Electric Power Systems Research 218 (2023) 109193

Contents lists available at ScienceDirect

Electric Power Systems Research

journal homepage: www.elsevier.com/locate/epsr

Evolution of knowledge mining from data in power systems: The Big Data Analytics breakthrough☆

Xavier Dominguez [a,*], Alvaro Prado [b], Pablo Arboleya [a], Vladimir Terzija [c]

[a] Department of Electrical Engineering, University of Oviedo, Gijón, Spain
[b] Departamento de Ingeniería de Sistemas y Computación (DISC), Universidad Católica del Norte (UCN), Antofagasta, Chile
[c] Center for Energy Science and Technology, Skolkovo Institute of Science and Technology, Moscow, Russia

ARTICLE INFO

Keywords:
Artificial Intelligence
Big data
Data analytics
Data mining
Distribution networks
Power systems

ABSTRACT

This paper presents an overview of the evolution of knowledge extraction from power systems data since 1980's up to date. As the existing literature in this application domain is vast and has exponentially grown over the last years, this work remarks the key relevant milestones and contributions that may allow readers to concisely capture the foundations and evolution on which the modern Big Data Analytics (BDA) framework is deployed in this field. Here, it is covered from the first Artificial Intelligent solutions that relied on rule-based expert systems, passing through the usage of Data Mining techniques and arriving to the BDA unfolding, including its trends and prospects in the power industry. Due to exponential increase of metering and communication infrastructure deployment, as well as the impact of distributed generation, one of the stages in the power delivery process that has been particularly revolutionized by the BDA breakthrough is the distribution sector. For this context, the latest and most noteworthy perspectives and experiences are also addresses to highlight the relevance and applicability of knowledge extraction from big data in future power distribution networks.

1. Introduction

Since the 1980's, electrical power systems are witnessing an exceptional increase in complexity and volume of their monitored data [1]. Such an increase has been caused by the continuous advancements of the information and communication technologies [2] and further deepened by several factors such as market deregulation, incorporation of renewable energy sources and the extended insertion of SCADAs and WAMs [2,3]. To manage the information overload within this context, Data Mining (DM) has been exhibited as a knowledge discovery conception based on convenient and versatile approaches to perform effective data processing, find patterns, meet correlations and make knowledge inference. By virtue of these capabilities, several applications in different areas of the power systems field have arisen within the DM framework. Moreover, in the last years the pursue for Smart Grids [4], Advanced Metering Infrastructure (AMI) [5], novel sensing and ICT technology [1], smart cities [6], the Electric Power Internet of Things (EPIoT) [7], electrified transportation [8], energy blockchain [9] among others, have significantly challenged the paradigms of the power industry regarding the knowledge extraction from continuously growing databases.

To cope with the current power systems requirements (e.g. optimal integration of sustainable energy sources, maintaining high levels of the system security, reliability and flexibility, efficient and economical power delivery, robust approaches for the system monitoring, protection and control, integration of different energy vectors), the latest research has migrated from the former perspective of manual-data handling towards an automated data management approach using the Big Data Analytics (BDA) paradigm [6]. Indeed, BDA arises as a promising framework that merges data science fields such as Big Data, Data Analytics and Artificial Intelligence (AI) (see Fig. 1) in order to attain real benefits from massive and unstructured yet information-rich data blocks. However, some challenges and affairs also arise due to the data volume and information criticality. For instance, the BDA implementation concerns are related with data privacy and security, limited real-time processing technology, inaccessible or constrained-quality data collection, lack of standards, policy and insufficient well-trained human resources.

Given the relevance of extracting representative information from data in power systems, hundreds of activities and incentives, using

☆ This work was partially supported by the Spanish Ministry of Economy and Competitivity under grant MCI-20-PID2019-111051RB-I00 and by the Departamento de Ingeniería de Sistemas y Computación from the Universidad Católica del Norte (Chile) under project 202203010029-VRIDT-UCN.

* Corresponding author.
E-mail addresses: uo233585@uniovi.es (X. Dominguez), alvaro.prado@ucn.cl (A. Prado), arboleyapablo@uniovi.es (P. Arboleya), V.Terzija@skoltech.ru (V. Terzija).

https://doi.org/10.1016/j.epsr.2023.109193
Received 19 October 2021; Received in revised form 25 July 2022; Accepted 5 February 2023
Available online 13 February 2023

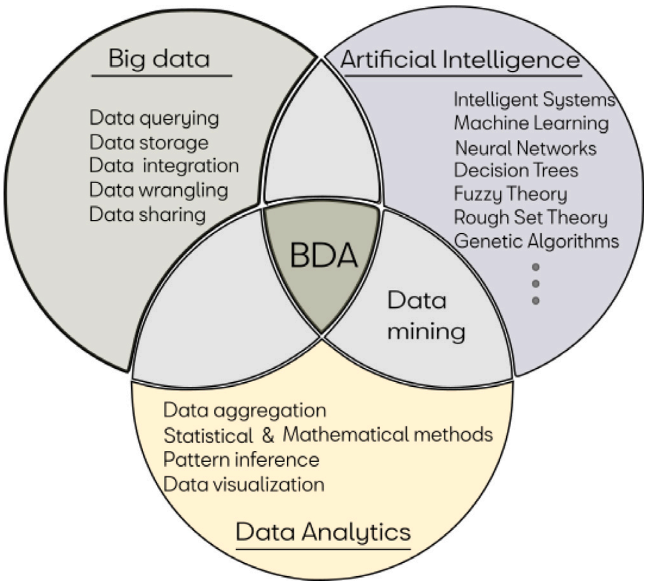

Fig. 1. Big Data Analytics Framework.

DM techniques, have been presented since the arrival of the digital era to propose advancements in the whole electricity chain process. Over the last few years, the amount of scientific projects regarding the integration of BDA in power systems can be overwhelming for a reader who is interested in promptly having a solid understanding of the evolutionary context and relevance of knowledge extraction from data in this field. Some review papers have been published related to BDA in power systems. Most of them focus on particular areas such as smart grids [10–12], distribution networks [13], smart meters data [14], household energy consumption [15] and electrical utilities [16]. Those works are mainly devoted to a specific research sphere and their time-span interest focuses only in the last years where BDA has become mainstream. However, none of them provide readers an encompassing understanding of the development of knowledge mining from data in the broad power systems spectrum from a historical perspective, since the early beginnings of the data digitization phenomena in the 1980's up to the ongoing big data explosion scenario. In addition, the explanations are backed with a curated selection of renowned highly cited publications. By doing all the previous, we aim to make it easier for readers to approach this intricate topic and facilitate the concise understanding of the context on which the modern BDA stands in this field, so that the implications of the current trends are understood in an integral manner and hence, the need towards more data analytics-driven decisions in the power industry is greatly apprehended by researchers and energy stakeholders.

With this background, this work aims to favor the comprehension of the up-growth, applications, current challenges, lessons learned and prospects of knowledge extraction from big data in power systems. Considering the aforesaid and focusing on a systematic literature review, the organization of this paper is as follows. The DM beginnings and its main application areas in power systems are exhibited in Section 2 and Section 3 respectively. The development of BDA in this field is described in Section 4. In Section 5, focusing on power distribution networks, the relevance and latest research on knowledge extraction from big data are exposed. Lessons learned, challenges and prospects are discussed in Section 6. Finally, the conclusions of this work are summarized in Section 7.

2. Beginnings of data mining in power systems

As a consequence of the significant computational improvements experienced during the 1980's, electrical facilities and power systems SCADA experienced a substantial growing digitization [2], deriving into the generation of medium and large-scale data banks. Simultaneously, the new computation capabilities permitted acceptable yet time-consuming offline [17] and real-time [18] simulations. The abundance of monitored and simulated data moved forward the exploration of data-driven Artificial Intelligence (AI) techniques to power systems applications, where the use of classical numerical methods were not able to fully meet requirements. For instance, in the course of that decade, none of the classical time-domain and direct strategies were capable to properly assess the online transient stability analysis of power systems [19] as a consequence of the intrinsic nonlinear modeling complexity.

