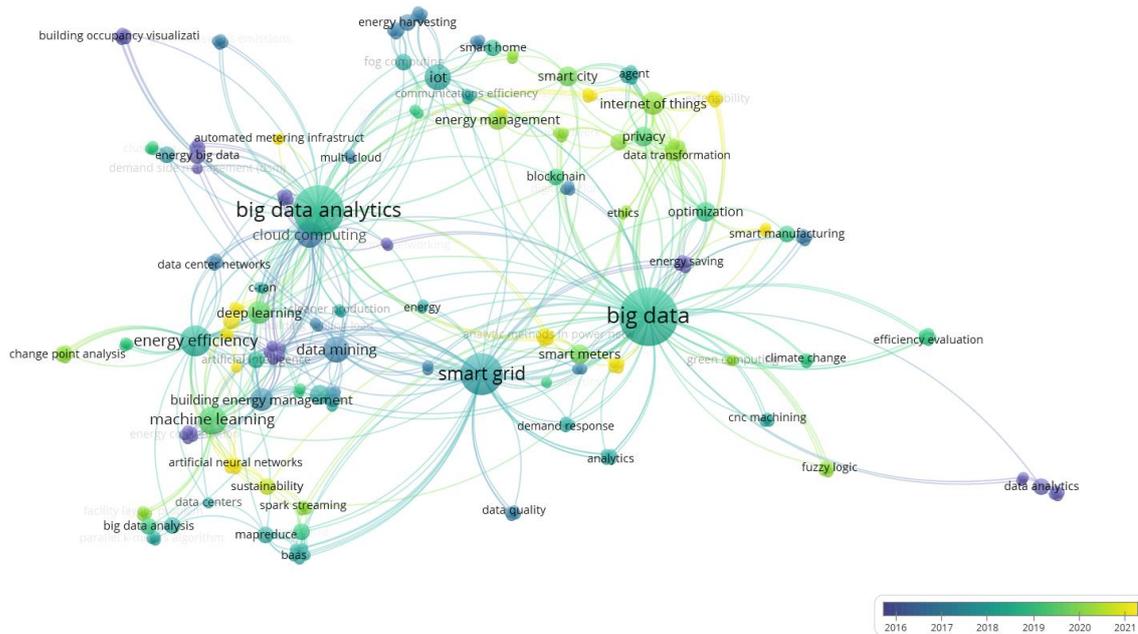


Figura 3.4. Análise temporal do uso das palavras-chave

Esta análise permite-nos ver como, ao longo do período entre 2016 e 2021, se tem verificado uma alteração ao nível das palavras-chave utilizadas no desenvolvimento de publicações que estudam estes temas. Palavras-chave como *data analytics* deram lugar a palavras como *big data analytics* e *energy saving* deu lugar a palavras como *energy efficiency*, pela legenda é possível observar o surgimento de novas palavras-chave como *energy transition*, muito devido à popularidade da transição para energias mais limpas verificada atualmente, ou *security and privacy* e *privacy-perserving data analysis* pela crescente preocupação com a segurança e privacidade dos dados recolhidos e trabalhados.

3.2.3.2 Filtro de resultados

Segundo a metodologia PRISMA, torna-se necessário, após a realização da exploração e identificação dos estudos disponíveis, realizar uma avaliação desses resultados com o objetivo de reduzir o número de estudos a analisar garantindo assim que os estudos que se encontram duplicados ou que simplesmente se revelam como não sendo relevantes para a matéria em discussão não fazem parte da RSL (Page et al., 2021).

Para a realização deste filtro, começou-se por enfrentar cada resultado com os critérios de inclusão e exclusão selecionados. Esta ponderação levou à remoção de um total de 8 dos 73 artigos iniciais, devido a estes 8 serem resultados que não cumprem os critérios

definidos por serem amostras ou artigos incompletos. Sendo assim são removidos 6 capítulos de livros, 1 editorial e 1 publicação de software resultando num total de 65 artigos.

Na fase seguinte do processo de seleção considerou-se o título das publicações e o seu resumo, caso se verificasse um afastamento do tema não prosseguiriam para a fase seguinte. A exploração desta fase de triagem resulta num corte de 34 artigos dos 65 acima mencionados permitindo uma focalização nos 31 artigos que argumentam as matérias em desenvolvimento nesta revisão.

No sentido de conseguir retirar conclusões acerca dos 31 artigos, procedeu-se à leitura dos mesmos na sua totalidade. A realização da leitura permite-nos realizar um teste de relevância aos 31 artigos, que consiste em durante a leitura dos artigos refletir e verificar se os mesmos respondem às questões do teste para assim selecionar entre os artigos potencialmente relevantes aqueles que verdadeiramente abordam as temáticas de *big data analytics* e energia e que refletem contributos à resposta da questão-chave colocada. Com a leitura realizada aos artigos obtêm-se as respostas às perguntas colocadas pelo teste de relevância, conseguindo assim retirar conclusões sobre os mesmos e sobre a sua inclusão nesta RSL.

As questões a considerar para o teste de relevância dos artigos podem ser visualizadas na tabela 3.2.

Tabela 3.2. Teste de Relevância para artigos encontrados

| Teste de Relevância para artigos encontrados |
|---|
| 1. “É realizada uma descrição do uso de <i>big data analytics</i> na energia ou a algum subtema relacionado com o setor energético ou que tem impacto sobre a energia?” |
| 2. “O artigo e a informação descrita relatam ao processo de análise, recolha, exploração ou descrição de <i>big data</i> e os seus efeitos na energia?” |
| 3. “O artigo e a informação apresentada acrescentam valor à pesquisa e contribuem de alguma forma para a resposta à questão-chave?” |

Através da descoberta dos artigos, os filtros realizados a essas publicações e o seu confronto com as questões do teste de relevância, procede-se à remoção de um total de

58 artigos garantindo assim, que a abordagem a esta matéria é apenas discutida com base em artigos relevantes para a mesma.

Um resumo deste processo de exploração e filtração dos resultados obtidos na base de dados pode ser visualizado através do diagrama presente na figura 3.5.

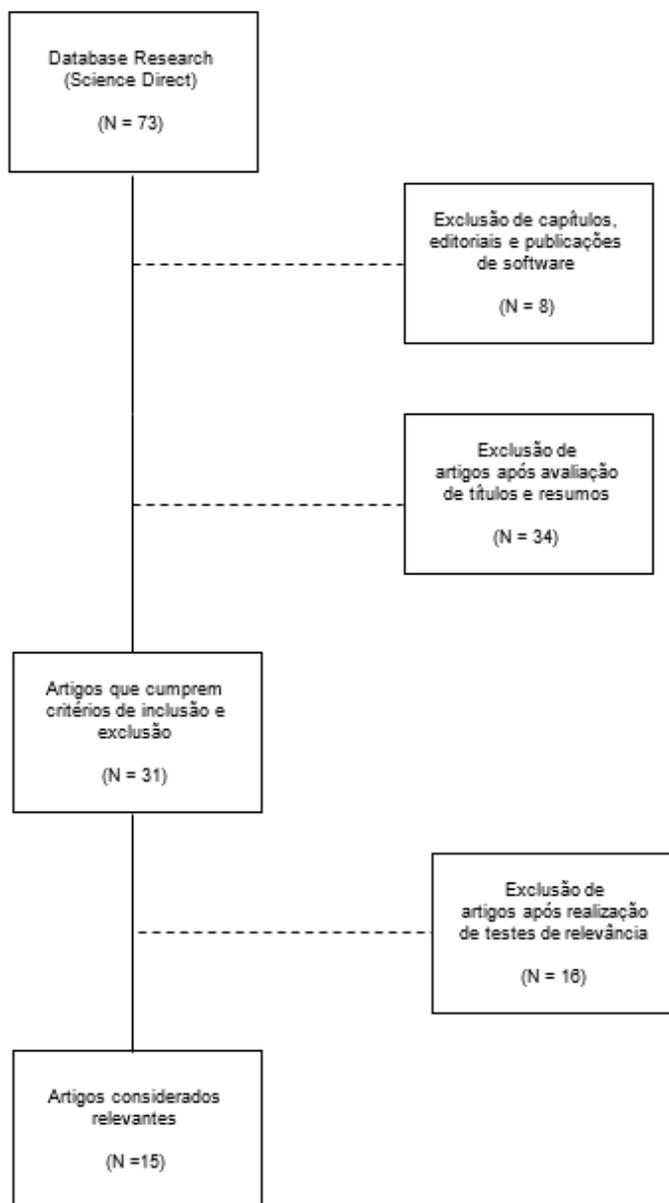


Figura 3.5. Diagrama de Resultados da Exploração da Base de dados

Com o teste de relevância são descobertos os artigos que devem fazer parte da RSL, ou seja, os que, com mais valor, contribuem para a mesma tendo em atenção ao tema a discutir e os objetivos definidos para a RSL.

O confronto dos artigos com estas questões resultou no estabelecimento de um total de 15 artigos que chegaram à fase final deste processo, sendo então estas as publicações que perante os parâmetros estabelecidos e aplicados contribuem com mais valor no que toca à investigação e exploração do uso de *big data analytics* ao nível da energia.

Os restantes artigos que não chegaram a esta fase do processo, tendo sido excluídos após realização dos testes de relevância podem ser observados na tabela 3.3 “Artigos Filtrados”, organizados pelas temáticas por estes abordadas.

Tabela 3.3. Artigos Filtrados

| Temas | Autores/Artigos |
|-----------------------|---|
| Sistemas de Energia | (Shiny et al., 2021); (Baker et al., 2017) |
| Gestão de Energia | (Sharmila et al., 2019); (Akhavan-Hejazi & Mohsenian-Rad, 2018); (M. wei Fan et al., 2019); (Le Ray & Pinson, 2020) |
| Privacidade | (Asikis & Pournaras, 2020); (Sheela & Sathesh Kumar, 2020); (Nadeem & Arshad, 2021) |
| Eficiência Energética | (Silva et al., 2020); (Rao et al., 2017); (G. Liu et al., 2018); (C. Fan et al., 2018); |
| Big Data | (Loughalam et al., 2017); (Elnakat & Gomez, 2016); (Kulatunga et al., 2017) |

A tabela 3.4 apresenta os artigos selecionados, os seus autores e ano de publicação juntamente com o objetivo da publicação, a(s) sua(s) área(s) de foco e as palavras-chave utilizadas pelas publicações consideradas para esta revisão.

Tabela 3.4. Publicações selecionadas para a extração de dados

| Publicações selecionadas para a extração de dados | | | |
|---|---|--|---|
| Autores e Ano | Objetivo do artigo | Áreas de foco | Palavras-Chave |
| (Zhou & Yang, 2016) | Estudo sobre formas de melhorar a eficiência energética e promover a conservação da energia através da análise dos comportamentos dos consumidores. | Eficiência Energética, Consumo Energético, Digitalização dos sistemas energéticos. | Household Energy Consumption Behavior, Energy Big Data, Big Data Analytics, Energy Information, Intervention Strategies |
| (Marinakis, Doukas, Tsapelas, Mouzakitis, et al., 2020) | Apresentação de uma arquitetura para uma plataforma capaz de suportar a criação, desenvolvimento e manutenção de serviços inteligentes de energia. | Gestão Energética, Previsão de consumo energético. | Big Data, Decision Support System, Energy Services, Intelligent Management, Smart Cities |
| (Munshi & Mohamed, 2017) | Apresentação de uma framework para o desenvolvimento de <i>smart grids</i> implementando o mesmo numa plataforma cloud | Smart Grid, Avanço tecnológico, Cloud | Smart grid, Smart meters, Big data, Data management, Dynamic demand response |
| (X. Liu & Nielsen, 2016) | Solução inovativa para aceleração ao processo de análise de <i>smart meter data</i> | Smart meter analytics, Integração de informação, Portal para visualização | ICT-solution, Smart meter data, Big data, Data analytics |
| (Wilcox et al., 2019) | Desenvolvimento de uma plataforma para <i>smart meter data analytics</i> apoiada de Hadoop | Armazenamento de dados, métodos, sistemas, ferramentas | Big data, Smart grid, Meter data analytics |
| (Marino & Marufuzzaman, 2020) | Desenvolvimento de um sistema <i>microgrid</i> apoiada de tecnologia Apache Spark para gestão energética | Apache Spark, Gestão Energética, Microgrid, Estratégia de big data | Sustainability, Microgrid, Big data, Spark streaming, Stochastic optimization, Wind power |
| (Yingfeng Zhang et al., 2018) | Implementação de uma <i>framework</i> para apoio na recolha de dados em indústrias com uso intensivo de energia | Eficiência energética, Consumo energético, produção limpa | Energy-intensive manufacturing industries, Big data analytics, Cleaner production, Data Mining |
| (Ardagna et al., 2021) | Abordagem para garantia de segurança de dados no setor energético fazendo a ponte entre os especialistas e os <i>data scientists</i> | Big data analytics, Segurança de dados, Smart Grid, Privacidade, Infraestrutura energética moderna | Artificial Intelligence, Big Data Analytics, Machine learning, Security and Privacy |

| Publicações selecionadas para a extração de dados | | | |
|---|---|--|--|
| Autores e Ano | Objetivo do artigo | Áreas de foco | Palavras-Chave |
| (Yassine et al., 2019) | Plataforma inovativa que tem o objetivo de recolher e providenciar informação acerca dos gastos de energia pelos consumidores | Smart grid, Smart home, Consumo energético, Gestão energética | Internet of things, Cloud computing, Fog computing, Big data analytics, Energy management, Smart homes |
| (Arif et al., 2021) | Apresentação de dois métodos desenvolvidos como resposta às perdas financeiras existentes nas empresas de energia devido a roubos de eletricidade | Smart Grids, Segurança, Gestão de Big Data, Deep Learning | Electricity theft detection, Smart grids, Machine learning, Deep Learning, Multi-layer perceptron |
| (Chou & Ngo, 2016) | Desenvolvimento de uma estrutura para <i>smart grid</i> que produza resultados ao nível do processo de poupança energética. | Poupança energética, Eficiência energética, Previsão e Otimização | Smart grid, Big data, Optimization, Time series data analytics, Energy saving, Home appliance, Web-based portal |
| (Chen et al., 2017) | Estudo sobre a importância e qualidade dos dados de consumo elétrico, e descoberta de dados problemáticos num ambiente <i>smart grid</i> . | Smart Grid, Energy Big Data, Data Quality, Consumo Energético | Electricity consumption data, Data quality, Outlier detection, Outlier data, Smart grid |
| (Renugadevi et al., 2021) | Discussão do uso de <i>smart grids</i> e o seu papel na introdução das <i>smart cities</i> , e como o <i>blockchain</i> pode responder às questões de segurança energética levantadas | Tecnologias inteligentes, desenvolvimento sustentável, gestão de recursos energéticos. | Smart city, IoT, Smart grid, Big data analytics, Blockchain |
| (Javaid et al., 2021) | Apresentação e simulação de uma técnica robusta de <i>big data analytics</i> para resolver problemas na deteção de situações de fraude interna e externa no uso de eletricidade | Deteção de roubo de eletricidade, dados de consumo, Smart grid, Técnicas de análise de dados | Big data analytics, Adaptive synthesis, Electricity theft detection, Deep learning, Deep siamese network |
| (Mokhtar et al., 2021) | Arquitectura baseada em <i>deep learning neural networks</i> para previsão da distribuição de voltagens através do uso de <i>smart meters</i> | Smart meters, Inteligência artificial, Redes inteligentes | Voltage prediction, Smart meters, Deep neural learning, Distribution network operation, Big Data Analytics, Analytic methods in power networks, Privacy-preserving data analysis |

3.2.4 Recolha de dados

A fase de recolha de dados numa RSL refere-se à interpretação e organização da informação investigada apresentando a mesma e tendo em conta a avaliação realizada até ao momento.

Para a recolha dos dados dos artigos selecionados para a revisão pretende-se considerar apenas a informação útil à pergunta colocada que origina a revisão, e, ao mesmo tempo, garantir a existência de variedade nos dados recolhidos. Para tal, os dados recolhidos a apresentar relatam a informação quantitativa, assim como, qualitativa para garantir uma mais completa demonstração da realidade atual de investigação presente na análise de grandes volumes de dados energéticos. Ao mesmo tempo são verificadas diferentes aplicações de modo a dar a conhecer melhor o espectro de alcance que a tecnologia de *big data analytics* atinge quando aplicada à energia, assim como, os efeitos repercutidos em diferentes áreas através da sua introdução e aplicação.

Durante essa recolha é fundamental garantir que em cada uma das publicações observadas, se efetua a recolha dos mesmos tipos de dados. Tendo esta informação em consideração serão extraídos para apresentação de resultados as seguintes informações: Título, autor, data de publicação, objetivos, considerações, metodologia, resultados e por fim as observações finais e limitações apontadas, se relevantes.

4 TRABALHO EMPÍRICO

Neste capítulo serão apresentados os resultados do trabalho de pesquisa realizado o que inclui uma descrição e síntese das temáticas abordadas dos artigos selecionados bem como a categorização dos resultados presentes nesses artigos no sentido de identificar como tem sido aplicado *big data analytics* na energia, assim como ferramentas utilizadas e as áreas afetadas pela aplicação de *big data analytics* na energia. Com a análise dos resultados pretende-se retirar conclusões sobre o estado atual de investigação sendo apresentadas as reflexões decorrentes da análise realizada através da discussão dos conteúdos explorados.

4.1 Sumário dos estudos selecionados

O processo de apresentação dos resultados obtidos é iniciado através de uma descrição das características dos resultados da pesquisa realizada. Como definido, os artigos que chegaram a esta fase do processo de filtro e avaliação encontram-se todos redigidos em inglês e com datas de publicação entre 2016 e 2022 (esse foi o intervalo de publicação definido), tendo o mais recente sido publicado a 2021 e o mais antigo publicado em 2016. A nível dos artigos incluídos na revisão verifica-se nos mesmos que estes apresentam o seu conteúdo com uma estrutura entendível para leitores que se encontrem em diferentes níveis de compreensão acerca do tema explorado.

