Renewable and Sustainable Energy Reviews 60 (2016) 615–630

Contents lists available at ScienceDirect

Renewable and Sustainable Energy Reviews

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

A knowledge discovery in databases approach for industrial microgrid planning

Carlos Gamarra [a], Josep M. Guerrero [b,*], Eduardo Montero [a]

[a] Department of Electromechanical Engineering, University of Burgos, Spain
[b] Department of Energy Technology, Aalborg University, Denmark

ARTICLE INFO

Article history:
Received 26 April 2015
Received in revised form
2 January 2016
Accepted 13 January 2016
Available online 8 February 2016

Keywords:
Microgrid planning
Knowledge discovery in databases
Energy Management Systems
Data Mining
Machine Learning
Sustainability

ABSTRACT

The progressive application of Information and Communication Technologies to industrial processes has increased the amount of data gathered by manufacturing companies during last decades. Nowadays some standardized management systems, such as ISO 50.001 and ISO 14.001, exploit these data in order to minimize the environmental impact of manufacturing processes. At the same time, microgrid architectures are progressively being developed, proving to be suitable for supplying energy to continuous and intensive consumptions, such as manufacturing processes.

In the merge of these two tendencies, industrial microgrid development could be considered a step forward towards more sustainable manufacturing processes if planning engineers are capable to design a power supply system, not only focused on historical demand data, but also on manufacturing and environmental data. The challenge is to develop a more sustainable and proactive microgrid which allows identifying, designing and developing energy efficiency strategies at supply, management and energy use levels.

In this context, the expansion of *Internet of things* and *Knowledge Discovery in Databases* techniques will drive changes in current microgrid planning processes. In this paper, technical literature is reviewed and this innovative approach to microgrid planning is introduced.

Contents

Abbreviations: (IoT), Internet of things; (MG), Microgrid; (KDD), Knowledge discovery in databases; (DM), Data mining; (DW), Data warehousing; (ICTs), Information and communication technologies; (ML), Machine learning; (MES), Manufacturing Execution System; (ENMS), Environmental Management System; (EMS), Energy Management System; (ERP), Enterprise Resource Planning system; (IoE), Internet of Energy; (DER), Distributed Energy Resources; (CERTS), Consortium for Electric Reliability Technology Solutions; (LMP), Locational Marginal Pricing; (M2M), Machine-to-Machine; (H2M), Human-to-Machine

* Correspondence to: Department of Energy Technology, Aalborg University, Microgrids Research Programme, Pontoppidanstraede 101, Room 79, 9220 Aalborg East, Denmark. Tel.: +45 9940 9726.
E-mail address: joz@et.aau.dk (J.M. Guerrero).
URL: http://www.microgrids.et.aau.dk (J.M. Guerrero).

http://dx.doi.org/10.1016/j.rser.2016.01.091
1364-0321/

1. Introduction

During 2008, the world's companies processed 63 terabytes of information annually on average and the world servers processed 12 gigabytes of information daily for the average worker (about 3 terabytes of information per worker per year) [1]. For sure Internet has changed the amount of data available for companies. Following this technological evolution (towards data acquisition, transmission and storage) new concepts have appeared around computer-based science in business environments. *Internet of things* (IoT) is perhaps one of the trending topics in this field nowadays. Many authors have approached it since this term arose in 1999. For example F. Mattern and C. Floerkemeier affirm in [2] that IoT *represents a vision in which the Internet extends into the real world embracing everyday objects. Physical items are no longer disconnected from the virtual world, but can be controlled remotely and can act as physical access points to Internet services.*

Hence, it can be expected that the progressive connection of everyday objects to internet will be used to *remotely determine their state so that information systems can collect up-to-date information on physical objects and processes* [2]. Also *devices should be able to communicate each other, and to develop a certain level of intelligence.* The IoT vision is grounded in the steady advances in electronics, communications and information technologies. Due to their diminishing size [3], falling price and declining energy consumption, processors, communications modules and other electronic devices are being increasingly integrated into everyday objects. Main objectives of the integration of this kind of devices are data gathering, measuring and communication. Perera et al. identify smart grid, smart homes and smart industries between main contributors to smart products sales market by 2016 [4].

As J. Short et al. point out in [1], there exist some differences between two related concepts: *data* and *information*. Since *data are collections of numbers, characters, images or other outputs from devices that represent physical quantities as artificial signals intended to convey meaning,* they define information as *a subset of data,* considering data as *the lowest level of abstraction from which information and knowledge are derived*. During the period from 1986–2007, general-purpose computing capacity grew at an annual rate of 58%, and the world's capacity for bidirectional telecommunication grew at 28% per year, closely followed by the increase in globally stored information (23%) [5].