In this context in the 1980's, one of the first extended techniques of AI in the power electrical field comprised the rule-based expert systems [20]. Their algorithms relied on Decision Trees (DT) [21] based on the "if-then" type rules that were derived from knowledge inference mechanisms fed from the expertise, practices and criteria from specialists, operators and field engineers [22]. Indeed, the use of this tactic in power systems was natural as most operating procedures had specific logical criteria and were developed following a flowchart scheme having unique responses for a set of given conditions [23]. As a consequence, numerous industrial application proposals and high academic interest were exhibited following this approach [24]. However, expert systems also presented some issues such as difficulty to handle unconsidered scenarios, costly maintenance and laboriousness to construct a reliable knowledge base [25]. Given these limitations and the improved computing possibilities of the early 1990's, research focused on more elaborated AI strategies such as Artificial Neural Networks (ANN) [26], fuzzy set systems [27], rough set theory [28], genetic algorithms [29] and Machine Learning (ML) techniques [30]. From this last approaches, genetic algorithms were the most popular provided their ease of use and effective adaptive search process [31]. In contrast to DTs that alone were unable to learn from experience and adapt to new scenarios, the inclusion of ANN and ML overcame this limitation. Thus, the extent of AI techniques in power systems was such, that by 1997, more than four hundred related articles were already traced [23].

In the second half of the 1990s, restructuring of the electrical sector [2] and the integration of renewable energy generation, took place [3,32]. Hence, the size and complexity of electrical infrastructures and databases increased even more, making crucial the refinement of AI approaches to systematize knowledge acquisition. To do so, researchers in power systems embraced the emerging multidisciplinary field known as Data Mining (DM) or also referred as Knowledge Discovery from Data. Foundations and techniques from statistics, database management, pattern recognition and AI were gathered by DM [33] in order to "identify valid, novel, potentially useful and ultimately understandable patterns in data" as one of the first most cited definitions stated [34]. In particular, DM was highly useful in power systems to handle their great scale nature (huge datasets and vast amount of state variables), complex statistical analysis (combining deterministic and stochastic events), mixture of discrete and analog information, variable temporal character (from milliseconds to years), effective visualization requirements, need of rapid decision-making and highly-uncertain data handling [35]. A timeline of the DM unfolding can be seen in Fig. 2.

3. Data mining goals and applications in power systems

First of all, it is important to highlight that mining data is not a lineal process but an iterative one. Besides, there is no matching of specific techniques for a given application as the literature reveals. Indeed, even when a method has been chosen, it is common to recurrently adjust the selected strategy or even use other complementary DM tools to infer different prospects of the data and improve results. Under this perspective, DM was conceived as to achieve three major goals: description, prediction and prescription [36,37]. From the former to

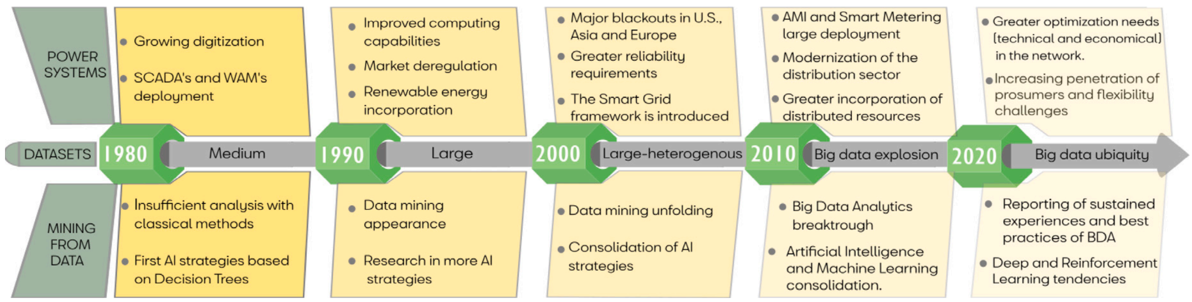

Fig. 2. A timeline of data mining in power systems.

the last one, the added value increases substantially as the two initial stages are focused on understanding what sort of phenomena took place and what causes were underlying so that future can be anticipated, while the latter stage is intended to take all the attained knowledge and transform it into actionable elements to better handle possible fore coming scenarios.

For description, DM techniques characterize the general properties of datasets to attain comprehension of systems by revealing patterns and relationships [38]. Here, segmentation, clustering and dimensionality-reduction tools such as k-means, density-based or Principal Components Analysis (PCA) are normally employed for this purpose with unsupervised manners. Later on, for prediction motives, the mining tools carry out an induction process on data to go beyond classification, regression or estimation and then infer unknown values future variables under study. For instance, in the first decade of this century, advanced prediction approaches were highly requested as a consequence of major blackouts [39] and the smart grid framework introduction [40] which demanded enhanced ancillary services and higher reliability standards. This prediction stage is commonly conducted with ML supervised approaches and nowadays also with cutting-edge Reinforcement Learning and novel ANN mechanisms rebranded as Deep Learning. On the other hand, to perform prescriptive analytics and when needed, wisely influence the future outcomes with suitable course of actions, the previous techniques are leveraged with the inclusion of rich broad-context structured and unstructured Big Data, so that it is ensured that the best possible deductions have been determined and can be applied in a self-healing closed-loop manner. This is not simple or possible to fully automate and therefore can need different degrees of human intervention and validation, specifically for critical duties which is for instance the case of dynamic security assessment in power systems. All the previous perspectives have been employed by the power systems community in a large variety of ambits going from system security and stability, passing through expansion planning, and arriving to monitoring and visualization issues, as exhibited in Table 1.

It is also worth to recall the multidisciplinary nature and wide scope of DM. For this reason, there is no clear boundary on the literature to decide either if DM was applied or not in a particular scientific article. For this reason, this section mainly refers to representative research in power systems where the authors have explicitly mentioned the intended use of DM by means of AI techniques, statistics and other allied disciplines. In this regard, in the last two decades, the DM process [38] and its different methods have been largely employed in the power systems scope. Nevertheless, ANNs with ML purposes have been preferred [41,42] as a consequence of its salient capabilities to learn from training, classify patterns and perform feature extraction [43,44]. After this strategy, DTs, fuzzy systems, statistical analysis (SA) and rough set theory have also been consistently employed in that order [42,45]. To provide the reader an insightful on the relevancy and applications of DM in power systems, Table 1 presents a summary of the most relevant application areas along with some highly cited contributions for each case. It is important to mention that DM solutions have specially succeeded in circumstances where the use of conventional model-driven approaches (based on physics-based methods) have not been able to alone cope with the new over floods of information, the novel technical challenges and impracticalities of previous tactics. In this respect, taking benefit of the newly generated data, engineers have been proposing data-driven solutions to come up with new modeling alternatives that see data as actionable and testable knowledge and not merely as silos of isolated information. Under this trend, mapping rules and relationships are derived by extracting patterns and knowledge from historical measurements or simulations. These data-driven alternatives exhibit high performance and strong decision making rationality [46]. Nonetheless, this last reference also recalls that both, model-driven and data-driven techniques should be conjugated in a systematic manner to complement their benefits and drawbacks in a fruitful way.

Table 1
DM contributions by application area.

Application area	Reference	DM Technique	Citations in Scopus (August/2021)
Dynamic security assessment	[47]	DT	108
	[48]	DT	104
	[49]	DT	38
	[50]	ML, SA	16
Faults detection and classification	[51]	DT	140
	[52]	DT, ML	114
	[53]	ML, SA	91
	[54]	Fuzzy	85
Protection design	[55]	ML	166
	[56]	DT	59
	[57]	DT	51
Load profiling and forecasting	[58]	SA	306
	[59]	DT, SA	96
	[60]	SA, Fuzzy	88
	[61]	SA	38
	[62]	SA, ML	34
Planning	[63]	DT, ML, SA	268
	[64]	Fuzzy	28
	[65]	SA	13
Electricity pricing	[66]	ML, SA	125
	[67]	SA, Fuzzy	1107
	[68]	ML	110
State estimation	[69]	SA, Fuzzy	28
Power quality	[70]	SA	12
	[71]	SA	11
Monitoring and visualization	[72]	SA, ANN, Fuzzy	130
	[47]	DT	108
	[73]	SA, DT	31
	[74]	SA	19

4. The big data analytics unfolding

In the last years, big data topics have become a part of the mainstream research in power systems as a vast amount of scientific papers