Nos artigos verificam-se aplicações práticas, consequentes resultados, assim como, exploração de ideias que impulsionem ambas as áreas e ainda preparação e estruturação das mesmas para a sua aplicação em situações reais. Ao nível das temáticas exploradas, podemos observar nos artigos diversos temas como as *smart grid*, a transição para energias renováveis, eficiência energética, a poupança na utilização da energia, o papel de *big data analytics* no estudo do comportamento do consumidor de energia, o valor que *big data* produz para o estudo do uso da energia, o valor dos *smart meters* para *big data analytics*, as preocupações existentes ao nível da segurança de dados e da própria energia, algumas ideias de como enfrentar desafios atuais existentes na área, assim como a *big data* permite alcançar melhores resultados ao nível de como a gestão inteligente da energia é realizada. Ao nível dos locais onde os estudos foram realizados estes diferem, existindo uma variedade geográfica considerável, tornando-se possível admitir este tópico

de estudo como sendo um que tem impacto e interesse a nível global reforçando a importância desta revisão e o valor da mesma.

4.2 Análise dos resultados obtidos

A análise das publicações presentes nesta RSL permite retirar evidências de determinadas matérias que têm origem ou são impulsionadas com a interligação entre *big data analytics* e energia. Ao nível da análise e consequente apresentação dos resultados opta-se por categorizar estes estudos em diferentes subsecções que traduzem as categorias dos temas identificados na RSL para uma melhor compreensão e destaque dos temas abordados pelas publicações filtradas.

Em relação às categorias, estas foram selecionadas tendo em conta os resultados obtidos na tabela 3.4 que traduzem os 15 artigos relevantes para a RSL e os temas explorados nesses artigos.

4.2.1 Estudo do consumidor de energia

No artigo de Zhou & Yang, (2016), “*Understanding household energy consumption behavior: The contribution of energy big data analytics*”, os autores exploram formas de melhoria para a correta utilização de energia e preservação da mesma. Através da introdução de tecnologias da informação no setor energético torna-se possível verificar uma revolução nos sistemas de energia existentes ao introduzi-los ao mundo digital. A introdução de novas ferramentas, como os *smart meters*, tecnologias de IoT e *big data analytics*, resultam numa nova visão sobre os dados existentes de energia, permitindo um estudo mais acertado acerca dos comportamentos dos consumidores de energia.

Os autores apresentam a garantia de que, graças à introdução das TIC, torna-se agora possível utilizar o conhecimento adquirido através das novas ferramentas para estudar os comportamentos, fatores psicológicos e restantes fatores, quer subjetivos, assim como, outros mais objetivos tais como tamanho das famílias, condições climatéricas, preços da energia.

A possibilidade de estudar estes dados a um nível mais aprofundado, em oposição ao que era possível durante a era dita tradicional, possibilita, segundo os autores, a formulação

de estratégias de intervenção como envio de *feedback*, estabelecimento de objetivos, notificações e advertências que consigam ter um efeito sobre os comportamentos dos consumidores e encorajar os mesmos para uma conservação do uso da energia.

No mesmo artigo os autores expõem que, através do uso de *big data analytics* associado ao estudo do uso de energia pelos consumidores, é possível observar dois diferentes paradigmas: paradigma comportamental e económico.

No que respeita ao paradigma comportamental consideram-se fatores externos como incentivos interpessoais, tais como normas, e intrapessoais, como hábitos, atitudes e valores, que podem ser influenciados através da introdução de estratégias ao nível das consequências, antecedentes e estrutura.

Para o paradigma económico exploram o comportamento dos consumidores para a tomada de decisão. Essa tomada de decisão tem por base as consequências económicas dessas decisões, argumentando que programas de *demand side management*, assim como a aquisição de equipamentos energeticamente mais eficientes e o uso de mecanismos mais complexos e dinâmicos para a determinação de preços da energia poderá surtir efeitos no consumo.

Os autores expõem que o conhecimento existente sobre os comportamentos de consumo dos utilizadores e dados de *energy big data* recolhida tem implicações práticas importantes para o estabelecimento de políticas de crescimento da eficiência energética e redução do consumo. Destacam o valor que estes dados têm para a mitigação de mudanças climáticas e promoção de desenvolvimento sustentável, enfatizando a necessidade dos governos, comunidades locais e empresas fornecedoras de energias através da promoção da alteração de comportamentos relacionados com o consumo energético motivarem os utilizadores para a alteração dos seus comportamentos.

Na conclusão do artigo, os autores terminam o mesmo apresentando a ideia de que, embora seja usada uma proporção substancial de energia pelas famílias, os padrões de consumo têm um grau de variedade elevado entre diferentes famílias devido aos diferentes fatores intrapessoais, interpessoais e externos. Terminam o seu artigo com a partilha das impressões por eles retiradas, em que com a entrada das novas tecnologias capazes de detetar, medir e comunicar com a rede, juntamente com os avanços em *cloud*

computing e big data analytics, têm aumentado as quantidades de dados de consumo energético recolhidos. Este aumento de conhecimento irá resultar, segundo os autores, em mais investigação nesta área, assim como novos desenvolvimentos associados a *energy big data analytics* e ao estudo de novas oportunidades para compreensão dos comportamentos de consumo energético nas casas dos consumidores.

Os autores Yassine et al., (2019), no seu artigo “*IoT big data analytics for smart homes with fog and cloud computing*” através de uma aplicação de conjunta de IoT e *big data analytics* apresentam uma estrutura de uma plataforma para a recolha de dados, processamento e reconhecimento em tempo real que permite obter informações acerca das ditas *smart homes* que produzem dados sobre o consumo de energia dos seus habitantes. Os autores indicam ainda que o uso da plataforma de recolha de dados resulta em benefícios conjuntos para ambos, fornecedores de energia e consumidores.

Ao nível da contribuição dos autores com o seu artigo, estes argumentam que os principais contributos passam pela apresentação de uma proposta para uma plataforma, justificando o processamento e análise de dados como fundamental para aplicações de gestão de energia em *smart grids* e o uso dado a essa energia. No referido artigo relatam, ainda, os detalhes, requisitos e *design* dos componentes da arquitetura da plataforma, focando-se nos requisitos para o processamento de várias fontes de dados. Após explicação da arquitetura procedem à realização de um caso prático onde analisam as rotinas de consumo energético e padrões existentes numa *smart home*.

A plataforma de *big data analytics* proposta pelos autores explora o uso de IoT *big data analytics* com *fog e cloud computing*. O sistema desenhado permite o processamento de grandes quantidades de *smart home IoT data* através de *fog nodes*. Segundo os autores, os *fog nodes* permitem expandir os serviços do sistema *cloud* para o ponto da rede que se encontra fisicamente mais próximo da *smart home*. Ao nível desta plataforma os autores destacam a distribuição de recursos, a possibilidade de realização de uma análise escalável, a performance da plataforma, os procedimentos para integração de dados de diferentes dispositivos e a visualização como sendo requisitos fundamentais para o funcionamento da sua proposta. Já ao nível de *design* da plataforma os componentes incluídos correspondem aos da *smart home*, ou seja, aparelhos, equipamentos e sistemas

de contagem de uso energético, os serviços de gestão de IoT e integração, os *fog nodes* e o sistema *cloud*.

No caso de estudo realizado pelos autores, estes realizam uma aplicação prática do modelo por eles desenvolvido através de uma *smart home* em Vancouver, no Canadá. Realizam uma análise de dados de *IoT smart home* recolhida para a previsão e comportamento dos ocupantes da casa quanto ao consumo energético, rotinas e padrões, discutindo os resultados obtidos no contexto de gestão do esforço energético exigido pelos utilizadores, permitindo assim evitar emergências energéticas e redução de custos de eletricidade. Este caso de estudo apresentado pelos autores é realizado com o objetivo de analisar os padrões de uso dos diferentes aparelhos elétricos existentes na casa, combinando essa informação com o contexto para o seu uso e a que horas eram utilizados. O estudo realizado contém várias fases, desde o processo de limpeza e preparação dos dados recolhidos, aplicação de *frequent pattern mining* para descobrir horas, dias e duração de uso através da examinação do estado on/off dos equipamentos. Numa fase final é aplicado o *cluster mining* para descobrir informações mais profundas ao nível do consumo energético dos equipamentos em tempos específicos, como o pico de horas de utilização entre os diferentes habitantes.

Na conclusão do artigo definem a sua aplicação como sendo uma que pode levar a diferentes tipos de resultados e potenciar diferentes desenvolvimentos consoante o estudo da informação recolhida. Segundo os autores a plataforma pode auxiliar na identificação de padrões de consumo de energia, planeamento de poupança de energia, prever a necessidade de manutenção de vários equipamentos existentes na habitação para assim evitar custos de manutenção e reparação elevados, garantir uma operacionalização mais eficiente dos equipamentos num ponto de vista de conservação da energia e possibilitar aos fornecedores planos de redução energético baseado no consumo energético dos utilizadores.

4.2.2 Smart meters e big data analytics

No artigo de X. Liu & Nielsen, (2016), intitulado “*A hybrid ICT-solution for smart meter data analytics*”, os autores iniciam o seu artigo apresentando os *smart meters* como sendo uma peça fulcral para a mensurabilidade da energia consumida num período de tempo

devidamente definido. Os dados recolhidos por estes equipamentos são normalmente agregados a dados sociais e económicos, assim como, as localizações geográficas dos contadores, condições climatéricas e informações dos seus utilizadores. Segundo os autores, a análise de dados destes contadores inteligentes pode auxiliar os fornecedores de energia na compreensão de padrões de consumo, responder melhor às necessidades energéticas em horas de ponta, apoiar na deteção de roubos energéticos e enviar *feedback* personalizado aos clientes.

Os autores neste artigo apresentam uma solução híbrida de tecnologias de informação criada para facilitar todo o processo de *smart meter data analytics*, incluindo a ingestão de dados, transformação, carregamento, análise e visualização. Esta solução permite aos utilizadores e fornecedores de energia a obtenção de informação quase em tempo real durante todo o processo. O artigo dos autores encontra-se dividido da seguinte forma: apresentação da proposta de solução híbrida que envolve diferentes tecnologias, como o Spark e Hive - para o processamento de dados - e o PostgreSQL e MADlib para a componente analítica, sendo o design constituído para considerar o suporte de *high-performance analytics* e *big data analytics*. O artigo inclui ainda a implementação de uma plataforma para processamento de dados, que pode ser facilmente expandida, e uma plataforma de análise de dados capaz de suportar *supply and demand-side analytics*, ou seja, que apresente a informação relacionada com os padrões de consumo dos consumidores, onde os fornecedores podem melhor planear para aumentos de necessidades energéticas, conhecer melhor os consumidores e ainda oferecer serviços personalizados. Por sua vez os consumidores têm uma compreensão do seu consumo o que os pode ajudar a monitorizar o seu gasto energético e poupar energia. Depois abordam a implementação de uma plataforma para análise de dados que consegue suportar os dois lados, o da oferta e da procura, do processo de análise, tendo esta plataforma o objetivo de auxiliar os fornecedores de energia a gerirem a energia produzida e, ao mesmo tempo, auxiliar os seus clientes, permitindo uma poupança monetária, uma vez que o encargo financeiro diminui graças a uma gestão mais eficiente da energia. A última parte do artigo consiste numa análise à plataforma proposta onde avaliam e discutem as tecnologias utilizadas para o desenvolvimento da plataforma, discutindo em que fases se devem utilizar certas tecnologias. Para exploração desta última componente, os autores avaliam

a sua proposta argumentando como esta se comporta confrontando a mesma com cenários de estudo da variabilidade de consumos por cada consumidor, sensibilidade térmica ao nível do consumo elétrico por consumidor, a criação de um perfil diário do uso energético dos consumidores e a deteção de anomalias no consumo de energia por consumidor. Para a testagem destes exemplos, os autores utilizam dados históricos para os três primeiros casos tendo apenas optado por dados em tempo real para a deteção de anomalias. Ao nível da discussão dos resultados, os autores explicam, na sequência das suas experiências que é favorável adotar o uso de *clusters* usando o Hive ou Spark para obter escalabilidade elevada e recomendam o uso do PostgreSQL/MADlib para gestão de dados socioeconómicos e estatísticos de *smart meters*.

Na conclusão do seu artigo os autores, referem as oportunidades originadas pelos *smart meters* e *smart meter analytics* para os fornecedores e clientes. Ao nível dos fornecedores, esta tecnologia permite uma melhoria no apoio ao cliente, redução de custos e melhor eficiência energética. Já para os clientes, o uso destes contadores permite-lhes poupar energia e reduzir a conta a pagar. Ao nível da solução proposta, esta apresentou, segundo os autores, resultados positivos de eficiência e eficácia para os testes por estes realizados. Estes testes permitem, na opinião dos autores concluir que a solução proposta consegue analisar grandes volumes de dados recolhidos durante um período definido, ou fazer análises através de um processo contínuo de carregamento de dados. Ao nível de possíveis melhorias futuras para a solução, os autores apontam para a incorporação de mais algoritmos para apoio na análise de dados, assim como, expandir o sistema proposto para que este possa suportar outros tipos de *smart meter data* como água, gás e aquecimento, garantir um maior suporte para gestão de redes de distribuição ligadas às *smart grids* e estudar como o seu sistema pode apoiar na gestão energética para os fornecedores de energia.

No artigo de Wilcox et al., (2019), “*A big data platform for smart meter data analytics*”, os autores relatam a falta de investigação para o desenvolvimento e uso de métodos, sistemas e ferramentas para oferecerem suporte ao nível do armazenamento de dados e análise de dados para *big data* gerada pelas *smart grids*. Para combater esta lacuna de investigação, apresentam uma solução para os problemas de *big data* relacionados com a gestão de dados dos contadores, incluindo na solução proposta um novo sistema,

descrição das escolhas ao nível da tecnologia utilizada e a sua implementação, terminando o mesmo com um teste realizado com dados reais de *smart meters* do Reino Unido.

Durante a exposição do tema, é argumentado pelos autores que a exploração de *meter data analytics* tem sido realizada com vários propósitos, mas focada apenas em conjuntos de dados de tamanho mais reduzido. Como tal, exploram a análise de um grande volume de *smart meter data* relativa a um grande período de tempo.

Para a gestão e DM de grandes volumes de dados propõem um sistema *core-client-broker*, a sua implementação traduz-se numa plataforma de *big data* de nome SMASH ou “*Smart Meter Analytics Scaled by Hadoop*”. Apresentam a arquitetura desta plataforma ilustrando os componentes mais relevantes da sua arquitetura como o armazenamento dos dados, a plataforma de *big data* incluindo o *hardware*, *middleware* e aplicações de DM, sendo todo este processo incluído numa interface que permite a análise final que resulta nas decisões a tomar tendo em conta os dados analisados. Ao nível da implementação do sistema, são descritos *hardware*, *software* e integração, onde explicam o raciocínio dos mesmos para a adoção de certos componentes do sistema como o uso do Hadoop e o fluxo de eventos ocorridos no seu sistema. Para a testagem e exploração da plataforma utilizam os dados recolhidos pelos *smart meters* existentes em diversas habitações presentes no Reino Unido, as quais recolhem dados em intervalos de trinta minutos. São descritos os resultados do uso da plataforma fazendo comparações entre os resultados obtidos com o uso da SMASH com os resultados publicados da Google Cloud, IBM Informix Timeseries, MongoDB e AMPLab. As comparações realizadas são feitas ao nível da velocidade de inserção de dados e velocidade de processamento dos mesmos. Por fim os autores fazem uma exposição de como os dados são visualizados através da SMASH são apresentados para visualização.

As conclusões indicam que a crescente recolha de grandes volumes de dados via *smart meters* transformou as necessidades de recolha de dados ao nível do período temporal da recolha, tendo esta passado de uma periodicidade correspondente a um quadrimestre para períodos de trinta minutos. Na opinião dos autores, os resultados obtidos através da plataforma permitem competir com os produtos presentes no mercado, sendo a plataforma descrita como uma alternativa fácil de operacionalizar e capaz de gerir a *energy big data*,

auxiliar na visualização dos dados recolhidos e preparados tendo potencial para providenciar suporte ao nível da tomada de decisão.

De acordo com os autores Mokhtar et al., (2021), no seu artigo “*Prediction of voltage distribution using deep learning and identified key smart meter locations*” a instalação massiva de *smart meters* e avanços na infraestrutura dos contadores são realizados com a intenção de garantir uma gestão mais inteligente da rede energética.

O artigo inicia com uma descrição sobre as *Low Voltage Networks*, redes com voltagem reduzida, e como estas estão a sofrer rápidas transformações, deixando de fluir apenas diretamente para os consumidores, passando os consumidores, através destas redes, a representar um papel mais ativo na produção de energia e distribuição da mesma. A estas mudanças acrescentam-se os crescentes requisitos energéticos do aumento de popularidade dos automóveis elétricos e transformações ao nível do uso da eletricidade para outros fins.

Segundo os autores, os *smart meters* e a infraestrutura construída para a sua gestão são uma parte crucial para o aumento da observabilidade de espaços previamente de dificuldade acrescida de visualização destas redes, tendo os mesmos um papel importante na gestão ativa e futura das *low voltage networks*, em especial para a prevenção de riscos para as mesmas. No entanto, segundo os autores, a expectativa de recolha sem erros de dados de todos os *smart meters* é irrealista, devido à dificuldade em garantir uma operabilidade sem problemas técnicos, assim como, devido às crescentes preocupações ao nível da privacidade dos utilizadores. Ao mesmo tempo os *smart meters* são importantes para as redes de distribuição, sendo através dos *smart meters* recolhidos dados importantes para a estimativa de voltagens em todos os pontos da rede. Tendo em consideração o problema da privacidade de dados, mas ao mesmo tempo a necessidade dos mesmos para a previsão do consumo, os autores questionam a necessidade de usar a totalidade dos dados para gestão das redes *low voltage* ou se seria possível calcular a distribuição da voltagem através de pontos-chave da própria rede.