KDD is essentially the process of discovering useful knowledge from a collection of data. A. Berstein et al. also define KDD as *the result of an exploratory process involving the application of various algorithmic procedures for manipulating data, building models from data, and manipulating the models* [6]. The exponential grow of the amount of data in many systems, no longer allows the manual search of underlying patterns, as it used to be. *The main objective of KDD is to extract high-level knowledge from these low-level information, or in other words, to automatically process large quantities of raw data, identify the most significant and meaningful patterns, and present these as knowledge appropriate for achieving the user goals* [7]. Relationship between KDD, IoT and Data Mining (DM) is described in an accurate way by N. Ramakrishan in [8]:

- IoT collects data from different sources, which may contain data for the IoT itself.
- KDD, when applied to IoT, will convert the data collected by IoT into useful information that can then be converted into knowledge.
- DM is responsible for extracting patterns or generating models from the output of the data processing step and then feeding them into the decision-making step, which takes care of transforming its input into useful knowledge.

There are critical steps along a KDD process. Yoong and Kerschberg assert in [9] that knowledge discovery critically depends on how well a database is characterized and how consistently the existing and discovered knowledge is evolved. The step definition of the KDD process can also have a strong impact on the final results of mining. For example, not all the attributes of the data are useful for mining. The consequence is that DM algorithms may have a hard time to find useful information if the selected attributes cannot fully represent the characteristics of the data [8]. DM is described by Fayad et al. in [10] as a step in the KDD process that consists of *applying data analysis and discovering algorithms that produce a particular enumeration of patterns (or models) over the data.* But every DM process requires a previous data processing step, also defined by Fayad et al. as *data warehousing* (DW). DW refers to *collecting and cleaning transactional data to make them available for online analysis and decision support*. DW helps set the stage for KDD in two important ways: data cleaning and data access.

Typical KDD process includes five general stages: selection, pre-processing, transformation, data mining and evaluation. But, instead of being based in the same principles, different authors propose different KDD processes. Fayad et al. [10] define KDD as an *iterative and interactive* process based in nine steps such as:

- *Developing an understanding of the application domain and the relevant prior knowledge and identifying the goal of the KDD process from the customer's viewpoint.*
- *Creating a target data set.*
- *Data cleaning and pre-processing.*
- *Data reduction and projection*
- *Matching the goals of the KDD process to a particular data mining method.*
- *Exploratory analysis and model and hypothesis selection.*
- *Data mining.*
- *Interpreting mined patterns.*
- *Acting on the discovered knowledge.*

  M. Last et al. introduce a specific time series databases KDD process [11] based on seven stages: data pre-processing, feature extraction, transformation, dimensionality reduction, prediction and rule extraction. A review on time series DM techniques is also presented by Fu [12]. Between mining tasks he highlights pattern discovery and clustering but also classification, rule discovery, summarization and other recent research directions. Finally, a deep review of 13 different KDD process models is presented by Kurgan and Musilek in [13]. Analyzing these KDD methodologies, data preparation can be considered the foundation, while *DM can be considered as the pillar of KDD*. The existence of similar, but at the same time different KDD methodologies makes sense since:

  - KDD techniques have not been widely applied to manufacturing processes, neither standardized yet.
  - KDD techniques are based on optimization problems between different alternatives, under different constraints and towards different goals depending not only on the characteristics of the manufacturing process, but also on environmental, social and legal conditions: there are some aspects in a KDD-based approach that might not be standardized.

On the basis of KDD, a growing body of emerging applications is changing the landscape of business decision support [14] such as: risk analysis, targeted marketing, customer retention [15], portfolio management and brand loyalty [16]. *Traditional DM approaches have proven to be efficient on modeling variables of interest*, so that these variables may be forecasted in future scenarios, and effective decisions taken based on that forecast. DM technologies are reviewed, described and classified [8] into *clustering, classification, and frequent patterns mining technologies, from the perspective of infrastructures and from the perspective of services* by Tsai et al. They point out that *most KDD systems available today and most traditional mining algorithms cannot be applied directly to process the large amount of data of IoT*. In this line of work, they define *three key considerations in choosing the applicable mining technologies for the problem to be solved by the KDD technology: objective, characteristics of data, and mining algorithm*. In other words, whether or not to develop a new mining algorithm can be easily justified by using these factors. Regarding mining algorithms, heuristic [17–20] and metaheuristic algorithms [21] are called to be trending techniques in order to solve a variety of mining and optimization problems in [22].