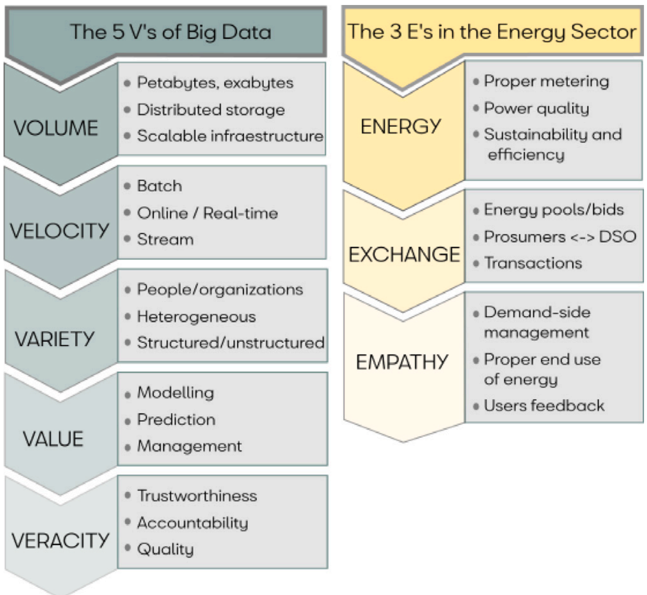

Fig. 3. Big Data characteristics.

in this field demonstrate. This is a result of the ongoing data-intensive era. For instance, in distribution networks, installed smart meters are expected to surpass 1.1 billion by 2022 while a three thousand fold increase in the amount of data is already being faced by utilities which now record energy consumption in a range from 15 min to one hour, contrary to once a month as they did a few years ago [13]. Regarding Transmission System Operators (TSOs), the situation is highly challenging as well as PMUs can take 30 to 60 samples per second. Thus, a single data concentrator gathering data from 100 PMUs (each one collecting 20 measurements at a 30-Hz sample rate) can generate up to 50 GB of data per day [75]. Nevertheless, big data goes far beyond the mere collection of structured and unstructured large quantities of data [10] and rather "describes a new generation of technologies and architectures, designed to economically extract value from very large volumes of a wide variety of data, by enabling high-velocity capture, discovery, and/or analysis" [76]. This popular definition evidences the main big data characteristics summarized in the 5 V's criteria. The first 4 V's were highlighted by the early works on BDA, being these volume, velocity, variety and value; while later research began to include also veracity as a key trait [77]. In the power energy sector other features referred as the 3 E's have also been exhibited as highly relevant, being these energy (that can be saved), exchange (data interchange between energy players to add value to big data) and empathy (increase user satisfaction) [13]. The aforementioned distinctive BDA features are represented in Fig. 3.

To face the inherent challenges of big data management, a complete BDA process (see Fig. 4) must encompass: (i) data collection, (ii) data storage, (iii) pre-processing, (iv) analytics, (v) visualization and (vi) decision making [6,78,79]. Just as DM permitted to extract knowledge from traditional datasets, emerging big data analytics (BDA) frameworks [6,7,11,79] are more than ever expected to attain real benefits from overwhelming but information-abundant datasets coming from different origins. It must be recalled that the overall intention of this mechanisms is to continuously convert raw data into knowledge (actionable useful information) that organizations can utilize to improve their processes and decision-making. In the Big Data era, this represents a significant challenge as traditional algorithms can normally obtain fragmented pieces of knowledge from single sources with uncertainty and different degrees of quality [80]. To overcome this situation, recent data knowledge engineering frameworks have been

Fig. 4. General BDA process.

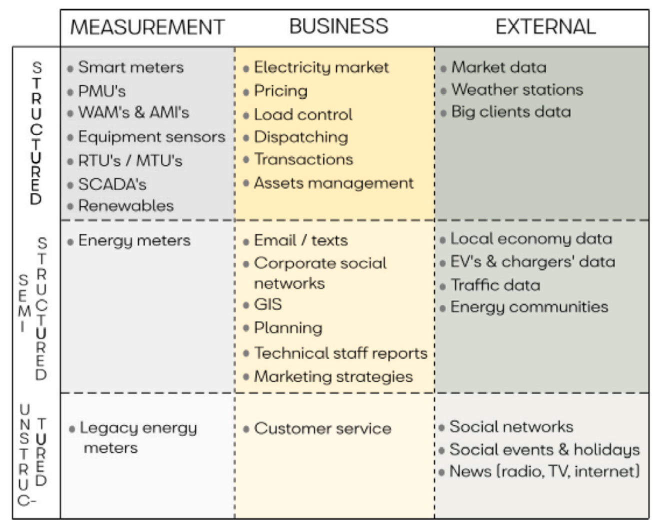

Fig. 5. Power systems data sources.

designed to include nonlinear fusion of fragmented knowledge, online learning from multiple heterogeneous sources and automated demand-driven knowledge navigation [81]. In the context of power systems, the numerous big data sources in modern grids can be categorized as Fig. 5 exhibits, this is considering its data structure [10] (structured, unstructured and semi-structured) and also its origin [82] (measurement data, business data and external data).

Undeniably, the potential advantages and applications of BDA are certainly present in the whole electricity chain process [7,10] ranging from renewable energy planning at the generation level to real-time interaction and energy saving at the demand-side management level. Nonetheless, the areas that have benefited the most are the ones where there is plenty of data, the analytical modeling is highly complex or unfeasible and also in the cases with high system uncertainties. Among these, in a simplified manner we have for instance:

- Faults detection and diagnosis so that the proper corrective measurements can be taken. This is a complex affair as online/offline and on-domain/off-domain data should be simultaneously considered.
- Dynamic security assessment which needs robust systems and near real-time processing as time-decision windows are in the order of seconds for this applications.
- Equipment diagnosis and predictive maintenance to extend the life-span of assets and safeguard the integrity of the system. Sophisticated BDA is needed to infer patterns from large historical data from assets in different technical ambits of the electricity value chain.

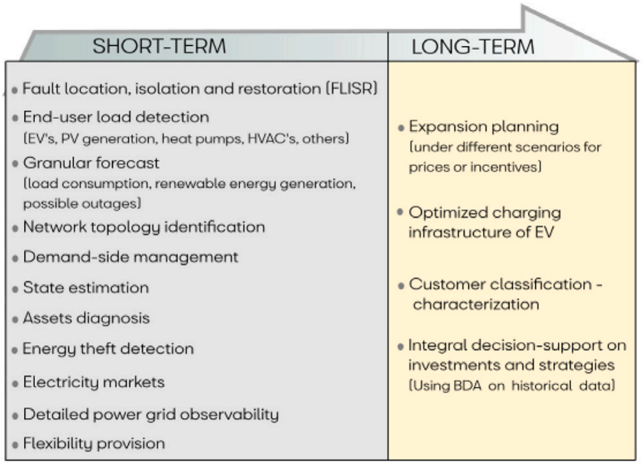

Fig. 6. BDA applications in power distribution systems.

- Power quality monitoring to veil for reliable power grids and to ensure client-satisfaction. Here, high-sampling data can be accessed from specialized monitoring equipment so that specialized BDA can take place.
- Renewable energy forecasting is highly useful to anticipate system fluctuations and feasible production levels.
- Load forecasting and demand flexibility to better organize the power dispatch of the system but also to optimize the distributed energy utilization and grid support with flexibility actions.
- Load disaggregation to attain fine-grained decomposition of clients consumption in order to have greater awareness about their power demand patterns and possible energy efficiency measures.
- Topology identification to curate GIS systems data and the electrical models (e.g. phase identification) and also to automatically determine the customers adoption of novel technologies (EVs, PVs, heat pumps, electronic loads, etc.).
- Energy theft detection to procure energy efficiency in the system and for having greater insights of the non-technical losses.
- Network planning support with the incorporation of BDA to refine and study possible outcomes of incentive schemes or tariffs.
- Smart grid communication systems to interconnect and coordinate the large number of elements in the network in an efficient robust disposition.

It is also noteworthy to remark the prominent role of BDA in distribution networks as next section elaborates. On the other hand, having rooted foundations in DM, BDA takes advantage of a wide variety of AI tools and intelligent processing methods in different power system extents as Ref. [83] exemplifies. Indeed, the application of AI is becoming ubiquitous in power systems as its deployment is not anymore constrained to the forefront of research and large corporations [84], it is now accessible to engineers and practitioners that can take benefit of user-friendly AI and ML tools now included in popular software and free-access libraries. Nonetheless, AI has become a buzz expression that may lead to misunderstandings. For instance, usually AI and Machine Learning (ML) are used interchangeably. However, the latter is only a subset of the possible approaches to held AI. Some other ways to achieve so, are for instance Metaheuristic methods (such as Particle Swarm Optimization (PSO), Genetic Algorithms (GA), rule-based systems (e.g. Fuzzy Logic) and the novel ANN-based Deep Learning tendencies [85].