Após exposição do problema, os autores procedem à descrição da metodologia utilizada onde propõem o uso de *deep learning neural networks* para a previsão da voltagem distribuída num circuito de *low voltage* utilizando apenas um número mínimo de *smart*

meters localizados em locais chave da rede. Após descrição da metodologia, iniciam uma simulação da distribuição da voltagem através de diferentes cenários de uso de *smart meters* em circuitos *low voltage* domésticos, tendo utilizado o OpenDSS para a simulação e explicando como diferentes tipos de propriedades podem afetar a simulação visto que detêm de diferentes necessidades energéticas. De seguida, procedem à realização de uma previsão da voltagem explicando os procedimentos utilizados. Após a previsão, o modelo preditivo é treinado utilizando a informação da procura energética equivalente a um mês tendo também sido simulada uma situação de falta de cobertura de *smart meters*. Com a simulação terminada, fazem a identificação dos locais chave da rede visto que os dados recolhidos destes locais são os únicos capazes de providenciar uma previsão efetiva.

Nas conclusões do processo de previsão evidencia-se que as medições realizadas pelos *smart meters* dos locais chave e o seu modelo de previsão baseado em *deep learning* são suficientes para uma previsão eficiente da distribuição de voltagem, chegando mesmo a considerar que os dados recolhidos ao nível das necessidades energéticas dos consumidores individuais não são necessários para o processo de previsão evitando assim potenciais problemas ao nível da privacidade dos dados. Os autores destacam o uso de redes *low voltage* como um elemento central de transição de energia, que terá de acomodar aumentos significativos em tecnologias de geração de energias, distribuição e armazenamento da mesma, do seu efeito na análise da necessidade de instalação de *smart meters* para garantia de uma melhor visão sobre a energia e da forma como esta é distribuída pelas redes.

4.2.3 Smart Grids

No artigo de Munshi & Mohamed, (2017), intitulado “*Big data framework for analytics in smart grids*”, é explorada a natureza das *smart grids* e como estas podem se revelar desafiantes tendo em atenção os requisitos necessários para a sua correta utilização, considerando a necessidade de técnicas avançadas de análise juntamente com infraestrutura capaz de estar em conformidade com a análise de grandes volumes de dados. No artigo é apresentada uma estrutura para uma plataforma que é proposta de forma a permitir efetivamente tirar proveito das crescentes quantidades de dados.

Para apresentação da sua solução, são descritas as tecnologias *open-source* e a forma como podem surgir como ferramentas que conseguem providenciar o desenvolvimento de uma plataforma a um custo reduzido e acessível. De seguida, os autores exploram a estrutura por eles desenvolvida que permite lidar com o ciclo de vida da *smart grid big data* desde a geração dos dados até à *data analytics* realizada.

Na descrição desta estrutura, são apresentados os componentes principais para a desenvolvimento de uma plataforma de *big data* para *smart grids*, desde a aquisição, distribuição e processamento, à consulta de dados, sua análise e visualização. Ao nível da geração dos dados, os autores consideram dados de carros elétricos, residências, edifícios, centrais elétricas, painéis solares, turbinas, fábricas e eventos naturais. Para a aquisição dos dados gerados por estas fontes utilizam o Flume para recolher, agregar e transferir os dados. Destacam o Hadoop como sendo extremamente popular tanto a nível industrial como educativo, justificando o seu valor para a gestão de *big data* em *smart grids* e sua análise utilizando o mesmo para o armazenamento, processamento e *querying* de dados. Os autores terminam a explicação da estrutura referindo-se à aplicação de DM, preparação visual e análise dos dados, assim como, formulação de previsões. Ao longo da descrição dos componentes vão apontando simultaneamente ferramentas *open-source* a considerar para cada uma das partes do processo de gestão e análise de dados.

Após a descrição dos principais componentes a utilizar, é descrita a implementação da estrutura numa plataforma de *cloud computing*. Terminada a demonstração da sua implementação e exploração dos componentes que compõem todo o processo de análise onde se inclui a exploração visual através do uso de ferramentas de *business intelligence*, nomeadamente o Tableau, os autores optam pela realização de duas aplicações práticas diferentes da estrutura por eles desenvolvida.

Para permitir testar a plataforma e visualizar o uso da energia, os autores realizaram a primeira aplicação da sua proposta numa única habitação com vários pequenos geradores de energia sobre a forma de turbinas e painéis solares e incluem um veículo elétrico dada a sua popularidade. Na sua segunda aplicação, os autores aplicam a sua estrutura a um banco de dados de *smart meters* que continha no mesmo dados de 6.436 casas e negócios.

Estas duas aplicações permitiram aos autores visualizar o estado da rede em duas circunstâncias completamente diferentes.

Na conclusão do seu artigo os autores debruçam-se sobre toda a investigação, quantitativa e qualitativa por eles realizada, argumentando que o facto de a sua estrutura conseguir promover a visualização do estado da rede energética em dois cenários muito diferentes justifica o seu valor para a *smart grid data analytics*. Terminam o seu artigo fazendo o apontamento de que a estrutura por eles desenvolvida, incluindo toda a sua configuração, código utilizado e aplicação poderá surtir efeitos positivos no desenvolvimento de outras plataformas de *big data*, associadas a outras áreas produzindo resultados ao nível dos negócios e promoção do desenvolvimento de ciências e tecnologias que melhorem a tomada de decisão, noutros setores apontando o setor governamental como exemplo.

No artigo de Chou & Ngo, (2016), “*Smart grid data analytics for increasing energy savings in residential buildings*” observamos a apresentação de outra estrutura para *smart grid big data analytics* e componentes necessários para o desenvolvimento de um sistema de suporte à decisão desenhado para a poupança energética.

Os autores iniciam o seu artigo fazendo uma exploração sobre os aumentos verificados ao nível do consumo energético e os efeitos que este consumo tem tido na emissão de gases de efeito de estufa, descrevendo os aumentos percentuais em diversos setores durante as últimas duas décadas. De seguida, descrevem os sistemas de *smart grid* e como estes oferecem soluções promissoras para dar resposta ao aumento de necessidades energéticas. Incluem alguns aspetos não referidos ainda nos artigos até aqui já explorados no âmbito deste trabalho de dissertação, como o seu papel na redução de custos de congestionamento, a capacidade de melhor acomodar as energias renováveis e o impacto que tem ao nível ambiental face às redes mais tradicionais.

Após exploração dos benefícios do sistema, iniciam a descrição da estrutura desenhada para *smart grid big data analytics* e como esta foi testada num edifício residencial. A descrição destaca o que a estrutura proposta permite alcançar, referindo que esta consegue recolher e continuamente analisar dados em tempo real ao nível da eletricidade consumida, identificar padrões de consumo de energia, prever futuros consumos realizados, gerar e catalogar os tempos ótimos para uso de diferentes aparelhos.

Os autores fazem uma aplicação da arquitetura por si desenhada através da integração da mesma num edifício com diferentes tipos de necessidades energéticas por divisão e pisos. Utilizam esta implementação prática como forma de justificar a arquitetura desenhada pelos próprios e o conhecimento que a mesma permite atingir justificando, assim, todos os pontos que argumentam no início do seu artigo acerca do valor dos sistemas de *smart grid*.

Ao nível da arquitetura desenhada, os autores apontam para três camadas: uma primeira que corresponde à camada de dados que compreende a gestão dos mesmos, descrição da infraestrutura aplicada e a sua configuração explicando ainda como todas as partes intervenientes no sistema comunicam entre si; na segunda camada, os autores apontam para a área de análise descrevendo as técnicas de DM aplicadas, *time series analysis* para identificação dos padrões de uso da energia e algoritmos dinâmicos multi-objetivo utilizados para a otimização e entrega de informação acerca dos tempos ótimos para utilização dos aparelhos; na terceira e última camada descrevem o portal ou área de interface que permite ao utilizador interagir com o sistema de suporte à decisão para poupança energética.

Ao terminarem a descrição do sistema, concluem destacando o potencial que o trabalho desenvolvido tem para servir como base ao desenvolvimento de um sistema mais completo de suporte à tomada de decisão inteligente. Descrevem este sistema como sendo capaz de, para além de melhor informar sobre o consumo ou alertar os consumidores para as melhores horas para o uso de energia, ser ainda capaz de formular estratégias para o seu uso podendo o mesmo ajudar numa utilização mais eficiente da energia, alocando a energia a utilizar através de *machine learning* e otimização, reduzindo custos de eletricidade, através de melhorias ao nível do uso de sistemas de iluminação, aquecimento, ventilação e ar condicionado.

No artigo de Marino & Marufuzzaman, (2020), com o título “*A microgrid energy management system based on chance-constrained stochastic optimization and big data analytics*” os autores descrevem as *smart grids* como sendo de tamanho consideravelmente grande, complexas e frágeis, e, tendo em conta estas características, descrevem as *microgrids* como sendo ferramentas poderosas capazes de alterar a forma

como a energia é produzida, distribuída e consumida. Uma grande diferença perante os artigos já aqui apresentados passa pelo foco no uso de energia eólica. Os autores descrevem as *microgrids* como sendo um pequeno ajuntamento de fontes ativas de distribuição de energia próximas dos consumidores finais que conseguem operar ligadas diretamente à rede ou num modo mais autónomo.

Segundo os autores, estas redes mais reduzidas - em tamanho - estão a ter um impacto considerável na forma como a eletricidade é produzida, no entanto, segundo a sua revisão da literatura, existe uma falta de pesquisa e aplicação de *big data analytics* junto das *microgrids*. O seu artigo é realizado no sentido de explorar as *microgrids* junto de tecnologias de *big data*. Para o fazer procuram responder como podem incorporar a *framework* do Apache Spark ao modelo de otimização existente da *microgrid* e explorar qual seria a redução do custo de energia ao utilizarem este processo de *big data analysis*.

Para responder às questões colocadas, iniciam a sua resposta com uma descrição de uma metodologia e arquitetura inovativa de tecnologia de *big data* capaz de absorver grandes volumes de dados climáticos, para prever continuamente a criação de energia através do vento. Descrevem o sistema, formulam o modelo apontando as variáveis e restrições definidas, exploram a primeira fase do uso do Apache Spark para introduzir um método focado nos resultados em tempo real. Os autores classificam o seu trabalho como de investigação pioneira ao nível da arquitetura e modelo introduzido na integração de tecnologias de *big data* na otimização de sistemas complexos de energia.

Na fase seguinte do seu artigo propõem um algoritmo *Simple Average Approximation*, capaz de providenciar soluções para resolver o modelo de otimização por eles construído, tendo ainda optado pela inclusão, no algoritmo, de detalhes que certifiquem a preferência pelo uso de energias renováveis.

Os autores desenvolvem ainda um caso de estudo real onde estudaram dados de uma empresa de fabricação baseada em San Diego, onde testaram a *performance* do modelo de otimização, em diferentes cenários, recorrendo ao uso de dados climáticos históricos. A aplicação dos autores resultou em resultados que indicam que o uso de dados em tempo real e tecnologias de *big data* permitem a obtenção de melhores resultados ao nível da redução dos custos de operacionalização da *microgrid* em comparação com técnicas de

análise tradicional. Segundo os autores, o utilizador do sistema, consegue, através do modelo introduzido obter mais informação, tal como visualizações em tempo real, o que permite obter vantagens ao nível dos processos de tomada de decisão. Mencionam ainda que a construção e desenvolvimento de novos modelos de previsão é importante porque permite a melhoria da previsão da energia eólica o que leva a maior poupança de custos energéticos. Por fim, apontam que o uso do modelo representa uma ferramenta efetiva na alocação da energia gerada e colmata o risco de redução da energia gerada face ao esforço realizado.

Ao nível das conclusões do artigo, os autores referem-se ao uso de tecnologias de *big data* e ao potencial que estas apresentam na otimização de geração e gestão energética e distribuição em sistemas energéticos coletivos ou em individuais como as *microgrids* e refletem ainda sobre o papel que *big data* tem na integração de energias renováveis. Apontam direções futuras para a arquitetura proposta, onde indicam a necessidade de comparar a mesma durante período ditos normais e períodos com disrupções como falhas de equipamentos ou na própria rede.

No artigo dos autores Chen et al., (2017), com o título “*Data quality of electricity consumption data in a smart grid environment*”, os autores exploram como o crescente desenvolvimento da economia conduz a uma necessidade acrescida de energia dado o papel da mesma no desenvolvimento da sociedade. Segundo os mesmos a construção de variadas *smart grids* e a aplicação das mesmas tem vindo a ser realizada de modo a criar uma forma de gerir os aumentos de necessidade de energia, tendo as mesmas contribuído com variadas informações ao nível do uso de energia, estimulado a acumulação de dados operacionais da mesma e dados da gestão da sua produção. No entanto, apontam para a existência de dados redundantes, dados em falta, ou *outliers* nos conjuntos de dados agregados.

Os autores mencionam como a escalada de dados de consumo elétrico recolhidos através de infraestruturas avançadas de contadores contidos nas *smart grids* tem aumentado cada vez mais. Segundo os mesmos a qualidade destes dados estão a ter um impacto direto na qualidade e eficiência do sistema energético e na aplicação baseada em *big data*, pelo que a investigação existente em qualidade de dados de consumo de energia torna-se

importante para os processos de DM e detém de um papel importante para a *energy big data analytics*.

Neste artigo, após apresentarem a importância da existência da qualidade dos dados energéticos utilizados descrevem a qualidade de dados de consumo elétrico, incluindo as características dos mesmos e a sua classificação num ambiente de *smart grid*, onde se focam especificamente em três tipos de problemas de qualidade de dados: 1) dados corrompidos; 2) incompletos e 3) *outliers*. Ao nível dos dados energéticos corrompidos, segundo os autores, são dados que possuem erros lógicos, que não cumprem as regras do negócio ou lógicas aplicadas, como por exemplo, os dados mensais de gastos energéticos conterem mais dias do que o que é suposto, enviesando os resultados obtidos ou possuírem valores inconsistentes, tais como valores negativos relativos ao uso de energia. No que respeita a dados incompletos, afirmam ser algo recorrente nas *smart grids* devido à complexidade do ambiente onde se centram, o que leva a que sejam aplicadas técnicas capazes de colmatar estas situações. Em relação aos dados que são categorizados como *outliers*, os autores mencionam ser estes os mais importantes, porque, ao contrário dos corrompidos e incompletos, no que toca a dados energéticos estes não são dados com erros, mas sim dados que exigem uma análise mais aprofundada porque podem conter informações de situações relevantes que não podem ser ignoradas. Após introdução destes três tipos de dados, o artigo foca-se nos processos de deteção de *outliers* nos dados de consumo de energia descrevendo os diferentes métodos existentes para a sua deteção. Analisam ainda as causas da existência destes tipos de dados, assim como, a sua significância no processo de análise. Ao nível da deteção destes *outliers*, os autores mencionam que estes são sumarizados através da aplicação de DM e métodos de estimação de estado. No que toca aos métodos de deteção de *outliers* através de DM, estes mencionam quatro métodos para a sua deteção: 1) baseados na *neural network*, 2) na *fuzzy theory* e *cluster analysis*, 3) no algoritmo estatístico, e em 4) *time series analysis*.

Nas conclusões do seu artigo, apontam para que, no processo de aplicação e partilha de *big data analytics* de dados elétricos, assegurar a qualidade dos mesmos é uma peça fundamental na tomada de decisão sobre os sistemas energéticos. Terminam o mesmo fazendo uma nota de futuros caminhos a seguir na exploração de processos de deteção e identificação de *outliers* nos dados de consumo elétrico, fazendo referência ao potencial

que as *smart grids* têm para recolha de dados energéticos em tempo real o que permite investigar futuramente a deteção de *outliers* em tempo real, assim como, ao nível de deteção dos mesmos quando provenientes de falhas nos equipamentos da rede, erros nos parâmetros e intrusões na rede.

No artigo de Renugadevi et al., (2021), intitulado “*IoT based smart energy grid for sustainable cities*”, os autores discutem os avanços tecnológicos relacionados com a *internet of energy*, o impacto dos mesmos nas *smart cities* e transformação das mesmas em cidades mais sustentáveis. Abordam a tecnologia de *smart grid* e aplicações da mesma e terminam com uma discussão acerca da inclusão da *blockchain* como ferramenta de apoio às questões de segurança existentes.

Os autores introduzem o tema abordando o impacto das *smart grids* para as *smart cities* e o relevante papel que *big data* e *big data analytics* têm na transformação das funcionalidades das *smart grids* para uso eficiente da energia nas *smart cities*.

De seguida apresentam uma explicação das diferentes tecnologias, desde a *smart grid* à IoT na *smart grid*, terminando com alguns desafios existentes para as *smart grids* e conhecimento por estas produzido.

Segundo os autores, os recursos energéticos têm de ser armazenados e geridos de forma eficiente para corresponder às necessidades da geração futura, como tal, procedem a explicar as *smart grids* e como são capazes de expandir a capacidade aumentando a possibilidade de atividades *smart* e fluxo de dados, os sistemas de gestão inteligente e o modelo para a gestão inteligente da energia.

Ao apresentar as motivações para aplicação das *smart grids* argumentam que a gestão da energia na *smart grid* permite alcançar quase todos os cantos das cidades para construção de uma infraestrutura de rede inteligente que promove a interação com o utilizador. Os autores terminam a sua explicação sobre o papel vital das *smart grids* enunciando as principais características que compõem o sistema de energia inteligente como a sinergia de diferentes tipos de energia, simetria de informação, distribuição da procura energética, a construção de sistemas mais fixos e transações abertas.