Perhaps one of the most important aspects in DM problems is how to evaluate a candidate model, and, obviously, this question depends on the type of DM task at hand. *Thus, most of the DM problems can be thought of as optimization problems*, where the aim is to evolve a candidate model that optimizes certain performance criteria. However, the majority of DM problems have multiple criteria to be optimized. Hence, *most of the DM problems are multi-objective in nature*. Multiobjective evolutionary algorithms for DM are surveyed in [23,24] into different categories regarding the DM task they face: feature selection, classification, clustering, association rule mining and other tasks. A collection of stages for a general KDD process is summarized in Fig. 1.

In addition, it must be highlighted that the application of KDD strategies not only requires technical but also human resources (qualified workers). It can be an intensive process regarding time and resources, and sometimes some of the optimal solutions can result obvious or simply incompatible with the manufacturing process (even with maintenance criteria or environmental goals). For example, the fewer units manufactured the less energy spent and the less environmental impact it is caused. Another possible recommendation could be to switch off a manufacturing process with a high thermal inertia, following energy saving criteria. But this action can sometimes be incompatible with manufacturing and quality goals, for example. *Thus, the identification of effective scenarios for the application of these techniques (regarding costs and benefits) is one of the main areas of development nowadays***.**

KDD field is in continuous evolution towards the direction pointed out by Cios and Kurgan in [25]. The future of KDD and DM process models *is in achieving overall integration of the entire process through the use of other popular industrial standards. Another currently very important issue is to provide interoperability and compatibility between different software systems and platforms, which also concerns KDD and DM models.*

As a conclusion, it can be asserted that KDD systems will help in achieving a higher-level automation of manufacturing processes, less assisted by manpower with higher professional qualification. Nowadays every manufacturing company has adopted ICT-based architectures. But, as it will be described below, the growing application of IoT-based strategies in manufacturing companies will be a valuable source of additional data, also regarding microgrid (MG) planning. Hence, combined planning approaches for energy supply and manufacturing process are expected to be able to reduce environmental impact, under economic profitability conditions. *This paper introduces a MG planning approach for industrial companies from a Knowledge discovery in databases (KDD) point of view***.**

2. KDD techniques applied to manufacturing processes

Industrial production accounts for 16% of Europe's gross domestic product and remains a key driver for innovation, productivity, growth and job creation. In 2009, 31 million people were employed in the European Union manufacturing sector, and each job in manufacturing generates at least an additional job in services [26]. *Research on manufacturing processes improvement is a critical issue for European Union government*. The point of departure of recent approaches is technological evolution in Information and Communication Technologies (ICTs). Data acquisition, communication, and decision-making based on the acquired data are essential functions in modern manufacturing processes nowadays. That is the reason why Bi et al. affirms that the *IoT is able to provide solutions to planning, scheduling, and controlling of manufacturing systems at all levels*, due to the close relations of the components in a manufacturing enterprise and the architecture of IoT [27].

New manufacturing-oriented IoT architectures have been proposed in recent years. Said and Masud review in [28] IoT architectures pointing a three and a five level architecture, including:

- Business layer defines IoT applications charge, management and users privacy.
- Application layer: determining the type of applications that will be used in the IoT.
- Processing layer is the responsible of processing the information gathered by perception layer.
- Transport layer receives and transmits the information from the perception layer to the processing layer and vice versa.
- Perception layer contains the technology used in the IoT, gathering information from field devices and transforming this data to signals.

A similar IoT architecture is illustrated in [27], classifying manufacturing components in three different levels and comparing them to IoT levels: machines and devices (ubiquitous computing), enterprise application (grid computing) and virtual enterprise and enterprise alley (cloud computing). *Analyzing these architectures, it can be asserted that the future vision of industrial IoT is yet to be defined.*

Based in these non-standardized architectures, IoT-based industrial applications are growing up strongly. For example, a farm operation monitoring system is proposed by Fukatsu and Nanseky in [29]. This system is based in "Field Servers" and a wearable device equipped with an RFID reader and motion sensors. They not only monitor the field environment, but also crop growth, insect infestation, and simple field operations controlling and measuring various sensors including some cameras. Da et al. also review some industrial applications for IoT, such as vigilance of workplace safety and food supply chain. They also cite some worksite safety management tasks such as disaster signals supervision in order to make warning, risk forecasting, and safety improvement of production in the mining field. In the food industry some IoT applications have been developed in order to add traceability capacities to the food supply chain*: from precise agriculture, to food production, processing, storage, distribution, and consuming* [30]. A novel enabling-approach of IoT for manufacturing processes is the virtual factory approach, in which different manufacturing processes can be modeled and executed as if they are being carried out in a single factory. Virtual factory management involves 4 steps: process definition, process forecasting and simulation, process execution and real-time monitoring. ICTs and IoT support is essential to success [31] in this process.