5. The relevancy of BDA in power distribution networks

Smart energy metering has been experiencing a continuous growth in the last years. For instance, only in the U.S. it was expected to have more than 100 million smart meters in operation by the end of 2020 [86] while at the UK this number was projected to be around 10 million [87]. The smart electric meter market worldwide represented USD 9 billion in 2019 and is expected to expand at over 5% Compound Annual Growth Rate (CAGR) up to 2026 [88]. It is worth to recall that in order to handle the data generated by smart meters, robust communication frameworks and infrastructures are required. To exemplify, Ref. [5] studies the IEEE 123-bus distribution model and states that to properly manage the common 15-minute non-periodic data from 4000 smart meters, five typical base stations are needed to wirelessly cover a geographical area of 2 km^2. On this issue, Ref. [89] proposes a system architecture for smart metering BDA to perform data storing, querying, analysis and visualization; while [90] evaluates different BDA frameworks.

On the other hand, we have been witnessing an increasing incorporation of Distributed Energy Resources (DER) in distribution networks mainly as a consequence of governmental incentives and the significant cost reductions of PV modules as prices have fallen roughly 80% since 2009 [91]. Subsequently, DERs will be able to supply by 2050 up to 45% of electricity needs in countries like Australia [92]. This decentralized energy generation incorporates additional advanced metering demands to achieve successful estimation and control over these systems. Handling the data created by the different metering devices at the distribution level nowadays represents a great challenge that will substantially deepen in the future.

In this overall context, it has been recently reported [93] the unreadiness of Distribution System Operators (DSO) to face these demands as they were mostly conceived as oversized, rigid-structured, conservative, non-automated passive agents [7,16,79] which are now unable to promptly integrate electric vehicle charging and a more active participation of end consumers now becoming energy producers (prosumers). Moreover, to this day modern distribution systems may present insufficient topology information (e.g. phase/transformer-to-client connection) [94], inaccurate state estimation [95] and power flow analysis [96] as well as a highly challenging reactive power control [97]. To progressively address these observability and controllability needs, Advanced Distribution Management Systems (ADMS) have been proposed as the backbone toward distribution grid modernization as they are expected to face the previous issues by holistically handling almost all aspects of a utility (e.g. assets and DER management, supervision and reliability, customer service, field services, transmission operations and planning/regulatory areas). Aside from technology upgrade, the need for collaborative cross-functional teams from the previous utility areas has been highlighted as a key enabler [98]. These teams will usually have to deal with non-standardized systems and architectures and also will be required to implement tailored solutions as every utility has different organizational and technological starting points.

To increase network awareness and permit the functional integration of different subsystems, applications and edge computing; BDA greatly supports ADMS to integrate disperse data models from SCADA, GIS and/or AMI in order to come up with solutions in two main temporal boundaries (see Fig. 6). The first, the short-term, is obviously the most common today as it is related to trending base-line applications such as fault location, isolation and site restoration (FLISR) but also includes actions such as active voltage regulation and direct load/DER control [99]. Other applications in this ambit are non-technical losses detection, end-user load detection, granular forecasting, network topology identification, assets diagnosis and state estimation. As part of the short-term evolution, advanced BDA is expected to support the optimization of feeder operations with the granular coordination of heterogeneous loads and DERs having flexibility provision. This will be possible by means of real-time network models (possibly non-linear) and Optimal Power Flow (OPF) methods. Lastly, in the long-term, BDA will be demanded to assist expansion planning (including EV charging infrastructures), provide integral decision-support and veil for the autonomy and resilience of the system in virtue of all possible hazards (natural disasters, cyber threats, operational issues or centralized-control failures). Within this context, BDA is intended to ultimately assist the optimal utilization of existing equipment and grids. This will lead to reduce expenditures in conventional bulky infrastructures such as centralized generators or substations and favor more digitized, observable and resilient power distribution networks as already pointed out. In this sector, to face most of the case studies, the use of ML has been widespread considering mainly five types of algorithms: (i) unsupervised [100], (ii) supervised [101], (iii) reinforcement learning [102], (iv) imitation learning [103] and (v) generative models [104]. Ultimately, the use of these techniques in different applications derive to lower and more efficient investments, more efficient energy production, transport and dispatch, better use of assets and greater mitigation of outages and contingencies. Further discussion of these applications and their benefits can be found in [13,105]. Lastly, it should be emphasized the relevancy of BDA to procure awareness, efficiency and sustainability at the end-user level; considering not only energy features but also social and behavioral aspects as those found in [15].

6. Lessons learned, challenges and future prospects

Being BDA a relatively novel framework, one of the biggest challenges consists on increasing long-term field experiences regarding its implementation and management in real applications. However, in the last years some sustained initiatives have come up with some results and lessons learned. For instance, BDA has promoted the development of several commercial data-driven smart energy management startups [78]. In this regard, BDA can be embraced as a key enabler to leverage businesses impact and profitability while improving social and sustainable development simultaneously. From the prior initiatives, it is important to highlight the use of core technologies such as cloud computing, distributed databases, IoT and novel intelligent sensing devices.

As expected, not only startups have taken advantage of BDA capabilities. Large companies involved in the power industry sector such as IBM, T-Systems, Siemens and Cisco are offering big data solutions for analysis and management. Furthermore, some collaboration projects between some of these companies and utilities or governmental agencies have taken place [7]. Also some BDA benefits have been exposed by operators related to predictive assets maintenance and transformers protection from geomagnetic disturbances using PMUs data [6]. These experiences may encourage other energy players to gradually adopt BDA to support their decision making. On the other hand, due to the nature of their business, utilities have evidenced the relevance of counting with dynamic interactive visualizations to emphasize the nature of data instead of presenting the information in pre-designed manners [7,79]. Moreover, 3D visualizations have been introduced to overcome the significant limitations of conventional 2D GISs when analyzing geo-spatial and power data [83]. Therefore, visual analytics will be given more relevance in the future when deploying big data-driven interfaces. Finally, it is also worth mentioning real-world implementation experiences [6] from academic research groups where for instance, successful peak-saving was achieved by means of a utility BDA-based platform [106] and a demonstrative electrical distribution system was developed to incorporate self-healing, flexible power transfer and demand-side management actions [107]. These projects evidence a couple of many successful partnerships between the academia and the private sector. Hence, these initiatives should be promoted and sustained on a long-term basis.

Recent demonstration of wide area monitoring and control systems for transmission and distribution networks, in which PMUs are used to support power system monitoring and operation, are a proof that adequate management of large quantity of data can lead to optimal utilization of the system assets, as well as flexible integration of renewable energy sources. In [108–110], both representative monitoring applications and data inquisition-management architectures are presented. Furthermore, in [111–113] a fast-response PMU-based wide area frequency control is presented for transmission networks. Such a control can be adapted and leveraged by BDA frameworks to be also used in controlling micro-grids, or islanded distribution networks.

In respect to the challenges and future prospects, most of the references already agree that privacy and information security are two of the biggest concerns. Despite the fact that the modernization and digitization of the grid translates into greater efficiency and system utilization, as a side-effect this also brings greater possible threats to data and communication infrastructures that are prone to greater menaces and manipulation for malicious purposes. In this regard, BDA is being consistently employed for the monitoring, intrusion detection and categorization of different power systems hazards in order to address this risks in a fast high-performance manner. Being the power network a key critical infrastructure for any government, it could not be stressed enough the relevancy that all the energy stakeholders should devote to the deployment of legislation and practices focused to safeguarding the integrity of the system. Hence, there is a need for superior cyber-security frameworks, regulations and laws to protect all the energy players and keep safe sensitive information and assets. Further elaboration on this demanding topic can be found in [114]. On the other hand, basic agreements about smart meter ownership have not been addressed yet such as who owns the data and how much of these data can be mined [14]. For instance, Ref. [78] states that consumers must have the right to own their measured and personal data which should only be used with explicit agreement. However, this still represents an open debate. Other demanding issues deal with data science and mathematical challenges to collect vast unstructured incomplete data to then be filtered, compressed [115] and processed with robust and rapid algorithms [13]. The difficulty to achieve real-time big data intelligence for instantaneous detection of anomalous occurrences has also been evidenced [12]. To tackle these requirements, new machine learning technologies such as deep learning and online learning have been remarked as opportune alternatives [14].