Em relação aos sistemas de energia inteligente baseados em IoT, segundo os autores, os principais componentes são os *energy cloud servers* e *energy edge servers*, capazes de dissociar os dados recolhidos pelos *smart meters* dos restantes equipamentos *IoT*. Na explicação dos sistemas, apresentam ainda o papel dos dispositivos de energia em todo o fluxo existente no sistema. Ao nível do modelo para a gestão inteligente da energia, segundo os autores, este é composto por quatro camadas: 1) camada de reconhecimento; 2) camada de rede; 3) camada de conhecimento; 4) camada de aplicação. Procedem à explicação do fluxo da *energy big data* e a sua passagem por todas estas camadas, assim como, os intervenientes que compõem as mesmas.

Na fase seguinte, o artigo expõe várias aplicações para a *smart grid*, evidenciando a expansão em todos os campos, desde residencial, industrial, ao campo de negócio. Na descrição das aplicações referem algumas já exploradas na presente dissertação, assim como, algumas ainda não mencionadas, por exemplo: o uso de tomadas inteligentes denominadas *smart plugs* como peça fundamental para a aplicação de uma rede inteligente baseada em IoT. Estas *smart plugs* atuam como sensores que ajudam a reduzir as faturas de energia e distribuição eficiente da carga elétrica, tendo ainda o benefício de poderem estar ligadas à Internet e configuradas para o envio de notificações lembrando os utilizadores de desligar de certos equipamentos ou máquinas ligadas a essas tomadas durante os períodos de maior procura de energia.

Por fim, os autores apresentam o uso de *blockchain* para garantir uma mais rede energética mais segura. Começam por explicar o papel que esta tecnologia tem desempenhado ao nível de preocupações com segurança fazendo, de seguida, uma comparação com a *internet of energy*. Segundo os autores, a *blockchain* tem a capacidade de alterar a indústria de fornecimento de energia e apontam para várias aplicações de *blockchain* associada à *internet of energy* como a gestão de ativos, maior segurança de autoria de recursos, *P2P energy trading*, gestão de ameaças, calendarização de cargas, pagamentos, entre outros. Terminam o seu artigo com a descrição de futuras direções para as redes *smart grid* baseadas em IoT mencionando as possíveis alterações com a introdução e expansão do 5G e 6G e como estes serão fundamentais para garantir o movimento de grandes quantidades de *smart meter data* no futuro.

4.2.4 Segurança da energia

No artigo de Arif et al. (2021), intitulado “*Towards Efficient Energy Utilization Using Big Data Analytics in Smart Cities for Electricity Theft Detection*”, os autores começam por definir que perdas de eletricidade podem ser classificadas como perdas técnicas e perdas não técnicas. As perdas técnicas são perdas inevitáveis que ocorrem devido à dissipação de energia nos transformadores, cabos e restantes equipamentos de transmissão. Já as perdas não técnicas ocorrem devido a uso ilegal da eletricidade, erros na faturação, *smart meters* com problemas ou que tenham sido intervencionados por terceiros. No seu artigo abordam a forma como falhas de segurança podem causar estas perdas não técnicas, as quais resultam em perdas financeiras elevadas devido aos roubos de eletricidade, sendo estes roubos normalmente causados pelo uso de energia sem que esta seja faturada aos consumidores. Segundo os autores, estas questões têm um impacto negativo quer para os fornecedores de energia ao nível do recebimento pela prestação dos seus serviços, quer para os restantes consumidores que se situam na zona onde acontece o comportamento fraudulento, já que podem impactar os mesmos ao nível das tarifas de energia devido às necessidades de potência para a zona estarem a ser adulteradas, assim como a qualidade de energia que é fornecida.

Os autores relatam que a expansão da *big data* e utilização de *smart grids* são benéficas na deteção destes roubos energéticos, ao conseguirem garantir uma visão que antes apenas era possível garantir através de inspeções graduais nos próprios locais. A utilização de *smart meters* acrescenta também a possibilidade de aplicação de *machine learning* e modelos de *deep learning*, capazes de detetar fraude energética de uma forma eficiente em comparação com as redes ditas tradicionais. Nestas redes tradicionais, a constante necessidade de efetuar inspeções ao local, numa tentativa de mitigar os roubos, eram difíceis de executar e dispendiosas sendo, em muitos casos, difícil de apontar o local exato onde se está a proceder ao roubo de eletricidade.

Ao nível do artigo, os autores contribuem para o estudo destas áreas com o desenvolvimento de uma nova técnica apoiada nos sistemas de *big data* desenhada para a realização do balanceamento dos dados durante o estado de pré-processamento. No artigo apresentam uma técnica e um método para classificação, capazes de classificar os

consumidores como sendo honestos e fraudulentos: a técnica é denominada *Tomek Link Borderline Synthetic Minority Oversampling Technique* apoiada de uma *Support Vector Machine* desenhada para manter o balanço dos dados durante o estado de pré-processamento. O método desenvolvido pelos autores passa pelo uso de uma *Temporal Convolutional Network* usada para dados sequenciais como o consumo de eletricidade e apoiada uma *Enhanced Multi-layer Perceptron* para os dados não sequenciais auxiliares. Os autores aplicam ainda uma estratégia para redução da variância permitindo assim obter melhores resultados, e terminam, com realização de simulações utilizando diferentes conjuntos de dados.

No que toca às fases do artigo, após esta descrição, procedem à explicação da metodologia apoiada no *related work* por eles agregado, apresentam as simulações realizadas e os seus resultados, terminando o seu artigo com uma discussão sobre os mesmos.

Ao nível dos resultados obtidos pela simulação realizada, os autores apresentam conclusões de que os modelos propostos e utilizados nas simulações permitem obter melhores resultados que os métodos base utilizados para deteção de roubos de eletricidade como investigação aos dados através de *neural networks*, entre outras.

Na conclusão do seu artigo, os autores referem algumas das limitações existentes no próprio modelo, acrescentando que existe uma troca de qualidade de respostas mais precisas dado o custo computacional necessário para a obtenção de melhores resultados mais precisos através da consideração de uma maior quantidade de dados de consumo de eletricidade. No que se refere a investigação futura, propõem a utilização do seu modelo para a exploração de dados climáticos e aplicarem os resultados dessa exploração para, assim, estenderem o modelo desenvolvido, aplicando-o novamente a um cenário de roubos de eletricidade.

No artigo dos autores Javaid et al. (2021), com o título “*An adaptive synthesis to handle imbalanced big data with deep siamese network for electricity theft detection in smart grids*”, os autores discutem a forma como, ao longo do tempo, têm sido desenvolvidas várias técnicas de análise de dados direcionadas de forma a estudar a deteção de roubos de energia elétrica. Ao nível da deteção do comportamento suspeito dos diferentes consumidores este é detetado através da análise dos padrões de consumo dos

consumidores. No entanto, segundo os autores, a aplicação destas técnicas de análise acaba por falhar devido a um uso impróprio das bases de dados e/ou seleção de um classificador apropriado.

Tendo estes princípios em consideração, os autores propõem um método robusto de análise de *big data* para a deteção de roubos de energia elétrica em *smart grids* para melhor diferenciar os consumidores justos dos que cometem fraudes/roubos, usando o consumo dos clientes justos como base para a justificação dos casos de roubo.

A proposta dos autores começa, à semelhança do artigo anterior no pré-processamento dos dados. Aqui os autores adotam o uso de um método de *adaptive synthesis* e integram uma *siamese network* aprofundada para a discriminação das características dos diferentes consumidores para assim atingir um rácio de deteção de roubos de eletricidade elevado. Numa fase final, usam diferentes métricas de performance para apoiar no estudo da *big data* gerada e compreender melhor os resultados obtidos.

Após apresentação da sua solução, os autores iniciam um processo de explicação do problema, análise do mesmo e apresentação da sua solução, testando a mesma através de diferentes simulações e explorando os resultados obtidos.

Segundo os autores, os resultados de uma análise de *time series* permitem observar os padrões de consumo dos diferentes consumidores, onde se verifica que os consumidores justos têm um padrão de consumo simétrico ao longo do tempo, em contraste o dos consumidores fraudulentos é assimétrico.

Na apresentação da sua solução, são exploradas pelos autores as diferentes fases do seu método: pré-processamento de dados, tratamento de dados em falta e harmonização dos diferentes dados recolhidos, tratamento do balanço das classes, extração de recursos através da *convolutional neural network*, estudo das sequências através de *long short term memory*, *machine learning* através da *deep siamese network* e, por fim, recolha de toda a informação, sendo aplicada uma fórmula que permite discriminar/classificar os consumidores.

Para as diferentes simulações, os autores realizam análises a dados recolhidos em tempo real de diferentes *smart meters*. Os resultados obtidos pelos autores nas três simulações

realizadas são favoráveis, tendo os autores optado pela realização de testes de divisão da base de dados utilizada em dois diferentes conjuntos de dados: um para treino e um segundo para teste, sendo divididos os rácios em cenários de treino e teste sendo 60%, 70% e 80% para o treino tendo utilizado os restantes para testes. Os autores terminam o seu artigo apresentando os resultados do seu modelo que consegue manter a *performance* nos três cenários explorados permitindo aos autores justificarem e validarem o modelo por eles apresentado.

No artigo dos autores Ardagna et al., (2021), com o título “*Big Data Analytics-as-a-Service: Bridging the gap between security experts and data scientists*” é explorada a forma como o uso de todo o potencial criado pela *big data* e tecnologias associadas - como *machine learning* e inteligência artificial – se tem tornado essencial para o desenvolvimento das empresas de diferentes dimensões . No entanto, o desenvolvimento de sistemas autónomos de *data analytics* têm sido alvo de ataques informáticos que causam perdas económicas aos negócios. Os autores acreditam que a existência de *experts* na área de segurança é insuficiente para solucionar este problema, sendo necessário aliar os especialistas em segurança aos *data scientists* de modo a garantir o desenho de sistemas que tenham em mente o tratamento de grandes volumes de dados - por vezes sensíveis - e ao mesmo tempo garantir a segurança dos mesmos.

Os autores desenvolvem o seu artigo com o objetivo de diminuir a distância entre as duas áreas e implementar uma proposta baseada numa metodologia e estrutura de um modelo de *Big Data-as-a-Service*, tendo focado o seu artigo no setor energético e nas *smart grids*, onde a monitorização do uso de energia é realizada em tempo real através da dupla comunicação existente entre a rede e os *smart meters* instalados. Segundo os autores, a *smart grid* tem um papel crucial na infraestrutura moderna de energia, no entanto existem dois desafios relacionados com a segurança: 1) a gestão dos equipamentos inteligentes; e 2) a perseveração da segurança e privacidade dos dados de consumo dos utilizadores sendo estes dados críticos para a utilização de *smart grids*, especialmente se forem envolvidos sistemas *cloud*.

Após discussão do problema, os autores iniciam a apresentação da sua solução onde admitem dar a conhecer um novo ciclo de desenvolvimento para a análise de *big data*

com o objetivo de suportar uma comparação mais rápida entre alternativas e aumentar a qualidade e otimização das soluções propostas. A solução descrita aproxima os *Business Analysts*, *Data Scientists*, *Big Data Developers* e *Big Data Engineers* envolvidos no design, execução e manutenção de complexas e contínuas *pipelines* de *big data analytics*. Esta solução é testada em vários cenários de segurança típicos das *smart grids*, onde os *experts* em segurança e analistas de dados têm de colaborar para o desenvolvimento de um processo de análise capaz de detetar incidentes de segurança numa perspetiva que preserve a privacidade dos dados em causa através da análise dos dados registados.

Após descrição da estrutura acima referida, os autores explicam como o seu modelo atua de forma a garantir a privacidade, descrevendo problemas de cibersegurança existentes nas *smart grids*, como a confidencialidade, integridade e disponibilidade e como o modelo apresenta uma solução para estas questões. Nessa descrição, os autores apresentam e explicam três *pipelines* para a garantia de segurança, sendo o primeiro utilizado para manter a anonimização dos dados, o segundo para a deteção de ataques e o terceiro para a deteção de contas esquecidas que podem ser utilizadas para aceder ao sistema.

Terminando a sua descrição, os autores procedem à apresentação da arquitetura e implementação tendo no artigo apresentado resultados visuais para diferentes cenários recolhidos.

Por fim, fazem uma avaliação da sua proposta comparando com outras propostas existentes ao nível das suas competências relacionadas com segurança, desde o processo de design, até quão apropriada esta é para a segurança de dados, o nível de customização, a velocidade a que pode ser implementada, se depende ou não de outros sistemas para a entrega de resultados e se tem a possibilidade de divulgar resultados sobre a forma de dados seguros.

4.2.5 Gestão inteligente da energia

No artigo de Yingfeng Zhang et al., (2018), intitulado “*A big data driven analytical framework for energy-intensive manufacturing industries*”, os autores exploram o uso de *big data analytics* associado à energia introduzindo uma proposta de uma *framework* para

redução do consumo energético e emissões nocivas existentes em indústrias fabris com usos intensivos de energia.

Os autores iniciam o seu artigo, explorando a ideia de que as indústrias de manufatura com uso intensivo de energia geram grandes volumes de dados energéticos através dos equipamentos, processo de produção e gestão de operações, sendo os dados gerados uma mistura de *structured data* sobre a forma de dados de consumo de energia em diversas dimensões (espacial, temporal e energia utilizada), *semi-structured data* (dados enviados e recebidos através das plataformas de gestão inteligente da energia) e *un-structured data* (por exemplo, notificações de email sobre o uso de energia e interações com os clientes sobre o seu uso de energia).

Após apresentação do tema, os autores exploram o funcionamento da *framework* para *big data analytics* aplicada à indústria de manufatura, que dão um uso intensivo à energia. Exploram ideias e formas de captura e perceção da mesma, tecnologias utilizadas para armazenamento e pré processamento desta energia, as técnicas de DM aplicadas centradas nas indústrias de manufatura como *regression*, *classification*, *clustering* e *prediction* e um exemplo de uma *framework* associada à indústria de cerâmica.

Após descrição do funcionamento, fazem o estudo de caso de uma aplicação da *framework* proposta, onde aplicam a mesma à cadeia de produção de uma indústria de manufatura de cerâmica. Ao longo da descrição dessa aplicação, são apresentadas as fases do processo de produção e a ferramenta utilizada que permite a recolha de dados de energia, referindo ainda o processo de DM utilizado aos dados recolhidos no processo.

Na apresentação de resultados da aplicação realizada, os autores discutem a forma como o sistema de gestão da energia desenvolvido permite estimar o consumo real durante o ciclo de produção através dos dados em tempo real da corrente elétrica, juntamente com os dados da voltagem utilizada. Ao nível dos resultados, o sistema implementado permitiu aos utilizadores descobrirem diferentes pontos no ciclo onde é possível poupar energia, reduzir custos de eletricidade e quais as peças do processo de produção que necessitam de reparação. Conseguiram, ainda, obter resultados ao nível das áreas mais eficientes em todo o processo, tendo obtido vários parâmetros que, sujeitos a análise e aplicados aos restantes elementos produtivos, levaram a uma redução de consumo energético e custos.

Segundo os autores, o uso deste sistema associado a *big data analytics* permitiu analisar e monitorizar os dados de várias fontes e dados heterogéneos gerados durante todo o processo. Através da *framework* utilizada, os autores reportam ter descoberto problemas de desperdício de energia e que, através da *energy big data analytics*, conseguiram compreender que apenas uma pequena falha em alguns elementos do processo produtivo pode ter impacto severo no consumo de energia. Num ponto de vista mais numérico e como resumo de todo o processo de análise da gestão energética, mencionam que a implementação de uma *framework* baseada em *big data analytics* para a gestão da energia resultou numa redução em 3% ao fim de um mês do consumo de energia para produção de uma unidade de produto e uma redução de 4% do total de gastos energéticos após 6 meses de uso.

Após apresentação dos resultados, os autores discutem os mesmos referindo-se à eficiência da *framework* proposta numa visão teórica e prática, assim como, as limitações existentes encontradas. Nas conclusões, resumem as contribuições da sua proposta em quatro distintas. A primeira passa pela justificação para o uso de uma *big data driven analytical framework* e a importância dos seus componentes chave. Anotam o segundo contributo da sua proposta como sendo a arquitetura utilizada pelos próprios ao nível da perceção e aquisição de *energy big data*. A terceira contribuição refere-se à estrutura desenvolvida para DM, para filtro de dados relevantes e padrões entre múltiplas fontes de dados e dados heterogéneos. Por fim, relatam a importância da sua contribuição para os diferentes departamentos responsáveis conseguirem tomar decisões mais conscientes e garantir um uso mais eficiente da energia, mais contenção de emissões nocivas e conservação de energia apontando estas como sendo os objetivos chave das *big data driven analytical frameworks* associadas à energia.

No artigo dos autores Marinakis, Doukas, Tsapelas, & Mouzakitidis, (2020), com o título “*From big data to smart energy services: An application for intelligent energy management*”, os autores focam-se na *energy big data* e nas diferentes formas em que esta pode existir, desde os dados dos *smart meters*, aos dados de outras fontes como: clima, gestão de ativos, certificados energéticos, auditoria a edifícios, planos desenvolvidos para a ação contra a crise climática e adoção de fontes de energia mais

sustentáveis, dados socioeconómicos, características dos utilizadores finais e níveis de conforto.

Como já referido, estes dados dão às empresas a possibilidade de extrair valor dos mesmos, associando aos dados ferramentas de gestão de energia e análise de dados. Segundo os autores, as tecnologias de informação prestam o maior contributo para o potencial verificado na *energy big data* sendo estas tecnologias o principal integrador da infraestrutura e soluções de gestão.