Under this context, data use is expected to grow in manufacturing processes, *but only under cost efficiency criteria and goal-oriented***.** Several information from productive process variables remains in databases, which at the end, store huge amounts of historical data. As a consequence, KDD techniques have a growing market in the field of manufacturing processes optimization and decision-making. A framework for implementation of KDD projects in firms is applied to the Indian company Ramco [32].

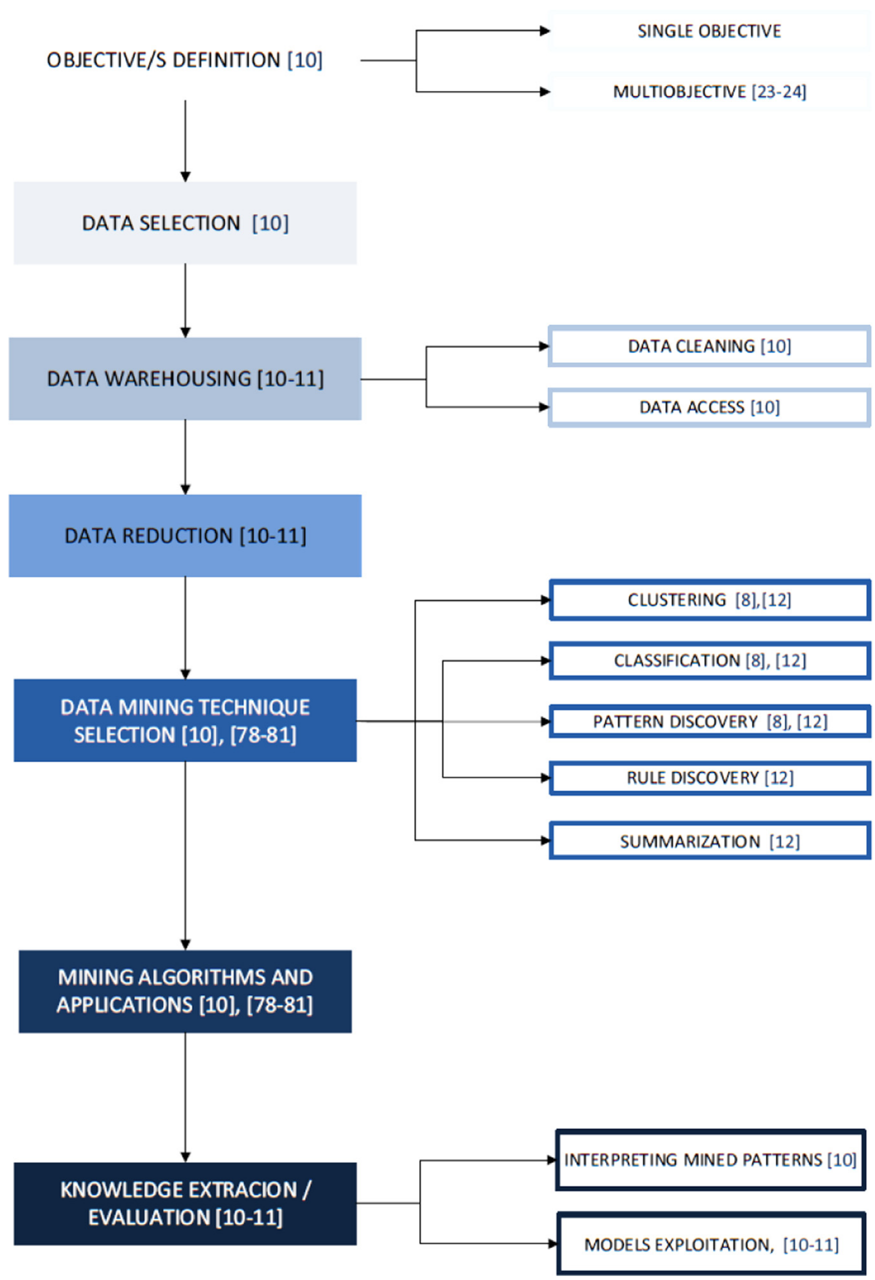

Fig. 1. A general KDD process scheme.