Complementary, the lack of universal data format standards has also been pointed out [10] along with the need of increased network bandwidth capacities [6]. Therefore, greater research on high-performance computation processing and the practical deployment of AI techniques are still needed. Otherwise, the final purpose of BDA to extract valuable information for successful decision-making may not be achieved given the trend of ever increasing amounts of data. Finally, to ensure a thriving deployment of BDA infrastructures, new retail business models [14], prolonged investments and well trained professionals with multidisciplinary backgrounds from energy and computer sciences will be required [6,78]. Last but not least, the creation of integrated frameworks for adequate incorporation of BDA, particularly focused on service-oriented architectures [116], can lead to their successful application in future power networks.

7. Conclusions

From the early beginnings of the power industry digitization in the 1980's, knowledge mining has played a prominent role to take benefit from complex and ever increasing datasets. Since the 1990's, the extended deployment of SCADAs along with market restructuring and distributed generation substantially increased the need for robust data analysis techniques. In this context, Data Mining (DM) emerged as a convenient approach that permitted the deployment of a broad range of applications suitable for transmission and distribution networks. The areas that have been supported the most are the ones where there is plenty of data, the classical modeling is eminently complex or unattainable and also in the cases with high system uncertainties. Nonetheless, the ongoing power system challenges that have taken place in the last years, provoked the evolvement of the DM framework into a Big Data Analytics (BDA) scope to face the inherent demands of the data-intensive era we are witnessing. In this respect, benefiting from the newly generated data, data-driven solutions have been employed to come up with new modeling alternatives that treat data as actionable knowledge and not merely as isolated silos of information. Therefore, the potential relying in Big Data has been leveraged with the incorporation of robust tools and techniques that extract useful patterns, correlations and insights; in order to perform the corresponding knowledge discovery from the actual overfloods of heterogeneous information. Under this approach, energy stakeholders are managing to make better utilization of existing assets and also to held more-informed investments. Indeed, all the electricity value chain has been benefited from BDA with numerous applications. Particularly, BDA is being contributing significantly to the modernization of the distribution sector which is experiencing significant market disruptions, technological challenges and new roles of end-clients. To provide solutions to these needs, the role of the different AI approaches (including Machine Learning) has been noticeable. Furthermore, the utilization of AI is not anymore limited to large companies as now it is also accessible to practitioners and engineers that can enjoy from ready-to-use free-access libraries and tools from popular programming languages and software. Under this scenario, BDA has emerged in power networks as a backbone framework from which we are beginning to witness the first sustained deployments and experiences, existing already some successful outcomes in real-world applications. Nonetheless, as a side-effect of the Big Data phenomenon, greater possible threats to critical data and communication infrastructures have arisen. However, BDA can also be used to alleviate this risk by detecting and mitigating cyber-attacks. However, legislation and regulations also need to take place for encouraging proper data security policies and the extension of best security practices. Despite of its challenges as an emerging technology, with no doubt BDA will play a major role to shape future power systems.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

References

[1] V. Terzija, G. Valverde, D. Cai, P. Regulski, V. Madani, J. Fitch, S. Skok, M.M. Begovic, A. Phadke, Wide-Area Monitoring, Protection, and Control of Future Electric Power Networks, Proc. IEEE 99 (1) (2011) 80–93, http://dx.doi.org/10.1109/JPROC.2010.2060450.

[2] F.F. Wu, K. Moslehi, A. Bose, Power System Control Centers: Past, Present and Future, Proc. IEEE 93 (11) (2005) 1890–1908, http://dx.doi.org/10.1109/JPROC.2005.857499.

[3] International Renewable Energy Agency (IRENA), Scaling Up Variable Renewable Power: The Role of Grid Codes, 2016, URL www.irena.org/Publications.

[4] I. Alotaibi, M.A. Abido, M. Khalid, A.V. Savkin, A Comprehensive Review of Recent Advances in Smart Grids: A Sustainable Future with Renewable Energy Resources, Energies 13 (23) (2020) http://dx.doi.org/10.3390/en13236269, URL https://www.mdpi.com/1996-1073/13/23/6269.

[5] A. Zobaa, T. Bihl (Eds.), Big Data Analytics in Future Power Systems, Taylor & Francis Group, 2019.

[6] IEEE Smart Grid Big Data Analytics, Machine Learning and Artificial Intelligence in the Smart Grid Working Group, Big Data Analytics, Machine Learning and Artificial Intelligence in the Smart Grid: Introduction, Benefits, Challenges and Issues, White Paper, 2017.

[7] R. Arghandeh, Y. Zhou (Eds.), Big Data Application in Power Systems, Elsevier Science, 2018.

[8] H. Jiang, K. Wang, Y. Wang, M. Gao, Y. Zhang, Energy Big Data: A Survey, IEEE Access 4 (2016) 3844–3861, http://dx.doi.org/10.1109/ACCESS.2016.2580581.

[9] M. Andoni, V. Robu, S. Flynn, D. Geach, D. Jenkinks, P. McCallum, P. A., Blockchain Technology in the Energy Sector: A Systematic Review of Challenges and Oportunities, Renew. Sustain. Energy Rev. 100 (2019) 143–174, http://dx.doi.org/10.1109/TSG.2020.3018633.

[10] Y. Zhang, T. Huang, E. Bompard, Big Data Analytics in Smart Grids: A Review, Energy Inform. (2018).

[11] A. Amr, M. Yasser, Big Data Framework for Analytics in Smart Grids, Electr. Power Syst. Res. 151 (2017).

[12] M. Ghorbanian, S.H. Dolatabadi, P. Siano, Big Data Issues in Smart Grids: A Survey, IEEE Syst. J. (2019) 1–12, http://dx.doi.org/10.1109/JSYST.2019.2931879.

[13] N. Yu, S. Shah, R. Johnson, R. Sherick, M. Hong, K. Loparo, Big Data Analytics in Power Distribution Systems, in: 2015 IEEE Power Energy Society Innovative Smart Grid Technologies Conference, ISGT, 2015, pp. 1–5, http://dx.doi.org/10.1109/ISGT.2015.7131868.

[14] Y. Wang, Q. Chen, T. Hong, C. Kang, Review of Smart Meter Data Analytics: Applications, Methodologies, and Challenges, IEEE Trans. Smart Grid 10 (3) (2019) 3125–3148, http://dx.doi.org/10.1109/TSG.2018.2818167.

[15] K. Zhou, S. Yang, Understanding Household Energy Consumption Behavior: The Contribution of Energy Big Data analytics, Renew. Sustain. Energy Rev. (2016).

[16] B. Schuelke-Leech, B. Barry, M. Muratori, B. Yurkovich, Big Data Issues and Opportunities for Electric Utilities, Renew. Sustain. Energy Rev. (2015).

[17] J. Mahseredjian, V. Dinavahi, J.A. Martinez, Simulation Tools for Electromagnetic Transients in Power Systems: Overview and Challenges, IEEE Trans. Power Deliv. 24 (3) (2009) 1657–1669, http://dx.doi.org/10.1109/TPWRD.2008.2008480.

[18] M.D. Omar Faruque, T. Strasser, G. Lauss, V. Jalili-Marandi, P. Forsyth, C. Dufour, V. Dinavahi, A. Monti, P. Kotsampopoulos, J.A. Martinez, K. Strunz, M. Saeedifard, X. Wang, D. Shearer, M. Paolone, Real-Time Simulation Technologies for Power Systems Design, Testing, and Analysis, IEEE Power Energy Technol. Syst. J. 2 (2) (2015) 63–73, http://dx.doi.org/10.1109/JPETS.2015.2427370.

[19] L. Wehenkel, T. Van Cutsem, M. Ribbens-Pavella, An Artificial Intelligence Framework for On-Line Transient Stability Assessment of Power Systems, IEEE Power Eng. Rev. 9 (5) (1989) 77–78, http://dx.doi.org/10.1109/MPER.1989.4310721.

[20] L. Chen-Ching, T. Dillon, State-of-the-Art of Expert System Applications to Power Systems, Int. J. Electr. Power Energy Syst. (1992) Special Issue Expert System Applications in Power Systems.

[21] L. Wehenkel, Automatic Learning Techniques in Power Systems, Springer, 1998.

[22] L. Wehenkel, M. Pavella, Decision Tree Approach to Power Systems Security Assessment, Int. J. Electr. Power Energy Syst. 15 (1) (1993) 13–36.