Os autores argumentam que os *Energy Management Systems* se encontram cada vez mais difundidos e antecipam o crescimento dos mesmos durante os próximos anos, fruto da integração de tecnologia para análise de dados nos modelos de gestão existentes. Tendo em conta esta informação, os autores apresentam uma proposta para uma arquitetura de *big data* de nível avançado capaz de suportar a criação, desenvolvimento, manutenção e exploração dos serviços de energia inteligente através da utilização de dados de vários domínios. O projeto apresentado pelos autores permite, ainda, a integração de dados heterogéneos, incluindo dados de sensores, dados públicos e dados históricos de diferentes fontes entre outros tipos de dados relacionados como dados ambientais, dados financeiros (preços de energia atuais e históricos), comportamento dos consumidores e dados demográficos.

Para além do desenvolvimento da plataforma, os autores desenvolvem ainda um sistema de suporte à decisão, em linha com as direções utilizadas para o desenvolvimento da plataforma, tendo como objetivo o uso deste para utilizar a energia de diferentes fontes aplicando-a a um contexto de *smart city* para a criação de planos de ação para a gestão energética. Segundo os autores, este sistema pode ser utilizado pelos gestores de energia e autoridades das cidades para gestão da performance energética dos seus edifícios.

Os autores começam por explicar a abordagem metodológica aplicada ao desenvolvimento da plataforma, tendo a mesma o objetivo de desbloquear a promessa que *big data* traz aos desafios existentes no setor energético, assim como apoiar a transformação de energia sustentável e a sua utilização para alcançar uma sociedade que cada vez mais reduz o uso de energias fósseis. Apontam e explicam os diferentes componentes da arquitetura da plataforma, sendo essas a camada de dados que se dividem

em: 1) interoperacional, capaz de receber dados das diversas fontes e de diferentes tipos; 2) de armazenamento dos dados, que utiliza soluções de armazenamento comprovadas na indústria; 3) a camada de controlo das políticas de acesso aos dados, responsável pelo isolamento dos dados de diferentes fornecedores e permitindo acesso a outras ferramentas consoante a detenção de direitos sobre esses dados ou avançadas permissões de utilização dos mesmos.

Na fase seguinte da arquitetura apresentam os serviços de análise, que auxiliam os beneficiários através da apresentação de análises que contribuem para a descoberta de conhecimento não antes constatado, ou outras informações gerais que ferramentas comuns não consigam prestar. Terminam a descrição dos serviços apresentando o *dashboard* que atua como o ponto de entrada para os utilizadores finais que podem utilizar a plataforma. Mencionam novamente o sistema de suporte à decisão baseado nos dados recolhidos, elaborando sobre as múltiplas fontes de dados a utilizar, o modelo de dados, os planos de ação para gestão energética e a forma como este sistema integra o *dashboard*.

Por fim, aplicam em situação real, onde apresentam o cenário por eles utilizado, as cidades onde o mesmo foi implementado, o processo de instalação e configuração, a implementação dos testes piloto e resultados obtidos. Ao nível da medição destes resultados, fazem um estudo do impacto da sua proposta em comparação com os resultados prévios à sua utilização. No que toca aos valores apresentados pelos autores, estes concluem que o sistema de suporte resultou numa redução significativa do consumo energético, emissões de gases estufa e custo da eletricidade, tendo este último alcançado uma redução de 20% e ao mesmo tempo resultado num aumento em 10% da produção de energia renovável.

Ao nível de perspetivas futuras, os autores terminam o seu artigo comunicando a ideia de no futuro pretenderem fazer uma ligação entre o sistema de suporte à decisão com o conceito de moeda energética, onde apresentam a ideia de os utilizadores finais conseguirem gerar moedas digitais através da redução ou moderação do seu consumo energético.

4.3 Discussão dos conteúdos explorados

Com a análise realizada aos artigos explorados foi possível verificar a situação atual no que toca à investigação e utilização de *big data analytics* na área da energia. A figura 4.1. e tabela 4.1 apresentam um resumo dos temas, tecnologias, autores e principais conclusões identificadas nos artigos analisados na RSL realizada.



Figura 4.1. Temas e tecnologias

Tabela 4.1. Autores e temas dos artigos abordados

| Resumo dos artigos analisados | | | |
|--|---|--|--------------------------|
| Tema | Artigo | Principais Conclusões | Autores |
| Estudo do consumidor de energia | <i>Understanding household energy consumption behavior: The contribution of energy big data analytics</i> | Os padrões de consumo têm um grau de variedade elevado entre diferentes famílias devido aos diferentes fatores intrapessoais, interpessoais e externos. | (Zhou & Yang, 2016) |
| Estudo do consumidor de energia | <i>IoT big data analytics for smart homes with fog and cloud computing</i> | Aplicação pode levar a diferentes tipos de resultados e potenciar diferentes desenvolvimentos consoante o estudo da informação recolhida. | (Yassine et al., 2019) |
| <i>Smart meters e big data analytics</i> | <i>A hybrid ICT-solution for smart meter data analytics</i> | Solução proposta consegue analisar grandes volumes de dados recolhidos durante um período definido, ou fazer análises através de um processo contínuo de carregamento de dados. | (X. Liu & Nielsen, 2016) |
| <i>Smart meters e big data analytics</i> | <i>A big data platform for smart meter data analytics</i> | Resultados obtidos através da plataforma permitem competir com os produtos presentes no mercado. Plataforma descrita como uma alternativa fácil de operacionalizar e capaz de gerir a <i>energy big data</i> , auxiliar na visualização dos dados recolhidos e preparados tendo potencial para providenciar suporte ao nível da tomada de decisão. | (Wilcox et al., 2019) |
| <i>Smart meters e big data analytics</i> | <i>Prediction of voltage distribution using deep learning and identified key smart meter locations</i> | Medições realizadas pelos <i>smart meters</i> dos locais chave e o seu modelo de previsão baseado em <i>deep learning</i> são suficientes para uma previsão eficiente da distribuição de voltagem. Informações dos consumidores individuais não são necessárias para o processo de previsão. | (Mokhtar et al., 2021) |
| <i>Smart Grids</i> | <i>Big data framework for analytics in smart grids</i> | Estrutura por si desenvolvida, incluindo toda a sua configuração, código utilizado e aplicação poderá surtir efeitos positivos no desenvolvimento de outras plataformas de <i>big data</i> , associadas a outras áreas produzindo resultados ao nível dos negócios e promoção do desenvolvimento de ciências e tecnologias que melhorem a tomada de decisão. | (Munshi & Mohamed, 2017) |
| <i>Smart Grids</i> | <i>Smart grid data analytics for increasing energy savings in residential buildings</i> | Sistema capaz de melhor informar sobre o consumo, alertar os consumidores para as melhores horas para o uso de energia e formular estratégias para o seu uso podendo o mesmo ajudar numa utilização mais eficiente da energia, alocando a energia a utilizar | (Chou & Ngo, 2016) |

| Resumo dos artigos analisados | | | |
|-------------------------------|---|--|---|
| Tema | Artigo | Principais Conclusões | Autores |
| Smart Grids | <i>A microgrid energy management system based on chance-constrained stochastic optimization and big data analytics</i> | Referem-se ao uso de tecnologias de <i>big data</i> e ao potencial que estas apresentam na otimização de geração e gestão energética e distribuição em sistemas energéticos coletivos ou em individuais como as <i>microgrids</i> e refletem ainda sobre o papel que <i>big data</i> tem na integração de energias renováveis. | (Marino & Marufuzzaman, 2020) |
| Smart Grids | <i>Data quality of electricity consumption data in a smart grid environment</i> | Assegurar a qualidade no processo de aplicação e partilha de <i>big data analytics</i> de dados elétricos é uma peça fundamental na tomada de decisão sobre os sistemas energéticos. | (W. Chen et al., 2017) |
| Smart Grids | <i>IoT based smart energy grid for sustainable cities</i> | Apresentam uma explicação detalhada de sistemas <i>smart grid</i> , gestão energética e sistemas de <i>smart grid</i> seguros para o desenvolvimento sustentável de <i>smart cities</i> . | (Renugadevi et al., 2021) |
| Segurança da Energia | <i>Towards Efficient Energy Utilization Using Big Data Analytics in Smart Cities for Electricity Theft Detection</i> | Referem algumas das limitações existentes no próprio modelo, acrescentando que existe uma troca de qualidade de respostas mais precisas dado o custo computacional necessário para a obtenção de melhores resultados mais precisos através da consideração de uma maior quantidade de dados de consumo de eletricidade. | (Arif et al., 2021) |
| Segurança da Energia | <i>An adaptive synthesis to handle imbalanced big data with deep siamese network for electricity theft detection in smart grids</i> | Apresentam os resultados do seu modelo que consegue manter a performance nos três cenários explorados permitindo aos autores justificarem e validarem o modelo por eles apresentado | (Javaid et al., 2021) |
| Segurança da Energia | <i>Big Data Analytics-as-a-Service: Bridging the gap between security experts and data scientists</i> | O potencial de <i>machine learning</i> e inteligência artificial é impactado pela dificuldade de desenhar e aplicar uma estrutura capaz de responder às expectativas do utilizador ao nível de precisão e segurança. O seu modelo de desenvolvimento responde a estas preocupações. | (Ardagna et al., 2021) |
| Gestão inteligente da energia | <i>A big data driven analytical framework for energy-intensive manufacturing industries</i> | Apontam que a sua proposta permite obter resultados ao nível de exploração de conhecimento escondido e transformar esse conhecimento em informações úteis para os gestores. | (Yingfeng Zhang et al., 2018) |
| Gestão inteligente da energia | <i>From big data to smart energy services: An application for intelligent energy management</i> | O modelo apresentado é capaz de providenciar novos caminhos na exploração de eficiência energética através da incorporação de novas abordagens e tecnologias. | (Marinakis, Doukas, Tsapelas, & Mouzakitis, 2020) |

Alguns dos artigos explorados na RSL também apresentavam algumas ideias de trabalho futuro a explorar pelos autores dos artigos analisados, sendo estas ideias possíveis observar na tabela 4.3.

Tabela 4.2. Ideias de trabalho futuro presentes nos artigos analisados

| Trabalho Futuro | Autores |
|--|---|
| Mais investigação nesta área, assim como novos desenvolvimentos associados a <i>energy big data analytics</i> e ao estudo de novas oportunidades para compreensão dos comportamentos de consumo energético nas casas dos consumidores. | (Zhou & Yang, 2016) |
| Incorporação de mais algoritmos para apoio na análise de dados, expandir o sistema proposto para suportar outros tipos de <i>smart meter data</i> como água, gás e aquecimento, garantir maior suporte à gestão de redes de distribuição ligadas às <i>smart grids</i> . | (X. Liu & Nielsen, 2016) |
| Comparar a estrutura durante período ditos normais e períodos com disrupções como falhas de equipamentos ou na própria rede. | (Marino & Marufuzzaman, 2020) |
| Investigar futuramente a deteção de <i>outliers</i> em tempo real, assim como, ao nível de deteção dos mesmos quando provenientes de falhas nos equipamentos da rede, erros nos parâmetros e intrusões na rede | (W. Chen et al., 2017) |
| Explorar possíveis alterações com a introdução e expansão do 5G e 6G e como estes serão fundamentais para garantir o movimento de grandes quantidades de <i>smart meter data</i> no futuro | (Renugadevi et al., 2021) |
| Propõem a utilização do seu modelo para a exploração de dados climáticos e aplicarem os resultados dessa exploração e, assim, estenderem o modelo desenvolvido, aplicando-o novamente a um cenário de roubos de eletricidade | (Arif et al., 2021) |
| Ligação do <i>Decision Support System</i> ao conceito de energia como moeda. Futura exploração de uma moeda digital. | (Marinakis, Doukas, Tsapelas, & Mouzakitis, 2020) |

Ao nível das temáticas a primeira temática explorada passa pelo uso de *big data analytics* como instrumento para compreender a utilização da energia por parte dos consumidores. Esta aplicação de *big data*, segundo os artigos explorados, permite alcançar diversos tipos de resultados positivos, tanto para os consumidores como para as próprias empresas que realizam a comercialização da energia: a empresa obtém vantagens por conseguir estabelecer um perfil para os seus clientes, podendo, através desse perfil, aconselhar o utilizador para uma mais correta utilização da energia. Esse uso permite à empresa estabelecer níveis de consumo energético mais concisos, o que gera um menor dispêndio de recursos para manter o cliente e, por sua vez, reajustar estes esforços para melhorar pontos da sua infraestrutura que necessitem de uma maior atenção. Nos artigos explorados denotam-se como relevantes para a construção do perfil do consumidor programas de

demand side management e tecnologias de *fog* e *cloud computing*. Outro ponto relevante do estabelecimento do perfil passa por conseguir obter *leads* uteis para a promoção de outros produtos/serviços que sejam do interesse dos clientes. Por fim, o uso de *big data analytics* permite obter uma noção do uso regular da energia por parte do utilizador, permitindo que, no momento de se verificar um aumento drástico da necessidade energética, seja possível proceder à identificação do problema, enviando um alerta e uma proposta de solução para o cliente. Do lado do cliente, este também beneficia da introdução de *big data analytics*. Nos dias de hoje a informação recolhida sobre o consumo energético é detalhadamente partilhada com o utilizador, pelo que o mesmo conseguirá identificar quais os pontos ou até eletrodomésticos da sua habitação que mais consomem energia. O utilizador também consegue analisar o seu próprio gasto energético em tempo real, conseguindo, assim, tomar consciência disso e desencadear atitudes para um uso mais eficiente da energia, reduzindo o consumo.

Os estudos analisados também permitem discutir o uso dos *smart meters* e o impacto que os mesmos produzem para a gestão de energia via a utilização de ferramentas de análise de dados. Os *smart meters* atuam logo no primeiro passo para exploração de dados, a recolha dos mesmos garantindo uma melhor compreensão do uso energético, explorar possíveis fraudes no uso da rede energética e ainda ser uma peça fundamental na comunicação com os utilizadores, promovendo o envio de *feedback* gerado relacionado ao seu gasto de energia. Os *smart meters* originam, ao mesmo tempo, uma redução de custos ao nível de manutenção dos próprios contadores, assim como aparelhos ligados à rede e um mais eficiente controlo do uso de energia. A instalação destes contadores possibilitou o desenvolvimento de plataformas que, através dos dados recolhidos, parametrizados e devidamente preparados, tornam possível a visualização do funcionamento da rede energética para estudo da mesma e preparação de intervenções para um uso mais eficiente da energia e tomada de decisões sobre como abordar o consumidor, decisões ao nível do estado da rede energética, deteção de situações de roubo ou melhoria da gestão. Ao nível das ferramentas mencionadas nos artigos verifica-se a aplicação de um variado leque de ferramentas, desde as *deep learning neural networks*, *Hadoop*, *Spark*, *Hive*, *MADlib* e o *OpenDSS*. Por fim, os *smart meters* são importantes

para a redução do uso de energia, sendo um contributo fundamental para o estabelecimento de *low voltage networks* - redes de potência reduzida - que conseguem manter o seu funcionamento normal devido à possibilidade de realização de uma análise que é possível ser feita via os dados recolhidos com os *smart meters*.

A implementação dos *smart meters* tem servido como um dos geradores de grandes volumes de dados, assim como tem melhorado a análise dos dados recolhidos permitindo obter resultados não possíveis através dos contadores tradicionais, ao permitir obter leituras mais precisas e dados de consumo energético em tempo real. O seu uso é, no entanto, apenas parte do processo de *big data analytics* na energia, sendo nas *smart grids* que se verificam efetivamente as melhorias ao nível de monitorização e controlo de energia. Em relação às *smart grids*, o desenvolvimento das mesmas é o exemplo mais claro da combinação de *big data analytics* com o setor energético, sendo o mesmo possível através do desenvolvimento da IoT e a sua crescente integração nas TICs. Nas aplicações exploradas verifica-se que as plataformas desenvolvidas para análise à *smart grid* são o ponto de encontro de todos os dados recolhidos da rede via os *smart meters* instalados ou outros componentes inteligentes que desempenham um papel ativo na rede. Nos estudos analisados é possível verificar algumas das ferramentas associadas à *smart grid* que permitem alcançar os objetivos que se pretende atingir com a análise de dados como o *Apache Spark*, *Hadoop*, *Tableau* para visualizações, e ainda os *decision support systems*.

É também fundamental destacar o papel que as *smart grids* têm na transição para o uso de energias renováveis e resposta às necessidades energéticas existentes na rede, através de um melhor controlo, visualização e análise da energia. Os estudos explorados permitem também verificar que os sistemas assentes na *smart grid* não têm de, necessariamente, cobrir toda a rede, sendo possível existirem sistemas mais concentrados denominados *microgrids* que permitem uma análise detalhada de uma parte mais concisa da rede energética. Por fim, o benefício associado à utilização de *smart grids* que aparenta ser mais relevante passa pela possível sinergia de diversas formas de energia na própria *smart grid* sendo esta característica similar a uma característica natural de *big data* revelando a ligação entre as duas áreas. O crescente interesse no desenvolvimento de

projetos para *smart cities* é também um ponto extremamente relevante para a importância das *smart grids* dado estas serem fundamentais para o desenvolvimento de *smart cities*.

O desafio mais gritante na aplicação das *smart grids* aparenta ser, para além da infraestrutura necessária para a sua execução, o tratamento dos dados gerados e distribuídos. Sendo necessário ter sempre em consideração a qualidade dos mesmos e minimizar ao máximo a existência de falhas na análise seja pela existência de *outliers*, dados incompletos ou dados redundantes que possam comprometer análises realizadas e enviar as informações que os dados nos permitem obter.