Authors divide this process into 10 steps, shaping 3 stages: strategic groundwork, data analytics and implementation. *Knowledge from databases should not only be extracted, but also capitalized:* efforts should be oriented towards cost critical issues and KDD process should be embedded into every company daily procedures. Following this guidelines, a novel KDD methodology is developed towards *modeling the knowledge requirements and the associated tasks for collecting the knowledge simultaneously in a company* by Tseng and Huang in [33].

The concept of networking devices to achieve higher levels of automated interaction is also driving changes in industrial networking. In addition to DM, (which is more focused on the automated analysis of large data in order to discover patterns and models that can be applied to forecasting and decision-making tasks), other mathematical disciplines, such as *Machine Learning* (ML), can help manufacturing industries to raise their performance. ML is focused on the construction and study of algorithms that can learn from data in order to adapt the behavior of a device to their environment (including their users). Main differences and similarities between these subfields are presented in [34]. Pan and Yang summarize the relationship between traditional ML and various transfer learning settings and categorize transfer learning under three sub-settings, inductive transfer learning, transductive transfer learning, and unsupervised transfer learning [35]. Additionally, a combination of DM and ML approaches (cycle-time factor identifying and predicting) in semiconductor manufacturing is presented in [36] with decision support purposes.

KDD applications to manufacturing processes have been reviewed in some papers such as [37–40]. Specially Liao et al. present a deep review on papers published from 2000 to 2011. Some of the most common applications and techniques will be mentioned below, allowing the reader to make a quick overview of this field.

KDD techniques application to manufacturing processes usually address the search of correlations between process variables [41–44] among production process data and control parameters. The discoveries of these correlations allow planners to incorporate the generated knowledge into a model of the manufacturing process, which can be exploited with multiple purposes. Since a manufacturing process is controlled in order to obtain a product with defined characteristics, a common application of these models is to identify critical parameters in order to get a better control of (and even forecast) the results of the manufacturing process. For example, DM techniques are proposed to study product variances in [45–48]. Moreover, the results of different control strategies or parameters are finally verified by quality control department.

Due to advances in data collection systems and analysis tools, DM has widely been applied for quality improvement in manufacturing and have been reviewed in [49] by Köksal et al.. Many papers afford quality issues using DM techniques [50–63]. Among them, semiconductor and electronics manufacturing can be considered a hot topic [51,53,58,59,61,63]. Quality problems together with new capabilities usually influence the design process of new products. DM techniques are used this process, such as it is described in [64] for a case of a motherboard design and assembly in a personal computer manufacturing process.

Not only quality but also energy and requirements for manufacturing processes have been addressed with DM techniques. The research on the process control and the energy saving of the aluminum electrolysis industry have inspired new specific DM algorithms [65]. Also Kusiak and Shong describe the existence of three different DM-based approaches to optimization of combustion efficiency: analytical models based on thermodynamics and chemistry, soft computing and hybrid systems [66]. Optimization DM-based models for improvement of a boiler–turbine system performance are formulated in [67,68]. In both papers historical process data are mined and the discovered patterns are selected for performance improvement of the system.

There is no doubt about the development or the innovative combination of techniques, enabling new approaches to KDD problems. Demand for real-time processing, on-demand processing as well as the in-transit processing of standard remote sensing data products are some of the development opportunities [69] in this field. At the same time, new manufacturing equipment is expected to be capable of (at least) reading and storing some basic data about their activity related with production, energy, time and other process-related parameters. This connected approach to manufacturing process management is very suitable to the IoT strategies application. Among manufacturing companies, the growing use of condition-based monitoring is widely accepted (if not always implemented) with different purposes such as maintenance, production quality and energy management optimization. Taking the idea a step further, manufacturers are looking forward to connect all these devices with higher decision-making levels [32]. This plant (or even multiple plant) connectivity from the device (or field) level to the enterprise (or decision-making) level connection involves connecting industrial devices to Manufacturing Execution Systems (MESs), Energy Management Systems (EMSs) or Enterprise Resource Planning systems (ERPs).