[23] M.A. Laughton, Artificial Intelligence Techniques in Power Systems, in: IEEE Colloquium on Artificial Intelligence Techniques in Power Systems (Digest No: 1997/354), 1997, pp. 1/1–119, http://dx.doi.org/10.1049/ic:19971179.

[24] Z. Zhang, G. Hope, O. Malik, Expert Systems in Electric Power Systems-A Bibliographical Survey, IEEE Trans. Power Syst. 4 (4) (1989) 1355–1362, http://dx.doi.org/10.1109/59.41685.

[25] Y. Kono, H. Suzuki, K. Matsumoto, Expert System Applications to Power Systems Operation, in: Proceedings. 5th IEEE International Symposium on Intelligent Control 1990, 1990, pp. 1139–1144 vol.2, http://dx.doi.org/10.1109/ISIC.1990.128597.

[26] R. Aggarwal, Y. Song, Artificial Neural Networks in Power Systems. III. Examples of Applications in Power Systems, Power Eng. J. 12 (6) (1998) 279–287, http://dx.doi.org/10.1049/pe:19980609.

[27] R.C. Bansal, Bibliography on the Fuzzy Set Theory Applications in Power Systems (1994–2001), IEEE Trans. Power Syst. 18 (4) (2003) 1291–1299, http://dx.doi.org/10.1109/TPWRS.2003.818595.

[28] G. Lambert-Torres, A.P.A. da Silva, V.H. Quintana, L.E.B. da Silva, Classification of Power System Operation Point Using Rough Set Techniques, in: 1996 IEEE International Conference on Systems, Man and Cybernetics. Information Intelligence and Systems (Cat. No.96CH35929), vol. 3, 1996, pp. 1898–1903 vol.3, http://dx.doi.org/10.1109/ICSMC.1996.565401.

[29] A. Mantawy, Genetic Algorithms in Applications, IntechOpen, Rijeka, Croatia, 2012, pp. 91–124.

[30] N. Hatziargyriou, Machine Learning Applications to Power Systems, in: Machine Learning and Its Applications: Advanced Lectures, Springer Berlin Heidelberg, 2001, pp. 308–317.

[31] V. Miranda, D. Srinivasan, L. Proenca, Evolutionary Computation in Power Systems, Int. J. Electr. Power Energy Syst. 20 (2) (1998) 89–98.

[32] A. Hirsch, Y. Parag, J. Guerrero, Microgrids: A review of Technologies, Key Drivers, and Outstanding Issues, Renew. Sustain. Energy Rev. 90 (2018) 402–411.

[33] D. Hand, Data Mining: Statistics and More? Amer. Statist. 52 (2) (1998) 112–118.

[34] U.M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, R. Uthurusamy (Eds.), Advances in Knowledge Discovery and Data Mining, American Association for Artificial Intelligence, Menlo Park, CA, USA, 1996.