Para além da realidade das *smart grids* a análise dos resultados também demonstra como, no que respeita ao uso de *big data analytics* na energia, também se levantam algumas preocupações ao nível da segurança de energia, dados e privacidade dos envolvidos no sistema. Nos artigos explorados abordam-se situações de roubos de energia que afetam os fornecedores numa perspetiva da sua distribuição não estar a ser contabilizada, não surtindo assim ganhos monetários para os próprios o que por sua vez afeta também os restantes clientes que podem sofrer com a sua existência vendo as tarifas por eles pagas alteradas, assim como, a própria qualidade da rede de energia estar a sofrer uma adulteração pondo em causa não só o uso da mesma mas também a vida útil de todos os equipamentos a ela ligados. Ao nível das ferramentas introduzidas nos artigos estudados que tocam nesta matéria verifica-se o uso de *deep learning*, *machine learning*, *adaptive synthesis e neural networks*. O impacto que *big data analytics* tem neste cenário é considerável dado que a existência de IENs permite uma mais fácil deteção deste tipo de comportamentos e por sua vez uma mais fácil resposta a estes problemas, assim como, o estabelecimento de protocolos a serem executados de modo a não prejudicar os restantes utilizadores caso venham a surgir estas situações. Na eventualidade do próprio sistema colocado em prática falhar na deteção das situações de roubo de energia, o apoio em ferramentas de análises de dados também permite uma análise mais robusta aos consumos realizados pelos utilizadores permitindo mais facilmente a sua deteção. Para este tipo de situações são aplicados métodos de análise de padrões de consumo, algo que apenas se torna possível através da existência de um perfil de consumo do próprio consumidor,

sendo possível definir esse perfil através da junção das áreas de *big data analytics* e energia.

Ao nível dos novos desenvolvimentos verificados com a junção das áreas *big data analytics* e energia, a maior preocupação levantada passa pela garantia da segurança dos dados privados dos consumidores, pelo que a própria infraestrutura e desenvolvimento das *smart grids* deve ter estabelecido este princípio como algo prioritário a garantir. Dada a natureza dos dados em tratamento, verifica-se como essencial a colaboração entre diferentes departamentos de modo que no desenvolvimento de uma *smart grid* e utilização de *smart meters* seja possível garantir que os dados a serem recolhidos não possam ser facilmente acessíveis, comprometendo informações privadas dos consumidores.

Ao nível da gestão da energia e a forma como esta pode ser potencializada através de *big data analytics*, foram analisadas duas perspetivas, uma perspetiva focada no seu uso industrial e uma perspetiva da gestão de energia geral. Na perspetiva industrial, verifica-se que, através da gestão de energia inteligente, torna-se possível obter uma melhor compreensão dos gastos energéticos durante as diferentes fases do processo industrial, possibilitando a realização de uma análise do consumo nas diferentes fases e, através dessa análise, coordenar estratégias que permitam uma poupança de custos de energia durante as diferentes fases do processo. Para a gestão inteligente da energia, nos artigos estudados ferramentas como o *NoSQL* e *Hadoop* foram mencionadas. A instalação de uma IEN garante a possibilidade de melhor compreender os momentos em que é necessário proceder à manutenção das diferentes ferramentas do processo ao permitir analisar através de *big data analysis* quais as partes do processo de fabrico que não se encontram a produzir aos níveis adequados. Ao mesmo tempo esta informação também permite fazer estimativas de quando será necessário substituir certos elementos visto o seu uso contínuo. Para a perspetiva mais geral da gestão inteligente da energia é possível entender que os sistemas de gestão de energia estão em crescimento, a adoção de dados de diferentes formas como dados climatéricos, entre outros, tem revolucionado a forma como a gestão é abordada, assim como, as decisões tomadas ao nível da energia. A integração de *big data analytics* e os resultados ao nível do desenvolvimento de novos

sistemas e plataformas têm dado asas à exploração de dados históricos, públicos que permitem a formulação de planos de gestão que contêm informações adicionais resultando na apresentação de uma “realidade” mais completa e informada que por sua vez permitirá a formulação de planos de ação mais exequíveis não só para questões de uma gestão consciente da energia, num olhar mais financeiro da mesma, mas também ao nível da crescente integração de energias renováveis que são impulsionadas através de desenvolvimentos como as *smart grids*, entre outros, fruto desta ligação entre *big data analytics* com a energia e a exploração no sentido de alcançar um uso mais consciente e eficiente da mesma.

5 CONCLUSÃO

A realização desta dissertação, sobre a forma de uma RSL, teve como objetivo responder à questão-chave, “Como tem sido aplicado *big data analytics* na área da energia?”. Este trabalho contribui para um entendimento das alterações verificadas através da realização de uma revisão sistemática da investigação no domínio de *big data analytics* na área de energia, procurando sistematizar a investigação relevante realizada entre os anos de 2016 a 2022, procedendo à categorização dos estudos descobertos com a pesquisa realizada e permitindo que trabalhos futuros possam ser desenvolvidos. Para começar procedeu-se a um enquadramento teórico de modo a categorizar os temas discutidos abordando-se as temáticas de *big data analytics* e as TICs, o seu desenvolvimento e ligação à área da energia.

Após enquadramento teórico, procedeu-se ao enquadramento metodológico onde se define a tipologia de estudo realizado como sendo uma RSL onde foi utilizada uma adaptação da *checklist* PRISMA para controlo do processo. De seguida são apresentados o protocolo e as suas fases, a estratégia de pesquisa, exploração de base de dados, filtro e avaliação de publicações, recolha dos dados, interpretação de resultados, terminando com as conclusões. A estratégia aplicada consistiu no estabelecimento de critérios de inclusão e exclusão, após apresentação dos critérios definiu-se a ScienceDirect como sendo a base de dados utilizada na RSL. Para o processo de seleção foi aplicada uma frase booleana para pesquisa de artigos e realizada uma exploração bibliométrica através do uso da ferramenta VOSViewer. De seguida procedeu-se à filtração de resultados através de uma comparação dos artigos encontrados com os critérios definidos, leitura dos artigos obtidos e realização de um teste de relevância para o tema explorado resultando num total de 15 artigos. O enquadramento metodológico terminou com a definição de considerações a ter no que toca à informação existente dentro dos 15 artigos onde se define que será realizada a extração de informações como o título, autor, data de publicação, objetivo, considerações, metodologia, resultados e por fim as observações finais e limitações apontadas se relevantes.

Após definição da metodologia a seguir, iniciou-se o trabalho empírico, onde se começa por sumarizar as temáticas exploradas dos estudos selecionados, de seguida realiza-se a

análise de 15 artigos, cuja agregação de informação dá lugar à resposta à questão de investigação colocada. A conclusão do trabalho empírico é realizada através da discussão dos conteúdos explorados durante a análise dos artigos obtidos sendo exploradas as temáticas verificadas nos 15 artigos analisados referindo o impacto de *big data analytics* em cada área explorada e os seus efeitos para o setor de energia e os seus intervenientes, a agregação de todo o trabalho realizado produz uma resposta à questão-chave colocada.

5.1 Contributos

Os contributos desta RSL baseiam-se principalmente na identificação e sistematização de como *big data analytics* tem sido aplicada à energia respondendo à questão-chave colocada, a indicação da evolução da investigação existente na área e identificação de oportunidades e desafios desenvolvidos com a junção das duas áreas.

A revisão contribuí para a investigação da junção dos temas *big data analytics* e energia com a análise e conclusões que são possíveis retirar com a exploração bibliométrica que permitiu concluir que a pesquisa que envolve o uso de *big data analytics* na energia se encontra em constante mudança, verificando-se através da exploração bibliométrica os avanços existentes na área ao longo dos anos, os quais são refletidos nas palavras-chave utilizadas nos artigos produzidos dentro da área de investigação. Nestes verifica-se que, é frequente encontrarem-se certas palavras-chave que apenas são utilizadas durante um curto período de tempo, entre 1 e 3 anos, dado que, devido aos avanços na área, novas/outras palavras começam a ser utilizadas. Esta RSL contribui também com uma categorização dos estudos descobertos com a pesquisa realizada organizando os mesmos por temáticas exploradas que servem de base ao desenvolvimento de trabalhos futuros que envolvam ambas as áreas *big data analytics* e energia. Na RSL foram identificadas as ferramentas e tecnologias utilizadas nos artigos explorados, verificando-se que o uso de certas ferramentas/tecnologias é comum entre os diferentes temas que fazem parte da análise de resultados obtidos, assim como, foi ainda apresentada duas tabelas onde é possível observar as principais conclusões existentes nos artigos e outra com o trabalho futuro apontado pelos autores nos seus artigos.

Ao nível das descobertas realizadas com a análise dos resultados, podem ser retiradas várias conclusões no que toca à forma como a aplicação de *big data analytics* interagiu com a área da energia e os contributos que *big data analytics* produz para a área.

Para começar, o uso de *big data analytics* na energia impacta tanto os fornecedores de energia como os consumidores, colocando os consumidores numa situação mais benéfica no que toca aos gastos em energia e os fornecedores quanto ao controlo dos gastos e previsões do fornecimento. Isto deve-se muito à possibilidade de, através de *big data analytics* conseguir traçar melhor o perfil do consumidor, permitindo aos fornecedores adotarem estratégias que respondem de forma mais eficiente às necessidades existentes no mercado, assim como explorar outras oportunidades como serviços de apoio técnico, agregação de *leads* e venda de novos produtos e serviços. O uso de *big data analytics* permite ainda às empresas do setor da energia ter um controlo mais apertado da sua produção e dos custos incorridos na própria produção, assim como na distribuição. Este controlo reflete-se numa tomada de decisão mais informada e um maior controlo sobre o estado dos equipamentos utilizados na rede. Ao nível dos utilizadores, estes conseguem, manter-se mais informados dos seus consumos energéticos e consultar estratégias para garantirem que conseguem retirar uma maior eficiência não só do uso de energia, mas também dos equipamentos que utilizam nas suas casas.

A pesquisa realizada leva-nos ainda a concluir que, à semelhança de outras indústrias, o setor energético sofreu grandes mudanças a nível da produção, distribuição e utilização da energia. Estas mudanças tiveram a sua origem na associação de técnicas, processos e sistemas planeados para análise de grandes volumes de dados (como podemos observar na análise dos resultados obtidos e discussão dos conteúdos explorados no trabalho empírico desta RSL), sendo estes possíveis de obter devido às crescentes melhorias tecnológicas verificadas nos últimos anos como por exemplo: a crescente incorporação de *smart meters*, utilização de equipamentos *smart* e o desenvolvimento de *smart grids* e *microgrids*. A associação de equipamentos *smart* que servem como a base de recolha de dados para a atividade de *big data analysis* também influenciam a segurança da energia, por exemplo, ao servirem para deteção remota de fraude existente na rede, levando a uma poupança de custos face aos métodos de redes tradicionais que antes envolviam o

deslocamento de forças humanas aos locais de atividade suspeita. Por fim, a própria gestão da energia foi também influenciada pela introdução de equipamentos *smart* e atividade de *big data analysis* permitindo um controlo mais apertado da mesma, resultando em redução de custos durante o ciclo energético e integração facilitada de energias de fonte renovável.

5.2 Limitações

Ao nível de limitações encontradas na produção desta dissertação, sentiu-se uma limitação em relação ao número de conectores a considerar para a construção da frase booleana utilizada na pesquisa, dado que a base de dados selecionada estabelece o limite de 8 conectores por pesquisa realizada. Visto que os artigos considerados se baseiam nas palavras-chave selecionadas, o número de artigos e perspectivas a explorar na área torna-se mais limitado se o seu universo for mais reduzido. A exclusão de estudos verificada tendo em conta este ponto pode influenciar a profundidade da revisão, pelo que é possível que durante o processo de seleção e filtro de estudos, a limitação de conectores que por sua vez influencia o número de palavras-chave para a frase booleana tenha levado a que não tenham sido considerados estudos que permitam uma resposta à questão-chave com um detalhe diferente ao apresentado nesta RSL.

5.3 Trabalho Futuro

No que respeita a desenvolvimento de trabalhos futuros, espera-se que os conteúdos discutidos nesta revisão deem lugar à exploração destes temas numa perspectiva mais profunda por parte de quem já investiga a ligação entre as duas áreas discutidas, assim como, inspirar “novos caminhos” na exploração destas temáticas. Numa perspectiva mais pessoal, um dos desenvolvimentos futuros pode vir a compreender estudos adicionais que considerem o contributo que a aplicação de *big data analytics* pode trazer ao estudo das alterações climáticas e como o conhecimento recolhido desta área afeta a tomada de decisões no que toca à produção de energia, solar, hídrica e eólica. A exploração dos impactos das alterações climáticas nesta área juntamente com a forma como *big data analytics* pode auxiliar na formulação de melhores estratégias de mitigação de perdas e formas de integração das energias renováveis para além da sua exploração através das

smart grids. Outros temas verificados na exploração de artigos desta revisão compreendem áreas como o desenvolvimento de moedas energéticas e o uso da *blockchain* que correspondem a áreas a considerar dado o crescente interesse nestas tecnologias, pelo que gostaria de verificar mais trabalhos que relacionem estas áreas de rápido crescimento que se encontram em constante mudança com estes novos temas emergentes.

REFERÊNCIAS BIBLIOGRÁFICAS

- Ahmed, E., Yaqoob, I., Hashem, I. A. T., Khan, I., Ahmed, A. I. A., Imran, M., & Vasilakos, A. V. (2017). The role of big data analytics in Internet of Things. *Computer Networks*, *129*, 459–471. <https://doi.org/10.1016/j.comnet.2017.06.013>
- Akhavan-Hejazi, H., & Mohsenian-Rad, H. (2018). Power systems big data analytics: An assessment of paradigm shift barriers and prospects. *Energy Reports*, *4*, 91–100. <https://doi.org/10.1016/j.egyr.2017.11.002>
- Ardagna, C. A., Bellandi, V., Damiani, E., Bezzi, M., & Hebert, C. (2021). Big Data Analytics-as-a-Service: Bridging the gap between security experts and data scientists. *Computers and Electrical Engineering*, *93*(May), 107215. <https://doi.org/10.1016/j.compeleceng.2021.107215>
- Arif, A., Alghamdi, T. A., Khan, Z. A., & Javaid, N. (2021). Towards Efficient Energy Utilization Using Big Data Analytics in Smart Cities for Electricity Theft Detection. *Big Data Research*, *27*, 100285. <https://doi.org/10.1016/j.bdr.2021.100285>
- Aromataris, E., & Pearson, A. (2014). The systematic review: An overview. *American Journal of Nursing*, *114*(3), 53–58. <https://doi.org/10.1097/01.NAJ.0000444496.24228.2c>
- Asikis, T., & Pournaras, E. (2020). Optimization of privacy-utility trade-offs under informational self-determination. *Future Generation Computer Systems*, *109*, 488–499. <https://doi.org/10.1016/j.future.2018.07.018>
- Baker, T., Asim, M., Tawfik, H., Aldawsari, B., & Buyya, R. (2017). An energy-aware service composition algorithm for multiple cloud-based IoT applications. *Journal of Network and Computer Applications*, *89*(August 2016), 96–108. <https://doi.org/10.1016/j.jnca.2017.03.008>
- Bansal, S. K. (2014). Towards a Semantic Extract-Transform-Load (ETL) framework for big data integration. *Proceedings - 2014 IEEE International Congress on Big Data, BigData Congress 2014*, 522–529. <https://doi.org/10.1109/BigData.Congress.2014.82>

- Bhattacharai, B. P., Paudyal, S., Luo, Y., Mohanpurkar, M., Cheung, K., Tonkoski, R., Hovsopian, R., Myers, K. S., Zhang, R., Zhao, P., Manic, M., Zhang, S., & Zhang, X. (2019). Big data analytics in smart grids: State-of-the-art, challenges, opportunities, and future directions. *IET Smart Grid*, 2(2), 141–154. <https://doi.org/10.1049/iet-stg.2018.0261>
- Bramer, W. M., de Jonge, G. B., Rethlefsen, M. L., Mast, F., & Kleijnen, J. (2018). A systematic approach to searching: An efficient and complete method to develop literature searches. *Journal of the Medical Library Association*, 106(4), 531–541. <https://doi.org/10.5195/jmla.2018.283>
- Bui, N., Castellani, A. P., Casari, P., & Zorzi, M. (2012). The internet of energy: A web-enabled smart grid system. *IEEE Network*, 26(4), 39–45. <https://doi.org/10.1109/MNET.2012.6246751>
- Chen, W., Zhou, K., Yang, S., & Wu, C. (2017). Data quality of electricity consumption data in a smart grid environment. *Renewable and Sustainable Energy Reviews*, 75(November 2016), 98–105. <https://doi.org/10.1016/j.rser.2016.10.054>
- Chin, W., Li, W., & Chen, H. (2017). Energy Big Data Security Threats in IoT-Based Smart Grid Communications. *IEEE Communications Magazine*, 55(October), 70–75.
- Chou, J. S., & Ngo, N. T. (2016). Smart grid data analytics framework for increasing energy savings in residential buildings. *Automation in Construction*, 72, 247–257. <https://doi.org/10.1016/j.autcon.2016.01.002>
- Chou, J. S., & Truong, N. S. (2019). Cloud forecasting system for monitoring and alerting of energy use by home appliances. *Applied Energy*, 249(February), 166–177. <https://doi.org/10.1016/j.apenergy.2019.04.063>
- Corizzo, R., Ceci, M., & Malerba, D. (2019). Big Data Analytics and Predictive Modeling Approaches for the Energy Sector. *Proceedings - 2019 IEEE International Congress on Big Data, BigData Congress 2019 - Part of the 2019 IEEE World Congress on Services*, 55–63. <https://doi.org/10.1109/BigDataCongress.2019.00020>