The challenge for the application of DM and ML techniques (as it is for the optimization techniques they are based on) is on achieving an optimal solution under not only technical and economical constraints, but also fulfilling environmental, social and other constraints of the manufacturing process. As it has been described in the previous section, data mining optimization processes sometimes results in non-viable or incompatible solutions. That is the main reason why constraints for manufacturing process have a very important role along the planning process. Following Fig. 1, constraints must be defined during *objective definition* stage and their fulfillment must be evaluated at *Knowledge extraction and evaluation* stage.

Some of the state, control variables and decision-making variables included in Fig. 2 will be processed as constrains along an optimization process (usually multi-objective optimization as it has been cited in Section 2). Among this variables there can be highlighted:

- Manufacturing equipment: units available, load status, maintenance status, energy consumption and environmental impact.
- Ancillary services: units available, load status, maintenance status, energy consumption and environmental impact.
- Production control: short term demand, production schedule and goals achievement.
- Quality control: data related with the fulfillment of quality standards.
- Environmental impact: exhaust gas emissions, waste materials and water consumption.
- Climatic and renewable resources data: potential wind capacity, solar irradiation, temperature and humidity.
- Energy market: power grid prices and energy costs for different microgrid planning scenarios.
- Warehouse status: raw materials and final product availability regarding short term requirements of manufacturing process.
- Legal and financial: future policies can make optimal scenarios for a manufacturing process change. Fund availability can also oscillate in order to face regular costs or additional investments.
- Sales and purchasing: raw materials supply regarding present and future sales (forecasting).
- Human resources: manpower planning, availability and costs.

Environmental and Energy Management Systems (ENMSs and EMSs) have claimed their role in industrial companies during last decades, mainly due to changes in environmental policies and to the raise of energy costs. More than 4.800 EMSs based in ISO 50.001 were certificated (a growth of 116% in a year) in 78 countries up to the end of December 2013 [70]. During 2014 the certification of this kind of systems raised up to 6.788, an additional 40%. At the same time, certified ENMSs based on ISO 14.000 grew in 2014 up to 324.148, (a 7% in a year) as the consolidated standard it is for evaluation and improvement of environmental policies. A new version of ISO 14.000 standard has been released in 2015. Certificated EMSs could be expected to keep growing during next years between a 15–20% per year, since they have many points in common with other quality or environmental international standards (such as ISO 14.000) *and energy represents a major cost for many manufacturing processes* (up to 80% for frozen processed food industry, or up to 70% for glass bottle manufacturing, for example).

Both ENMSs and EMSs are based on ICTs and software tools which track the use of energy and raw materials. At the same time, ENMSs quantify environmental impact using data, for example, of water consumption or emission of exhaust gas. The key performance indicators (KPIs) used by both systems support managers along the process of defining the environmental and energy performance of the manufacturing process, and also identifying improvement and saving opportunities. The aforementioned appearance of innovative architectures and techniques (such as IoT and DM) makes it possible for these systems to evolve quickly towards forecasting, modeling and optimizing capabilities. *Correspondingly, the point of smart industries is not only on connecting smart devices each other, but also on using gathered information in order to model and optimize a manufacturing process***.** This is the natural evolution of manufacturing management systems such as MES, EMS and ERPs. In a short period of time, data acquisition, communication and analysis tools are expected to become essential functions for these systems, and KDD techniques are expected to take part of integrated manufacturing management systems, assisting company managers along decision-making processes. Some tries of DM-based applications have been developed regarding energy efficiency for industrial companies, such as an energy audit web-based application [71]. But there is a still a lot