[35] T. Olaru, P. Geurts, L. Wehenkel, S. Roy, T. Ho, Data Mining Tools and Application in Power System Engineering, in: Proceedings of the 13th Power System Computation Conference, PSCC99, 1999, pp. 324–330.
[36] A. Shi-Nash, D. Hardoon, Data Analytics and Predictive Analytics in the Era of Big Data, in: Internet of Things and Data Analytics Handbook, 2021, pp. 329–345.
[37] J. Han, M. Kamber, P. Jian, Data Mining: Concepts and Techniques, Morgan Kauffman Publisher, 2012.
[38] M. Kantardzic, Data Mining: Concepts, Models, Methods, and Algorithms, Wiley-IEEE Press, 2011.
[39] G. Andersson, P. Donalek, R. Farmer, N. Hatziargyriou, I. Kamwa, P. Kundur, N. Martins, J. Paserba, P. Pourbeik, J. Sanchez-Gasca, R. Schulz, A. Stankovic, C. Taylor, V. Vittal, Causes of the 2003 major grid blackouts in North America and Europe, and recommended means to improve system dynamic performance, IEEE Trans. Power Syst. 20 (4) (2005) 1922–1928, http://dx.doi.org/10.1109/TPWRS.2005.857942.
[40] M. Olken, Smart Grid Technology, IEEE PES Power Energy Mag. 10 (4) (2012) 4–6.
[41] Z.A. Vale, C. Ramos, S. Ramos, T. Pinto, Data Mining Applications in Power Systems — Case Studies and Future Trends, in: 2009 Transmission Distribution Conference Exposition: Asia and Pacific, 2009, pp. 1–4, http://dx.doi.org/10.1109/TD-ASIA.2009.5356830.
[42] A. Kumar-Mishra, A. Kumar-Saxena, Data Mining Technology and its Applications to Power Systems, Int. J. Comput. Appl. 131 (8) (2015).
[43] R. Aggarwal, Y. Song, Artificial Neural Networks in Power Systems. I. General Introduction to Neural Computing, Power Eng. J. 11 (3) (1997) 129–134, http://dx.doi.org/10.1049/pe:19970306.
[44] K. Swarup, K. Shanti, Artificial Neural Network Using Pattern Recognition for Security Assessment and Analysis, Neurocomputing 71 (4–6) (2008) 983–998, http://dx.doi.org/10.1016/j.neucom.2007.02.017.
[45] H. Mori, State-of-the-Art Overview on Data mining in Power Systems, in: 2006 IEEE Power Engineering Society General Meeting, 2006, http://dx.doi.org/10.1109/PES.2006.1709104.
[46] Q. Wang, F. Li, Y. Tang, Y. Xu, Integrating Model-Driven and Data-Driven Methods for Power System Frequency Stability Assessment and Control, IEEE Trans. Power Syst. 34 (6) (2019) 4557–4568, http://dx.doi.org/10.1109/TPWRS.2019.2919522.
[47] M. He, V. Vittal, J. Zhang, Online Dynamic Security Assessment with Missing PMU Measurements: A Data Mining Approach, IEEE Trans. Power Syst. 28 (2) (2013) 1969–1977, http://dx.doi.org/10.1109/TPWRS.2013.2246822.
[48] M. He, J. Zhang, V. Vittal, Robust Online Dynamic Security Assessment Using Adaptive Ensemble Decision-Tree Learning, IEEE Trans. Power Syst. 28 (4) (2013) 4089–4098, http://dx.doi.org/10.1109/TPWRS.2013.2266617.
[49] E.M. Voumvoulakis, A.E. Gavoyiannis, N.D. Hatziargyriou, Decision Trees for Dynamic Security Assessment and Load Shedding Scheme, in: 2006 IEEE Power Engineering Society General Meeting, 2006, p. 7, http://dx.doi.org/10.1109/PES.2006.1709418.
[50] J.C. Cepeda, D.G. Colomé, N.J. Castrillón, Dynamic Vulnerability Assessment Due to Transient Instability Based on Data Mining Analysis for Smart Grid Applications, in: 2011 IEEE PES Conference in Innovative Smart Grid Technologies Latin America (ISGT la), 2011, pp. 1–7, http://dx.doi.org/10.1109/ISGT-LA.2011.6083211.
[51] D.P. Mishra, S.R. Samantaray, G. Joos, A Combined Wavelet and Data-Mining Based Intelligent Protection Scheme for Microgrid, IEEE Trans. Smart Grid 7 (5) (2016) 2295–2304, http://dx.doi.org/10.1109/TSG.2015.2487501.
[52] E. Casagrande, W.L. Woon, H.H. Zeineldin, D. Svetinovic, A Differential Sequence Component Protection Scheme for Microgrids With Inverter-Based Distributed Generators, IEEE Trans. Smart Grid 5 (1) (2014) 29–37, http://dx.doi.org/10.1109/TSG.2013.2251017.
[53] L. Xu, M. Chow, L.S. Taylor, Power Distribution Fault Cause Identification With Imbalanced Data Using the Data Mining-Based Fuzzy Classification *E*-Algorithm, IEEE Trans. Power Syst. 22 (1) (2007) 164–171, http://dx.doi.org/10.1109/TPWRS.2006.888990.
[54] S. Kar, S.R. Samantaray, M.D. Zadeh, Data-Mining Model Based Intelligent Differential Microgrid Protection Scheme, IEEE Syst. J. 11 (2) (2017) 1161–1169, http://dx.doi.org/10.1109/JSYST.2014.2380432.
[55] S. Pan, T. Morris, U. Adhikari, Developing a Hybrid Intrusion Detection System Using Data Mining for Power Systems, IEEE Trans. Smart Grid 6 (6) (2015) 3104–3113, http://dx.doi.org/10.1109/TSG.2015.2409775.
[56] E.E. Bernabeu, J.S. Thorp, V. Centeno, Methodology for a Security/Dependability Adaptive Protection Scheme Based on Data Mining, IEEE Trans. Power Deliv. 27 (1) (2012) 104–111, http://dx.doi.org/10.1109/TPWRD.2011.2168831.
[57] H.G. Far, A.J. Rodolakis, G. Joos, Synchronous Distributed Generation Islanding Protection Using Intelligent Relays, IEEE Trans. Smart Grid 3 (4) (2012) 1695–1703, http://dx.doi.org/10.1109/TSG.2012.2208659.
[58] V. Figueiredo, F. Rodrigues, Z. Vale, J.B. Gouveia, An Electric Energy Consumer Characterization Framework Based on Data Mining Techniques, IEEE Trans. Power Syst. 20 (2) (2005) 596–602, http://dx.doi.org/10.1109/TPWRS.2005.846234.
[59] B.D. Pitt, D.S. Kirschen, Application of Data Mining Techniques to Load Profiling, in: Proceedings of the 21st International Conference on Power Industry Computer Applications. Connecting Utilities. PICA 99. To the Millennium and beyond (Cat. No.99CH36351), 1999, pp. 131–136, http://dx.doi.org/10.1109/PICA.1999.779395.
[60] H.-C. Wu, C.-N. Lu, A Data Mining Approach for Spatial Modeling in Small Area Load Forecast, IEEE Trans. Power Syst. 17 (2) (2002) 516–521, http://dx.doi.org/10.1109/TPWRS.2002.1007927.
[61] A.H. Nizar, Z.Y. Dong, M. Jalaluddin, M.J. Raffles, Load Profiling Method in Detecting Non-Technical Loss Activities in a Power Utility, in: 2006 IEEE International Power and Energy Conference, 2006, pp. 82–87, http://dx.doi.org/10.1109/PECON.2006.346624.
[62] C. Alzate, M. Sinn, Improved Electricity Load Forecasting Via Kernel Spectral Clustering of Smart Meters, in: 2013 IEEE 13th International Conference on Data Mining, 2013, pp. 943–948, http://dx.doi.org/10.1109/ICDM.2013.144.
[63] A. Kusiak, H. Zheng, Z. Song, Short-Term Prediction of Wind Farm Power: A Data Mining Approach, IEEE Trans. Energy Convers. 24 (1) (2009) 125–136, http://dx.doi.org/10.1109/TEC.2008.2006552.
[64] H. Mori, N. Kosemura, T. Kondo, K. Numa, Data Mining for Short-Term Load Forecasting, in: 2002 IEEE Power Engineering Society Winter Meeting. Conference Proceedings (Cat. No.02CH37309), vol. 1, 2002, pp. 623–624 vol.1, http://dx.doi.org/10.1109/PESW.2002.985075.
[65] J. Ferreira, S. Ramos, Z. Vale, J. Soares, A Data-Mining-Based Methodology for Transmission Expansion Planning, IEEE Intell. Syst. 26 (2) (2011) 28–37, http://dx.doi.org/10.1109/MIS.2011.4.
[66] J.H. Zhao, Z.Y. Dong, Z. Xu, K.P. Wong, A Statistical Approach for Interval Forecasting of the Electricity Price, IEEE Trans. Power Syst. 23 (2) (2008) 267–276, http://dx.doi.org/10.1109/TPWRS.2008.919309.
[67] A. Motamedi, H. Zareipour, W.D. Rosehart, Electricity Price and Demand Forecasting in Smart Grids, IEEE Trans. Smart Grid 3 (2) (2012) 664–674, http://dx.doi.org/10.1109/TSG.2011.2171046.
[68] J.H. Zhao, Z.Y. Dong, X. Li, K.P. Wong, A Framework for Electricity Price Spike Analysis With Advanced Data Mining Methods, IEEE Trans. Power Syst. 22 (1) (2007) 376–385, http://dx.doi.org/10.1109/TPWRS.2006.889139.
[69] S.-J. Huang, J.-M. Lin, Enhancement of Anomalous Data Mining in Power System Predicting-Aided State Estimation, IEEE Trans. Power Syst. 19 (1) (2004) 610–619, http://dx.doi.org/10.1109/TPWRS.2003.818726.
[70] M. Guder, O. Salor, I. Çadirci, B. Ozkan, E. Altintas, Data Mining Framework for Power Quality Event Characterization of Iron and Steel Plants, IEEE Trans. Ind. Appl. 51 (4) (2015) 3521–3531, http://dx.doi.org/10.1109/TIA.2015.2406751.
[71] A. Asheibi, D. Stirling, D. Robinson, Identification of Load Power Quality Characteristics using Data Mining, in: 2006 Canadian Conference on Electrical and Computer Engineering, 2006, pp. 157–162, http://dx.doi.org/10.1109/CCECE.2006.277720.
[72] M. Schlechtingen, I.F. Santos, S. Achiche, Using Data-Mining Approaches for Wind Turbine Power Curve Monitoring: A Comparative Study, IEEE Trans. Sustain. Energy 4 (3) (2013) 671–679, http://dx.doi.org/10.1109/TSTE.2013.2241797.
[73] A. Kaci, I. Kamwa, L. Dessaint, S. Guillon, Synchrophasor Data Baselining and Mining for Online Monitoring of Dynamic Security Limits, IEEE Trans. Power Syst. 29 (6) (2014) 2681–2695, http://dx.doi.org/10.1109/TPWRS.2014.2312418.
[74] M. Kazerooni, H. Zhu, T.J. Overbye, Literature Review on the Applications of Data Mining in Power Systems, in: 2014 Power and Energy Conference at Illinois, PECI, 2014, pp. 1–8, http://dx.doi.org/10.1109/PECI.2014.6804567.
[75] L. Xie, Y. Chen, P.R. Kumar, Dimensionality Reduction of Synchrophasor Data for Early Event Detection: Linearized Analysis, IEEE Trans. Power Syst. 29 (6) (2014) 2784–2794, http://dx.doi.org/10.1109/TPWRS.2014.2316476.
[76] J. Gantz, D. Reinzel, Extracting Value from Chaos, Proc. IDC IView (2011).
[77] H. Hu, Y. Wen, T. Chua, X. Li, Toward Scalable Systems for Big Data Analytics: A Technology Tutorial, IEEE Access 2 (2014) 652–687, http://dx.doi.org/10.1109/ACCESS.2014.2332453.
[78] K. Zhou, C. Fu, S. Yang, Big Data Driven Smart Energy Management: From Big Data to Big Insights, Renew. Sustain. Energy Rev. (2016).
[79] J. Zhu, E. Zhuang, J. Fu, J. Baranowski, A. Ford, J. Shen, A Framework-Based Approach to Utility Big Data Analytics, IEEE Trans. Power Syst. 31 (3) (2016) 2455–2462, http://dx.doi.org/10.1109/TPWRS.2015.2462775.
[80] X. Wu, X. Zhu, G.-Q. Wu, W. Ding, Data Mining with Big Data, IEEE Trans. Knowl. Data Eng. 26 (1) (2014) 97–107, http://dx.doi.org/10.1109/TKDE.2013.109.
[81] X. Wu, H. Chen, G. Wu, J. Liu, Q. Zheng, X. He, A. Zhou, Z.-Q. Zhao, B. Wei, M. Gao, Y. Li, Q. Zhang, S. Zhang, R. Lu, N. Zheng, Knowledge Engineering with Big Data, IEEE Intell. Syst. 30 (5) (2015) 46–55, http://dx.doi.org/10.1109/MIS.2015.56.
[82] Z. Ten, Application Technology of Big Data in Smart Distribution Grid and its Prospect Analysis, Power Syst. Technol. 38 (2014) 3305–3312.
[83] Y. Guo, Z. Yang, S. Feng, J. Hu, Complex Power System Status Monitoring and Evaluation Using Big Data Platform and Machine Learning Algorithms: A Review and a Case Study, Complexity 151 (2018).