- Dang-Ha, T. H., Olsson, R., & Wang, H. (2015). The role of big data on smart grid transition. *Proceedings - 2015 IEEE International Conference on Smart City, SmartCity 2015, Held Jointly with 8th IEEE International Conference on Social Computing and Networking, SocialCom 2015, 5th IEEE International Conference on Sustainable Computing and Communic*, 33–39. <https://doi.org/10.1109/SmartCity.2015.43>
- Diamantoulakis, P. D., Kapinas, V. M., & Karagiannidis, G. K. (2015). Big Data Analytics for Dynamic Energy Management in Smart Grids. *Big Data Research*, 2(3), 94–101. <https://doi.org/10.1016/j.bdr.2015.03.003>
- Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research*, 133(March), 285–296. <https://doi.org/10.1016/j.jbusres.2021.04.070>
- Duan, L., & Xiong, Y. (2015). Big data analytics and business analytics. *Journal of Management Analytics*, 2(1), 1–21. <https://doi.org/10.1080/23270012.2015.1020891>
- El Khaouat, A., & Benhlima, L. (2017). Big data based management for smart grids. *Proceedings of 2016 International Renewable and Sustainable Energy Conference, IRSEC 2016*, 1044–1047. <https://doi.org/10.1109/IRSEC.2016.7983902>
- Elnakat, A., & Gomez, J. D. (2016). The flame dilemma: A data analytics study of fireplace influence on winter energy consumption at the residential household level. *Energy Reports*, 2, 14–20. <https://doi.org/10.1016/j.egyr.2016.01.002>
- European Commission. (2018). *Big data solutions for energy | Programme | H2020 | CORDIS | European Commission*. https://cordis.europa.eu/programme/id/H2020_DT-ICT-11-2019
- Fan, C., Xiao, F., Li, Z., & Wang, J. (2018). Unsupervised data analytics in mining big building operational data for energy efficiency enhancement: A review. *Energy and Buildings*, 159, 296–308. <https://doi.org/10.1016/j.enbuild.2017.11.008>
- Fan, M. wei, Ao, C. chu, & Wang, X. rong. (2019). Comprehensive method of natural

- gas pipeline efficiency evaluation based on energy and big data analysis. *Energy*, 188, 116069. <https://doi.org/10.1016/j.energy.2019.116069>
- Fawzy, D., Moussa, S., & Badr, N. (2016). The evolution of data mining techniques to big data analytics: An extensive study with application to renewable energy data analytics. *Asian Journal of Applied Sciences*, 4(3), 756–766.
- Ge, M., Bangui, H., & Buhnova, B. (2018). Big Data for Internet of Things: A Survey. *Future Generation Computer Systems*, 87, 601–614. <https://doi.org/10.1016/j.future.2018.04.053>
- Ghorbanian, M., Dolatabadi, S. H., & Siano, P. (2019). Big Data Issues in Smart Grids: A Survey. *IEEE Systems Journal*, 13(4), 4158–4168. <https://doi.org/10.1109/JSYST.2019.2931879>
- Hariri, R. H., Fredericks, E. M., & Bowers, K. M. (2019). Uncertainty in big data analytics: survey, opportunities, and challenges. *Journal of Big Data*, 6(1). <https://doi.org/10.1186/s40537-019-0206-3>
- Harris, J. D., Quatman, C. E., Manring, M. M., Siston, R. A., & Flanigan, D. C. (2014). How to write a systematic review. *American Journal of Sports Medicine*, 42(11), 2761–2768. <https://doi.org/10.1177/0363546513497567>
- Hu, J., & Vasilakos, A. V. (2016). Energy Big Data Analytics and Security: Challenges and Opportunities. *IEEE Transactions on Smart Grid*, 7(5), 2423–2436. <https://doi.org/10.1109/TSG.2016.2563461>
- Jaradat, M., Jarrah, M., Bousselham, A., Jararweh, Y., & Al-Ayyoub, M. (2015). The internet of energy: Smart sensor networks and big data management for smart grid. *Procedia Computer Science*, 56(1), 592–597. <https://doi.org/10.1016/j.procs.2015.07.250>
- Javaid, N., Jan, N., & Javed, M. U. (2021). An adaptive synthesis to handle imbalanced big data with deep siamese network for electricity theft detection in smart grids. *Journal of Parallel and Distributed Computing*, 153, 44–52. <https://doi.org/10.1016/j.jpdc.2021.03.002>

- Jiang, H., Wang, K., Wang, Y., Gao, M., & Zhang, Y. (2016). Energy big data: A survey. *IEEE Access*, 4, 3844–3861. <https://doi.org/10.1109/ACCESS.2016.2580581>
- Jin, X., Wah, B. W., Cheng, X., & Wang, Y. (2015). Significance and Challenges of Big Data Research. *Big Data Research*, 2(2), 59–64. <https://doi.org/10.1016/j.bdr.2015.01.006>
- Junaidi, N., & Shaaban, M. (2018). Big Data Applications in Electric Energy Systems. *2018 International Conference on Computational Approach in Smart Systems Design and Applications, ICASSDA 2018*, 1–5. <https://doi.org/10.1109/ICASSDA.2018.8477607>
- Kafle, Y. R., Mahmud, K., Morsalin, S., & G. E., T. (2015). Towards an Internet of Energy. *2016 IEEE International Conference on Power System Technology (POWERCON), Sustainabl.*
- Kambatla, K., Kollias, G., Kumar, V., & Grama, A. (2014). Trends in big data analytics. *Journal of Parallel and Distributed Computing*, 74(7), 2561–2573. <https://doi.org/10.1016/j.jpdc.2014.01.003>
- Ku, T. Y., Park, W. K., & Choi, H. (2017). IoT energy management platform for microgrid. *2017 IEEE 7th International Conference on Power and Energy Systems, ICPEs 2017, 2017-Decem*, 106–110. <https://doi.org/10.1109/ICPEsYS.2017.8215930>
- Kulatunga, C., Bhargava, K., Vimalajeewa, D., & Ivanov, S. (2017). Cooperative in-network computation in energy harvesting device clouds. *Sustainable Computing: Informatics and Systems*, 16, 106–116. <https://doi.org/10.1016/j.suscom.2017.10.006>
- Kumar, O., & Goyal, A. (2016). Visualization: A novel approach for big data analytics. *Proceedings - 2016 2nd International Conference on Computational Intelligence and Communication Technology, CICT 2016*, 121–124. <https://doi.org/10.1109/CICT.2016.32>
- Le Ray, G., & Pinson, P. (2020). The ethical smart grid: Enabling a fruitful and long-

- lasting relationship between utilities and customers. *Energy Policy*, 140(March), 111258. <https://doi.org/10.1016/j.enpol.2020.111258>
- Li, T., Tang, J., Jiang, F., Xu, X., Li, C., & Ding, T. (2019). Research on Storage and Processing Method for Renewable Energy Big Data. *2019 IEEE PES Innovative Smart Grid Technologies Asia, ISGT 2019*, 3680–3684. <https://doi.org/10.1109/ISGT-Asia.2019.8881713>
- Liu, G., Yang, J., Hao, Y., & Zhang, Y. (2018). Big data-informed energy efficiency assessment of China industry sectors based on K-means clustering. *Journal of Cleaner Production*, 183, 304–314. <https://doi.org/10.1016/j.jclepro.2018.02.129>
- Liu, W., Zhao, J., & Wang, D. (2021). Data mining for energy systems: Review and prospect. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 11(4), 1–18. <https://doi.org/10.1002/widm.1406>
- Liu, X., & Nielsen, P. S. (2016). A hybrid ICT-solution for smart meter data analytics. *Energy*, 115, 1710–1722. <https://doi.org/10.1016/j.energy.2016.05.068>
- Louhghalam, A., Akbarian, M., & Ulm, F. J. (2017). Carbon management of infrastructure performance: Integrated big data analytics and pavement-vehicle-interactions. *Journal of Cleaner Production*, 142, 956–964. <https://doi.org/10.1016/j.jclepro.2016.06.198>
- Lv, Z., & Singh, A. K. (2021). Big Data Analysis of Internet of Things System. *ACM Transactions on Internet Technology*, 21(2), 1–15. <https://doi.org/10.1145/3389250>
- Lyu, W., & Liu, J. (2021). Artificial Intelligence and emerging digital technologies in the energy sector. *Applied Energy*, 303(February), 117615. <https://doi.org/10.1016/j.apenergy.2021.117615>
- Ma, Z., Xie, J., Li, H., Sun, Q., Si, Z., Zhang, J., & Guo, J. (2017). The role of data analysis in the development of intelligent energy networks. *IEEE Network*, 31(5), 88–95. <https://doi.org/10.1109/MNET.2017.1600319>
- Makrani, H. M., Tabatabaei, S., Rafatirad, S., & Homayoun, H. (2018). Understanding the role of memory subsystem on performance and energy-efficiency of Hadoop

- applications. *2017 8th International Green and Sustainable Computing Conference, IGSC 2017, 2017-October*, 1–6. <https://doi.org/10.1109/IGCC.2017.8323591>
- Marinakakis, V., Doukas, H., Tsapelas, J., & Mouzakitis, S. (2020). From big data to smart energy services: An application for intelligent energy management. *Future Generation Computer Systems*, *110*, 572–586. <https://doi.org/10.1016/j.future.2018.04.062>
- Marinakakis, V., Doukas, H., Tsapelas, J., Mouzakitis, S., Sicilia, Á., Madrazo, L., & Sgouridis, S. (2020). From big data to smart energy services: An application for intelligent energy management. *Future Generation Computer Systems*, *110*, 572–586. <https://doi.org/10.1016/j.future.2018.04.062>
- Marino, C. A., & Marufuzzaman, M. (2020). A microgrid energy management system based on chance-constrained stochastic optimization and big data analytics. *Computers and Industrial Engineering*, *143*(February), 106392. <https://doi.org/10.1016/j.cie.2020.106392>
- Mark van Rijmenam. (2014). *Why The 3V's Are Not Sufficient To Describe Big Data*. Dataflog.Com. <https://dataflog.com/read/3vs-sufficient-describe-big-data/166>
- Marlen, A., Maxim, A., Ukaegbu, I. A., & Kumar Nunna, H. S. V. S. (2019). Application of Big Data in Smart Grids: Energy Analytics. *International Conference on Advanced Communication Technology, ICACT, 2019-Febru*, 402–407. <https://doi.org/10.23919/ICACT.2019.8701973>
- Mokhtar, M., Robu, V., Flynn, D., Higgins, C., Whyte, J., Loughran, C., & Fulton, F. (2021). Prediction of voltage distribution using deep learning and identified key smart meter locations. *Energy and AI*, *6*(April), 100103. <https://doi.org/10.1016/j.egyai.2021.100103>
- Motlagh, N. H., Mohammadrezaei, M., Hunt, J., & Zakeri, B. (2020). Internet of things (IoT) and the energy sector. *Energies*, *13*(2), 1–27. <https://doi.org/10.3390/en13020494>
- Mukherjee, S., Shaw, R., Haldar, N., & Changdar, S. (2015). A Survey of Data Mining

- Applications and Techniques. *International Journal of Computer Science and Information Technologies*, 6(5), 4663–4666.
- Munshi, A. A., & Mohamed, Y. A. R. I. (2017). Big data framework for analytics in smart grids. *Electric Power Systems Research*, 151, 369–380. <https://doi.org/10.1016/j.epsr.2017.06.006>
- Nadeem, A., & Arshad, N. (2021). A data-driven approach to reduce electricity theft in developing countries. *Utilities Policy*, 73(October), 101304. <https://doi.org/10.1016/j.jup.2021.101304>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *International Journal of Surgery*, 88(March). <https://doi.org/10.1016/j.ijisu.2021.105906>
- PRISMA*. (2020). <https://prisma-statement.org/>
- Rao, P., Muller, M. R., & Gunn, G. (2017). Conducting a metering assessment to identify submetering needs at a manufacturing facility. *CIRP Journal of Manufacturing Science and Technology*, 18, 107–114. <https://doi.org/10.1016/j.cirpj.2016.10.005>
- Renugadevi, N., Saravanan, S., & Naga Sudha, C. M. (2021). IoT based smart energy grid for sustainable cites. *Materials Today: Proceedings*, xxx. <https://doi.org/10.1016/j.matpr.2021.02.270>
- Schuelke-Leech, B. A., Barry, B., Muratori, M., & Yurkovich, B. J. (2015). Big Data issues and opportunities for electric utilities. *Renewable and Sustainable Energy Reviews*, 52, 937–947. <https://doi.org/10.1016/j.rser.2015.07.128>
- Sestino, A., Prete, M. I., Piper, L., & Guido, G. (2020). Internet of Things and Big Data as enablers for business digitalization strategies. *Technovation*, 98(July), 102173. <https://doi.org/10.1016/j.technovation.2020.102173>
- Sharmila, P., Baskaran, J., Nayanatara, C., & Maheswari, R. (2019). A hybrid technique

- of machine learning and data analytics for optimized distribution of renewable energy resources targeting smart energy management. *Procedia Computer Science*, 165, 278–284. <https://doi.org/10.1016/j.procs.2020.01.076>
- Sheela, S., & Sathesh Kumar, K. (2020). Privacy –Preserved in cloud based data sharing for energy management system using big data analytics. *Materials Today: Proceedings*, xxxx. <https://doi.org/10.1016/j.matpr.2020.09.621>
- Shiny, S. S. G., Sathya Priya, S., & Murugan, K. (2021). Repeated game theory-based reducer selection strategy for energy management in SDWSN. *Computer Networks*, 193(April), 108094. <https://doi.org/10.1016/j.comnet.2021.108094>
- Shobol, A., Ali, M. H., Wadi, M., & Tur, M. R. (2019). Overview of big data in smart grid. *8th International Conference on Renewable Energy Research and Applications, ICRERA 2019*, 1022–1025. <https://doi.org/10.1109/ICRERA47325.2019.8996527>
- Siddaway, A. P., Wood, A. M., & Hedges, L. V. (2018). How to Do a Systematic Review: A Best Practice Guide for Conducting and Reporting Narrative Reviews, Meta-Analyses, and Meta-Syntheses. *Annual Review of Psychology*, 70(1), 747–770.
- Siddiqua, A., Karim, A., & Gani, A. (2017). Big data storage technologies: a survey. *Frontiers of Information Technology and Electronic Engineering*, 18(8), 1040–1070. <https://doi.org/10.1631/FITEE.1500441>
- Silva, B. N., Khan, M., & Han, K. (2020). Integration of Big Data analytics embedded smart city architecture with RESTful web of things for efficient service provision and energy management. *Future Generation Computer Systems*, 107, 975–987. <https://doi.org/10.1016/j.future.2017.06.024>
- Statista. (2022). *Total data volume worldwide 2010-2025* | . <https://www.statista.com/statistics/871513/worldwide-data-created/>
- Sun, Z., & Huo, Y. (2021). The Spectrum of Big Data Analytics. *Journal of Computer Information Systems*, 61(2), 154–162. <https://doi.org/10.1080/08874417.2019.1571456>

- Syed, D., Zainab, A., Ghrayeb, A., Refaat, S. S., Abu-Rub, H., & Bouhali, O. (2021). Smart Grid Big Data Analytics: Survey of Technologies, Techniques, and Applications. *IEEE Access*, 9, 59564–59585. <https://doi.org/10.1109/ACCESS.2020.3041178>
- Syed, D., Zainab, A., Refaat, S. S., Abu-Rub, H., & Bouhali, O. (2020). Smart Grid Big Data Analytics: Survey of Technologies, Techniques, and Applications. *IEEE Access*, 1–1. <https://doi.org/10.1109/access.2020.3041178>
- Tsai, C. W., Lai, C. F., Chao, H. C., & Vasilakos, A. V. (2015). Big data analytics: a survey. *Journal of Big Data*, 2(1), 1–32. <https://doi.org/10.1186/s40537-015-0030-3>
- ur Rehman, M. H., Yaqoob, I., Salah, K., Imran, M., Jayaraman, P. P., & Perera, C. (2019). The role of big data analytics in industrial Internet of Things. *Future Generation Computer Systems*, 99, 247–259. <https://doi.org/10.1016/j.future.2019.04.020>
- Walch, A., Castello, R., Mohajeri, N., & Scartezzini, J. L. (2020). Big data mining for the estimation of hourly rooftop photovoltaic potential and its uncertainty. *Applied Energy*, 262(November 2019), 114404. <https://doi.org/10.1016/j.apenergy.2019.114404>
- Wang, K., Li, H., Feng, Y., & Tian, G. (2017). Big Data Analytics for System Stability Evaluation Strategy in the Energy Internet. *IEEE Transactions on Industrial Informatics*, 13(4), 1969–1978. <https://doi.org/10.1109/TII.2017.2692775>
- Wilcox, T., Jin, N., Flach, P., & Thumim, J. (2019). A Big Data platform for smart meter data analytics. *Computers in Industry*, 105, 250–259. <https://doi.org/10.1016/j.compind.2018.12.010>
- Wu, J., Guo, S., Li, J., & Zeng, D. (2016). Big Data Meet Green Challenges: Greening Big Data. *IEEE Systems Journal*, 10(3), 873–887. <https://doi.org/10.1109/JSYST.2016.2550538>
- Wu, S., Li, F., Mehrotra, S., & Ooi, B. C. (2011). Query optimization for massively

- parallel data processing. *Proceedings of the 2nd ACM Symposium on Cloud Computing, SOCC 2011*. <https://doi.org/10.1145/2038916.2038928>
- Yaqoob, I., Hashem, I. A. T., Gani, A., Mokhtar, S., Ahmed, E., Anuar, N. B., & Vasilakos, A. V. (2016). Big data: From beginning to future. *International Journal of Information Management*, 36(6), 1231–1247. <https://doi.org/10.1016/j.ijinfomgt.2016.07.009>
- Yassine, A., Singh, S., Hossain, M. S., & Muhammad, G. (2019). IoT big data analytics for smart homes with fog and cloud computing. *Future Generation Computer Systems*, 91, 563–573. <https://doi.org/10.1016/j.future.2018.08.040>
- Zhang, Yang, Huang, T., & Bompard, E. F. (2018). Big data analytics in smart grids: a review. *Energy Informatics*, 1(1), 1–24. <https://doi.org/10.1186/s42162-018-0007-5>
- Zhang, Yingfeng, Ma, S., Yang, H., Lv, J., & Liu, Y. (2018). A big data driven analytical framework for energy-intensive manufacturing industries. *Journal of Cleaner Production*, 197, 57–72. <https://doi.org/10.1016/j.jclepro.2018.06.170>
- Zhou, K., Fu, C., & Yang, S. (2016). Big data driven smart energy management: From big data to big insights. *Renewable and Sustainable Energy Reviews*, 56(2016), 215–225. <https://doi.org/10.1016/j.rser.2015.11.050>
- Zhou, K., & Yang, S. (2016). Understanding household energy consumption behavior: The contribution of energy big data analytics. *Renewable and Sustainable Energy Reviews*, 56, 810–819. <https://doi.org/10.1016/j.rser.2015.12.001>
- Zhou, K., Yang, S., & Shao, Z. (2016). Energy Internet: The business perspective. *Applied Energy*, 178, 212–222. <https://doi.org/10.1016/j.apenergy.2016.06.052>

APÊNDICES

APÊNDICE 1. CLUSTERS OBTIDOS NA EXPLORAÇÃO BIBLIOMÉTRICA.