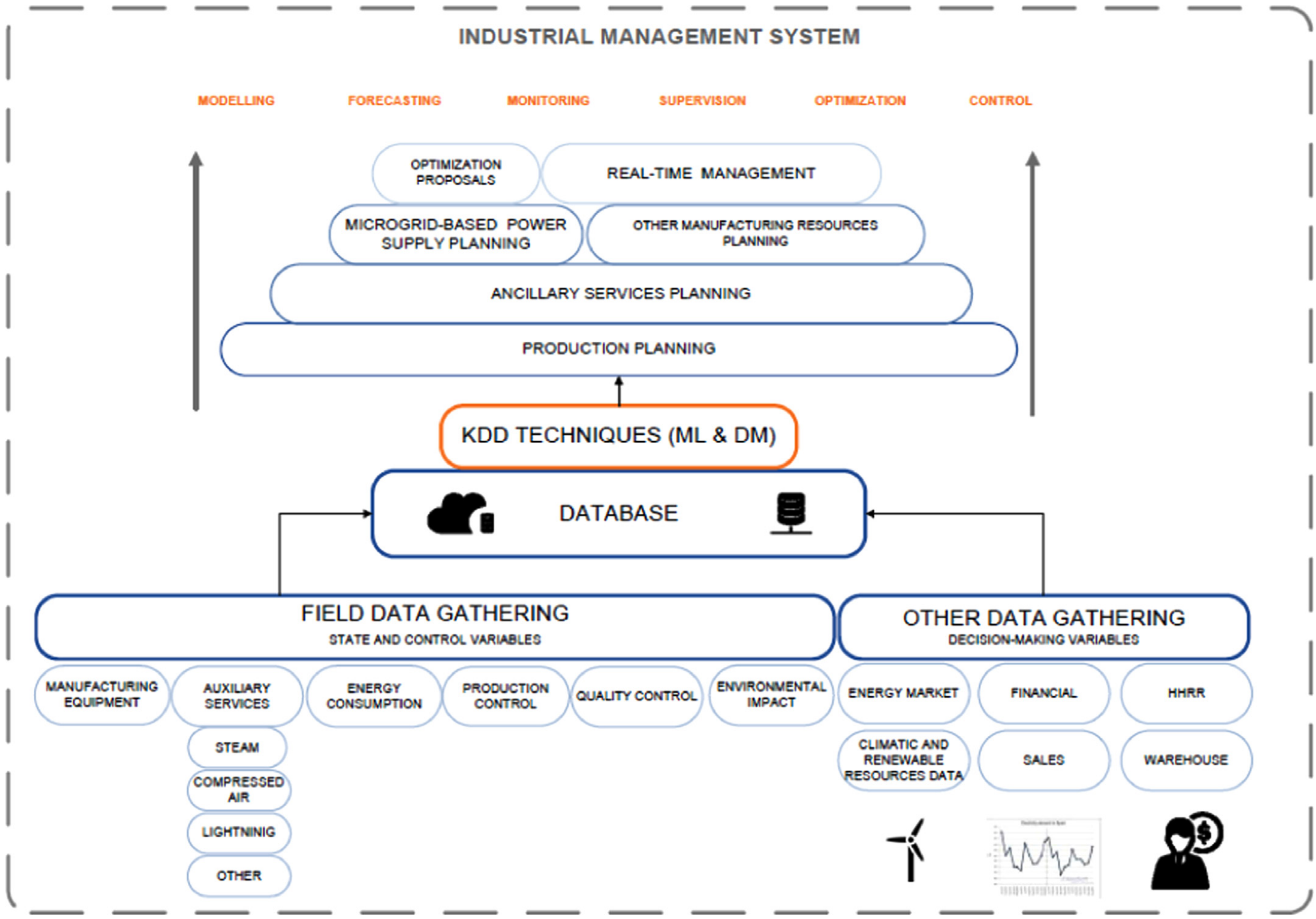

Fig. 2. Microgrid planning process scheme.

of work to do in order to integrate KDD techniques into future manufacturing management systems.

3. MG as sustainable energy supply system for manufacturing processes

The power grid, it is widely agreed, *is the next big thing in computing* [72]. Nowadays, it can be said that the traditional electrical grid has achieved quality and quantity supply goals (from the conventional and renewable generation of electricity, to power transportation and distribution), *but it has to improve substantially from the end user point of view and the functionalities expected of it*. At each step along the way, large (and growing) data volumes are created by energy-related industrial or even domestic equipment. In this context, KDD techniques are suitable to improve the overall efficiency of an energy supply, or additionally to support the development of new services for energy consumers. This concept can be applied not only to improve electrical networks performance but also to manage the actions of all the systems connected to them (those which generate electricity, which use electricity and which take both actions). Therefore, IoT architectures and KDD techniques are expected to drive the evolution of power systems into smart systems. Moreover, the visions of smart grids and IoT have recently been combined into the concept *Internet of Energy* (IoE) [73]. There is a global effort to incorporate sensors, actuators and data networks into power grids. This IoT application to power grids offers deep monitoring and controls, but needs advanced analytics over millions of data streams for efficient and reliable operational decisions [74].

Analytic tools and applications have a growing importance in every power system. *The analytics layer covers the new solutions that vendors are bringing to the market* [75]. In the smart grid space, there are three domains that will increasingly rely on analytics: the enterprise, grid operations and consumer analytics.