[84] M. Godoy (Ed.), Artificial Intelligence for Smarter Power Systems, The Institution of Engineering and Technology (IET), 2021.
[85] S. Stock, D. Babazadeh, C. Becker, Applications of Artificial Intelligence in Distribution Power System Operation, IEEE Access 9 (2021) 150098–150119, http://dx.doi.org/10.1109/ACCESS.2021.3125102.
[86] T&D World. Smart Meter Deployment Projected to Reach 107 Million as of Year-End 2020, 2021, URL https://bit.ly/2LymSZ3 (Accessed 01 January 2021).
[87] UK Department for Business, Energy and Industrial Strategy. Smart Meter Statistics in Great Britain: Quarterly Report to end March 2020, 2021, URL https://bit.ly/2Xobtxy, (Accessed 01 January 2021).
[88] Global Market Insights. Smart Electric Market, 2021, URL https://bit.ly/2LdI3jc, (Accessed 01 January 2021).
[89] T. Wilcox, N. Jin, P. Flach, J. Thumim, A Big Data Platform for Smart Meter Data Analytics, Comput. Ind. (2019).
[90] M. Ansari, V. Vakili, B. Bahrak, Evaluation of big data frameworks for analysis of smart grids, J. Big Data 6 (2019).
[91] IENE Institute of Energy. IRENA Forecasts 59% Solar PV Price Reduction by 2025, 2021, URL https://bit.ly/3nsJnM0, (Accessed 01 January 2021).
[92] International Renewable Energy Agency (IRENA). Market Integration of Distributed Energy Resource, 2021, URL https://bit.ly/38n5SxH, (Accessed 01 January 2021).
[93] G. Prettico, M. Flammini, N. Andreadou, S. Vitiello, G. Fulli, M. Masera (Eds.), Distribution System Operators Observatory 2018 - Overview of the electricity distribution system in Europe, Publications Office of the European Union, 2019.
[94] Z.S. Hosseini, A. Khodaei, A. Paaso, Machine learning-enabled distribution network phase identification, IEEE Trans. Power Syst. (2020) 1, http://dx.doi.org/10.1109/TPWRS.2020.3011133.
[95] K. Dehghanpour, Z. Wang, J. Wang, Y. Yuan, F. Bu, A survey on state estimation techniques and challenges in smart distribution systems, IEEE Trans. Smart Grid 10 (2) (2019) 2312–2322, http://dx.doi.org/10.1109/TSG.2018.2870600.
[96] Y. Wang, H. Wu, H. Xu, Q. Li, S. Liu, A general fast power flow algorithm for transmission and distribution networks, IEEE Access 8 (2020) 23284–23293, http://dx.doi.org/10.1109/ACCESS.2020.2970205.
[97] Z. Tang, D.J. Hill, T. Liu, Distributed Coordinated Reactive Power Control for Voltage Regulation in Distribution Networks, IEEE Trans. Smart Grid 12 (1) (2021) 312–323, http://dx.doi.org/10.1109/TSG.2020.3018633.
[98] E. Lightner, Evolution of Distribution: Impacts on Grid Planning and Operations [In My View], IEEE Power Energy Mag. 18 (1) (2020) 92–96, http://dx.doi.org/10.1109/MPE.2019.2949317.
[99] A. Dubey, A. Bose, M. Liu, L.N. Ochoa, Paving the Way for Advanced Distribution Management Systems Applications: Making the Most of Models and Data, IEEE Power Energy Mag. 18 (1) (2020) 63–75, http://dx.doi.org/10.1109/MPE.2019.2949442.
[100] W. Wang, N. YU, AMI Data Driven Phase Identification in Smart Grid, in: Second International Conference on Green Communications, Computing and Technologies, 2017.
[101] W. Wang, N. Yu, A Machine Learning Framework for Algorithmic Trading with Virtual Bids in Electricity Markets, in: 2019 IEEE Power Energy Society General Meeting, PESGM, 2019, pp. 1–5, http://dx.doi.org/10.1109/PESGM40551.2019.8973750.
[102] D. Cao, W. Hu, J. Zhao, G. Zhang, B. Zhang, Z. Liu, Z. Chen, F. Blaabjerg, Reinforcement Learning and Its Applications in Modern Power and Energy Systems: A Review, J. Mod. Power Syst. Clean Energy 8 (6) (2020) 1029–1042, http://dx.doi.org/10.35833/MPCE.2020.000552.
[103] Y. Zhang, F. Qiu, T. Hong, Z. Wang, F.F. Li, Hybrid Imitation Learning for Real-Time Service Restoration in Resilient Distribution Systems, IEEE Trans. Ind. Inform. (2021) 1, http://dx.doi.org/10.1109/TII.2021.3078110.
[104] L. Ge, W. Liao, S. Wang, B. Bak-Jensen, J.R. Pillai, Modeling Daily Load Profiles of Distribution Network for Scenario Generation Using Flow-Based Generative Network, IEEE Access 8 (2020) 77587–77597, http://dx.doi.org/10.1109/ACCESS.2020.2989350.
[105] M. Kezunovic, P. Pinson, Z. Obradovic, S. Grijalva, T. Hong, R. Bessa, Big Data Analytics for Future Electricity Grids, Electric Power Syst. Res. J. 189 (2020) 106788.
[106] Hawai'i Natural Energy Institute, Maui Smart Grid Demonstration Project: Managing Distribution System Resources for Improved Service Quality and Reliability, Transmission, Congestion Relief, and Grid Support Functions - Final Technical Report, 2014, URL https://www.hnei.hawaii.edu/wp-content/uploads/Maui-Smart-Grid-Final-Report.pdf.
[107] The Robert W. Galvin Center for Electricity Innovation, Perfect Power Prototype for Illinois Institute of Technology - Technical Report, 2014, URL https://www.osti.gov/servlets/purl/1191135.
[108] P. Wall, D. Papiya, J. Zhaoyang, M. Priyanka, Y. James, D. Wilson, H. Karine, S. Clark, M. Osborne, P. Ashton, V. Terzija, Deployment and Demonstration of Wide Area Monitoring System in Power System of Great Britain, J. Mod. Power Syst. Clean Energy 4 (3) (2016) 506–518.
[109] N. Alexandru, M. Albu, R. Hair, V. Terzija, An Enterprise Platform for Wide Area Monitoring of Load Dynamics Using Synchronized Measurement Technology, IFAC-PapersOnLine 49 (27) (2016) 85–90.
[110] S. Clark, D. Wilson, N. Al-Ashwal, F. Macleod, P. Mohapatra, J. Yu, P. Wall, P. Dattaray, V. Terzija, P. Ashton, M. Osborne, Addressing Emerging Network Management Needs with Enhanced WAMS in the GB VISOR Project, in: 2016 Power Systems Computation Conference, PSCC, 2016, pp. 1–7, http://dx.doi.org/10.1109/PSCC.2016.7540948.
[111] H. Qiteng, M. Nedd, S. Norris, I. Abdulhad, M. Karimi, V. Terzika, M. Benjamin, B. Keith, C. Booth, An Enterprise Platform for Wide Area Monitoring of Load Dynamics Using Synchronized Measurement Technology, J. Eng. 2019 (16) (2019) 1696–1702.
[112] H. Qiteng, M. Karimi, m. Sun, S. Norris, O. Bagleybter, D. Wilson, I. Abdulhadi, V. Terzija, B. Marshall, C. Booth, Design and Validation of a Wide Area Monitoring and Control System for Fast Frequency Response, IEEE Trans. Smart Grid 11 (4) (2020) 3394–3404.
[113] C. Yi, R. Azizipanah-Abarghooee, S. Azizi, L. Ding, V. Terzija, Smart Frequency Control in Low Inertia Energy Systems Based on Frequency Response Techniques: A Review, Appl. Energy 279 (2020) 115798.
[114] O.A. Alimi, K. Ouahada, A.M. Abu-Mahfouz, A Review of Machine Learning Approaches to Power System Security and Stability, IEEE Access 8 (2020) 113512–113531, http://dx.doi.org/10.1109/ACCESS.2020.3003568.
[115] F. De Caro, A. Andreotti, R. Araneo, M. Panella, A. Vaccaro, D. Villacci, A review of the enabling methodologies for knowledge discovery from smart grids data, in: 2020 IEEE International Conference on Environment and Electrical Engineering and 2020 IEEE Industrial and Commercial Power Systems Europe (EEEIC / I CPS Europe), 2020, pp. 1–6, http://dx.doi.org/10.1109/EEEIC/ICPSEurope49358.2020.9160678.
[116] A. Vaccaro, M. Popov, D. Villacci, V. Terzija, An Integrated Framework for Smart Microgrids Modeling, Monitoring, Control, Communication, and Verification, Proc. IEEE 99 (1) (2011) 119–132, http://dx.doi.org/10.1109/JPROC.2010.2081651.