Quadro 1. Clusters obtidos na exploração bibliométrica

| | |
|-------------------------|---|
| Cluster 1 (16 items) | artificial neural network; baas; catalysis; complementary operation of multi from energy system; complex event processing; lambda architecture; mapreduce; microgrid; microservice architecture; power systems; renewable energies; spark; spark streaming; stochastic optimization; sustainability; wind power |
| Cluster 2 (14 items) | artificial intelligence; building energy management; decision support system; energy consumption; machine learning; metaheuristic optimization; nature-inspired metaheuristic optimization; pattern prediction; pattern recognition; security and privacy; smart grid data; time-series data; time-series technique; web-based system |
| Cluster 3 (13 items) | cloud network; data sharing; data transformation; differential privacy; diversity; energy management; masking; privacy-preserved; security analysis; software defined wireless network system; trade-off; utility; wireless sensor networks |
| Cluster 4 (13 items) | adaptive decisions; computation offloading; connected devices; cooperative computing; energy harvesting; fog computing; in-network analytics; internet of things; low cost; low-latency applications |
| Cluster 5 (12 items) | analytic methods in power networks; data ethics; data management; deep neural learning; distribution network operation; dynamic demand response; energy transition; privacy-preserving data analysis; smart meters; surveillance capitalism; trust; voltage prediction |
| Cluster 6 (12 items) | big data analytical applications; building energy efficiency; building operational performance; cleaner production; cloud platform; Data Mining; data-driven models; energy-intensive manufacturing industries; gradual pattern mining; interpretable machine learning; knowledge discover; unsupervised Data Mining |
| Cluster 7 (12 items) | addictive manufacturing; energy saving; home appliance; industrial analytics; metering assessment; optimization; small to midsized manufacturing; smart manufacturing; submetering; sustainable manufacturing; time series data analytics; web-based portal |
| Cluster 8 (12 items) | communications efficiency; distributed learning; heterogeneous wireless; iot; multi-cloud; restful architecture; rfid; sensors; service composition; smart cities; smart home; web of things |
| Cluster 9 (11 items) | cloud computing; data center networking; data center networks; datacenter; electrical and optical packet switching; energy and power optimization; interconnection networks; qap; routing; virtual machine; virtual machine placement |

| | |
|--------------------------|--|
| Cluster 10 (10 items) | adaptive synthesis; building cooling load; building energy prediction; c-ran; deep learning; deep Siamese network; electricity theft detection; mobile network; multi-layer perceptron; smart grids; |
| Cluster 11 (9 items) | agent; aggregation; ethics; grouping; network; privacy; sensor; smart city; smart meter |
| Cluster 12 (9 items) | change point analysis; co concentration; energy efficiency; energy efficiency analysis; industrial 4.0; life cycle cost analysis; occupancy-based control; real-time big data analytics; set-point temperature |
| Cluster 13 (9 items) | clustering; demand side management; distributed systems; energy big data; energy informatics; household energy consumption; intervention strategies; optimized distribution; smart energy management |
| Cluster 14 (9 items) | big data analysis; data centers; facility layout problem; factor analysis; hadoop; parallel k-means algorithm |
| Cluster 15 (8 items) | Data quality; electricity consumption; energy; internet of energy; meter data analytics; outlier data; outlier detection; smart grid |
| Cluster 16 (8 items) | certificate authority; extensibility; industrial devops; internet of things; scalability; smart building; smart infrastructure; stream analytics |
| Cluster 17 (7 items) | data analytics; energy conservation; fireplace; ict-solution; residential energy consumption; smart meter data; winter energy efficiency |
| Cluster 18 (7 items) | automated metering infrastructure; big data analytics; electricity theft; greenhouse gas emissions; network analysis; pavement vehicle-interaction; power-law distribution |
| Cluster 19 (6 items) | big data; climate change; green computing; low carbon; sustainable consumption; workflow mapping |
| Cluster 20 (5 items) | building occupancy visualization; building performance assessment; human presence; occupancy extraction; visual analytics |
| Cluster 21 (5 items) | efficiency evaluation; energy aspect; equivalent fuel gas; natural pipeline; volumetric work |
| Cluster 22 (4 items) | analytics; numerical weather analysis; power system simulation; smarter energy research |
| Cluster 23 (4 items) | fuzzy logic; multi-agent systems; smart data; wireless sensor network |

| | |
|-------------------------|---|
| Cluster 24 (4 items) | blockchain; digitalization; ontology; urban energy system |
| Cluster 25 (3 items) | cnc machining; energy efficient manufacturing; intelligent immune mechanism |
| Cluster 26 (3 items) | demand response; pattern mining; time series |

ANEXOS

ANEXO 1. CHECKLIST PRISMA

| Section and Topic | Item # | Checklist item | Location where item is reported |
|-------------------------------|--------|--|---------------------------------|
| TITLE | | | |
| Title | 1 | Identify the report as a systematic review. | Título |
| ABSTRACT | | | |
| Abstract | 2 | See the PRISMA 2020 for Abstracts checklist. | Abstract |
| INTRODUCTION | | | |
| Rationale | 3 | Describe the rationale for the review in the context of existing knowledge. | 1.1 |
| Objectives | 4 | Provide an explicit statement of the objective(s) or question(s) the review addresses. | 1.2 |
| METHODS | | | |
| Eligibility criteria | 5 | Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses. | 3.2.1 |
| Information sources | 6 | Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted. | 3.2.2 / 3.2.3 |
| Search strategy | 7 | Present the full search strategies for all databases, registers and websites, including any filters and limits used. | 3.2.1 |
| Selection process | 8 | Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process. | 3.2.3 |
| Data collection process | 9 | Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process. | 3.2.4 |
| Data items | 10a | List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect. | 3.2.3 |
| | 10b | List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information. | 3.2.3 |
| Study risk of bias assessment | 11 | Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process. | 3.2.3 |
| Effect measures | 12 | Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results. | - |
| Synthesis | 13a | Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study | 4.1 |

| Section and Topic | Item # | Checklist item | Location where item is reported |
|-------------------------------|--------|--|---------------------------------|
| methods | | intervention characteristics and comparing against the planned groups for each synthesis (item #5)). | |
| | 13b | Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions. | 3.2.3.1 |
| | 13c | Describe any methods used to tabulate or visually display results of individual studies and syntheses. | 3.2.3.1 |
| | 13d | Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used. | 3.2.3.2 |
| | 13e | Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression). | 3.2.3.2 |
| | 13f | Describe any sensitivity analyses conducted to assess robustness of the synthesized results. | 3.2.3.2 |
| Reporting bias assessment | 14 | Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases). | 3.2.4 |
| Certainty assessment | 15 | Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome. | 3.2.2 |
| RESULTS | | | |
| Study selection | 16a | Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram. | 3.2.3 |
| | 16b | Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded. | 3.2.3 |
| Study characteristics | 17 | Cite each included study and present its characteristics. | 4.2 |
| Risk of bias in studies | 18 | Present assessments of risk of bias for each included study. | 3.2.3.2 |
| Results of individual studies | 19 | For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots. | 4.2 |
| Results of syntheses | 20a | For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies. | 4.2 |
| | 20b | Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect. | - |
| | 20c | Present results of all investigations of possible causes of heterogeneity among study results. | 4.2 |
| | 20d | Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results. | - |
| Reporting biases | 21 | Present assessments of risk of bias due to missing results | |

| Section and Topic | Item # | Checklist item | Location where item is reported |
|--|--------|--|---------------------------------|
| | | (arising from reporting biases) for each synthesis assessed. | |
| Certainty of evidence | 22 | Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed. | 3.2.4 / 4.1 |
| DISCUSSION | | | |
| Discussion | 23a | Provide a general interpretation of the results in the context of other evidence. | 4.3 / 5. |
| | 23b | Discuss any limitations of the evidence included in the review. | - |
| | 23c | Discuss any limitations of the review processes used. | 5.2 |
| | 23d | Discuss implications of the results for practice, policy, and future research. | 5.3 |
| OTHER INFORMATION | | | |
| Registration and protocol | 24a | Provide registration information for the review, including register name and registration number, or state that the review was not registered. | - |
| | 24b | Indicate where the review protocol can be accessed, or state that a protocol was not prepared. | - |
| | 24c | Describe and explain any amendments to information provided at registration or in the protocol. | - |
| Support | 25 | Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review. | - |
| Competing interests | 26 | Declare any competing interests of review authors. | - |
| Availability of data, code and other materials | 27 | Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review. | 3.2.2 |

ANEXO 2. CHECKLIST PRISMA (PORTUGUÊS)

| Secção e Tópico | Item # | Verificação do Item | Local onde o item está |
|--|--------|---|------------------------|
| TÍTULO | | | |
| Título | 1 | Identifica a publicação como uma revisão sistemática | Título |
| RESUMO | | | |
| Resumo | 2 | Ver a lista de verificação PRISMA 2020 para Resumos. | Resumo |
| INTRODUÇÃO | | | |
| Fundamentação | 3 | Fundamenta a revisão no contexto do conhecimento existente. | 1.1 |
| Objetivos | 4 | Apresenta explicitamente o(s) objetivo(s) ou questão(ões) respeitantes à revisão | 1.2 |
| MÉTODOS | | | |
| Crítérios de elegibilidade | 5 | Especifica os critérios de inclusão e exclusão para a revisão e forma como os estudos foram agrupados para as sínteses. | 3.2.1 |
| Fontes de Informação | 6 | Especifica todas as bases de dados, registos, websites, organizações, listas de referências e outras fontes pesquisadas ou consultadas para identificação dos estudos. Especifica a última data em que cada fonte foi pesquisada ou consultada. | 3.2.2 / 3.2.3 |
| Estratégia de pesquisa | 7 | Apresenta as estratégias de pesquisa completas para todas as bases de dados, registos e websites, incluindo todos os filtros e limites utilizados. | 3.2.1 |
| Processo de seleção | 8 | Especifica os métodos utilizados para decidir se um estudo satisfaz os critérios de inclusão da revisão, incluindo quantos revisores fizeram a triagem de cada registo e publicação selecionada, se trabalharam de uma forma independente e, se aplicável, os detalhes de ferramentas de automatização utilizadas no processo. | 3.2.3 |
| Processo de recolha de dados | 9 | Especifica os métodos utilizados para recolha de dados das publicações, incluindo quantos revisores recolheram a informação de cada publicação, se trabalharam de uma forma independente, todos os processos de obtenção ou confirmação de dados por parte dos investigadores do estudo e, se aplicável, detalhes de ferramentas de automatização utilizadas. | 3.2.4 |
| Dados dos itens | 10a | Lista e define todos os resultados para os quais os dados foram pesquisados. Especifica se foram pesquisados todos os resultados compatíveis com cada domínio em cada estudo (p ex. para todas as medidas, momentos, análises) e, se não, especifica os métodos utilizados para decidir quais resultados a recolher. | 3.2.3 |
| | 10b | Lista e define todas as outras variáveis para as quais os dados foram pesquisados (p. ex. características dos participantes e intervenções, fontes de financiamento). Descreve os pressupostos utilizados sobre informação em falta ou pouco clara. | 3.2.3 |
| Avaliação do risco de viés nos estudos | 11 | Especifica os métodos utilizados para avaliar o risco de viés dos estudos incluídos, incluindo detalhes sobre o(s) instrumento(s) utilizado(s), quantos revisores avaliaram cada estudo e se trabalharam de forma independente e ainda, se aplicável, detalhes de ferramentas de | 3.2.3 |

| Secção e Tópico | Item # | Verificação do Item | Local onde o item está |
|------------------------------------|--------|--|------------------------|
| | | automatização utilizadas no processo. | |
| Medidas de efeito | 12 | Especifica para cada resultado a(s) medida(s) de efeito (p. ex. risco relativo e diferença de média) utilizada(s) na síntese ou apresentação dos resultados. | - |
| Método de síntese | 13a | Descreve os processos utilizados para decidir os estudos elegíveis para cada síntese (p. ex. apresentar as características da intervenção apresentada no estudo e comparar com os grupos planeados para cada síntese (item #5)). | 4.1 |
| | 13b | Descreve todos os métodos necessários de preparação de dados para apresentação ou síntese, tais como lidar com os dados em falta no resumo da estatística, ou conversões de dados. | 3.2.3.1 |
| | 13c | Descreve todos os métodos utilizados para apresentar ou exibir os resultados individuais de estudos e sínteses. | 3.2.3.1 |
| | 13d | Descreve todos os métodos utilizados para resumir os resultados e fornece uma justificação para a(s) escolha(s). Se foi realizada uma meta-análise, descreve o(s) modelo(s) e método(s) para identificar a presença e extensão da heterogeneidade estatística, e de software utilizado(s). | 3.2.3.2 |
| | 13e | Descreve todos os métodos utilizados para explorar possíveis causas de heterogeneidade entre os resultados do estudo (p. ex. análise de subgrupos, meta-regressão). | 3.2.3.2 |
| | 13f | Descreve todas as análises de sensibilidade realizadas para avaliar a robustez a síntese dos resultados. | 3.2.3.2 |
| Avaliação do viés reportado | 14 | Descreve todos os métodos utilizados para avaliar o risco de viés devido à falta de resultados numa síntese (decorrente de viés de informação). | 3.2.4 |
| Avaliação do grau de confiança | 15 | Descreve todos os métodos utilizados para avaliar a certeza (ou confiança) no corpo de evidência de um resultado. | 3.2.2 |
| RESULTADOS | | | |
| Seleção dos estudos | 16a | Descreve os resultados do processo de pesquisa e seleção, desde o número de registos identificados na pesquisa até ao número de estudos incluídos na revisão, idealmente utilizando um fluxograma. | 3.2.3 |
| | 16b | Cita estudos que parecem satisfazer os critérios de inclusão, mas que foram excluídos, e explica as razões da exclusão. | 3.2.3 |
| Características dos estudos | 17 | Cita cada estudo incluído e apresenta as suas características. | 4.2 |
| Risco de viés nos estudos | 18 | Apresenta a avaliação de risco de viés para cada estudo incluído. | 3.2.3.2 |
| Resultados individuais dos estudos | 19 | Para todos os resultados de cada estudo, apresenta: (a) resumo da estatística para cada grupo (quando apropriado) e (b) uma estimativa do efeito e a sua precisão (p. ex. intervalo de confiança/credibilidade), utilizando idealmente tabelas ou gráficos estruturados. | 4.2 |
| Resultados das sínteses | 20a | Para cada síntese, resumo das características e risco de viés entre os estudos selecionados. | 4.2 |

| Secção e Tópico | Item # | Verificação do Item | Local onde o item está |
|---|--------|--|------------------------|
| | 20b | Apresenta os resultados de todas as sínteses estatísticas realizadas. Se foi feita uma meta-análise, apresenta para cada resultado o resumo da estimativa e a sua precisão (p. ex. intervalo de confiança/credibilidade) e medidas de heterogeneidade estatística. Se forem comparados grupos, descreve a direção do efeito. | - |
| | 20c | Apresenta os resultados de todas as investigações de possíveis causas de heterogeneidade entre os resultados do estudo. | 4.2 |
| | 20d | Apresenta resultados de todas as análises de sensibilidade realizadas para avaliar a robustez dos resultados sintetizados. | - |
| Vieses reportados | 21 | Apresenta a avaliação do risco de viés devido à falta de resultados (resultantes de viés de informação) para cada síntese avaliada. | |
| Nível de significância | 22 | Apresenta a avaliação de certeza (ou confiança) no corpo de evidência para cada resultado avaliado. | 3.2.4 / 4.1 |
| DISCUSSÃO | | | |
| Discussão | 23a | Fornecer uma interpretação geral dos resultados no contexto de outra evidência. | 4.3 / 5. |
| | 23b | Discute todas as limitações da evidência, incluídas na revisão. | - |
| | 23c | Discute todas as limitações dos processos de revisão utilizados. | 5.2 |
| | 23d | Discute as implicações dos resultados para a prática, política e investigação futura. | 5.3 |
| OUTRAS INFORMAÇÕES | | | |
| Registo do protocolo | 24a | Fornecer informação sobre o registo da revisão, incluindo o nome e número de registo, ou refere que a revisão não está registada. | - |
| | 24b | Indica local de acesso ao protocolo da revisão, ou refere que o protocolo não foi preparado. | - |
| | 24c | Descreve e explica todas as alterações à informação fornecida no registo ou no protocolo. | - |
| Apoios | 25 | Descreve as fontes de financiamento ou apoio sem financiamento que suportam a revisão, e o papel dos financiadores ou patrocinadores da revisão. | - |
| Conflito de interesses | 26 | Declara todos os conflitos de interesses dos autores da revisão. | - |
| Disponibilidade dos dados, códigos e outros materiais | 27 | Reporta quais dos seguintes materiais estão acessíveis publicamente e onde podem ser encontrados: modelo de formulários de recolha de dados extraídos dos estudos incluídos, dados utilizados para análise; código analítico, qualquer outro material utilizado na revisão. | 3.2.2 |