- Enterprise analytics are focused on moving from traditional historical analytics to real-time predictive analytics, complete situational awareness, business intelligence, real-time visualization and simulation of the grid
- Grid operations analytics are focused on grid optimization and operational intelligence, asset management analytics, crisis management analytics, decision-making analytics, outage management analytics, weather and location data, mobile workforce management and energy theft.
- Consumer analytics are focused on behavioral, demand response, load flow and distributed generation analytics, and social media data integration.

At the same time, data infrastructure and data management enables smart capabilities for power systems, providing data to analytics. A cloud-based software platform for data-driven smart grid management is presented in [74] by Simmhan et al. But cloud services to smart things may face latency and intermittent connectivity. In this context fog devices are introduced in [76], between cloud and smart devices. Their main advantage is the high-speed Internet connection to the cloud, and physical proximity to users, enabling real-time applications. For example, a demand response management algorithm can be implemented with a cloud computing approach, but the bandwidth cost would be high if each supplier and customer communicates directly with the cloud.

As it has been cited before, analytics are the foundation of some advanced features development, such as self-healing, mutual operation and participation of the users, electricity quality, distributed generations and demand response, sophisticated market and effective asset management [77]. KDD applications to power systems based in those analytics have been reviewed by different authors in 1997 [78], 2006 [79], 2009 [80] and 2014 [81]. A comparison between these papers regarding year, number of references and applications is presented in Table 1. Perhaps the most complete review among them is the one by Kazerooni et.al. [81]. They present a classification of four major areas of data mining and related works in power systems: visualization, clustering, outliers detection and classification. At the same time these applications could be divided into utility-side (electricity generation and distribution) and demand-side (energy consumption). Not only utility managers are willing to be able to exploit KDD-based cost-saving opportunities. Customers are also willing to use friendly control devices, and to consume and control energy generation from environmental-friendly Distributed Energy Resources (DER). Following this approach a wireless communication network to sense, estimate and control DER states is proposed in [82] by Rana and Li. Smart meters are a key element in order to develop this demand-based approach, since they are a powerful source of consumption data. These meters are increasingly replacing traditional meters and measuring a detailed profile of consumption data. With such data, utility companies are in possession of the raw material needed to improve efficiencies and customer services [83]. Intelligent use of data from smart meter allow power grid planners to develop deeper electricity consumption analysis [84]. A wide range of modeling, simulation and forecasting tasks can be applied in order to develop scheduling strategies, making power equipment work at higher performance levels: identifying low efficiency scenarios and supporting improvement strategies.

Environmental impact of manufacturing processes can be minimized developing advanced MG systems, not only due to the improvement of energy efficiency. For example, energy consumption for some critical ancillary services o manufacturing equipment could be supported by renewable power sources in order to increase the reliability and minimize the environmental impact of the systems. There can be found profitable applications for renewable power sources regarding environmental and economical aspects in almost every manufacturing process, especially under self consumption strategies. Net metering strategies could also be studied in some special cases such as manufacturing processes with low power capacity demands, photovoltaic or wind power potential and space availability, or either equipped with small combined heat and power systems.

Nowadays energy consumption has a very significant role in inputs cost of manufacturing companies. Furthermore power outages can cause substantial economical losses in some manufacturing processes. As a consequence, *innovative energy supply and backup architectures and systems are being developed* together with this cost rising for industrial companies. Following this line of research, recent approaches to MG planning for industrial facilities have been published in technical literature. Pipattanasompor et al. analyze [85] the optimal DG mix at various facility outage costs with and without an emission restriction. They also discuss the impact of varying the grid reliability and the capital costs of DG units on the decision to invest in backup power. Leif et al. [86] propose a technique for sizing and scheduling electricity supply at industrial sites with combined heat and power and wind generation. An industrial size MG is scheduled [87] and evaluated regarding different performance indicators such as technical, economic and environmental. The techno-economic potential for a predominantly renewable electricity-based MG serving an industrial-sized drink water plant in the Netherlands is studied [88] by Soshinskaya et al.

EMSs are ready to gather and manage data, not only from the manufacturing process, but also from power supply equipment: power generation machines (from conventional to renewable power sources), distribution substations, power transformers and grids and smart meters (that control and monitor power consumption at the point of delivery). An example of EMS based on microgrid architecture is presented in [89]. Furthermore, an integrated analysis of stored data from different units of the manufacturing process can reveal several opportunities to exploit at different layers in order to improve the global efficiency (such as energy generation, energy distribution, energy consumption, production planning, resources planning and environmental impact planning). *As a consequence, production and energy supply could be planned in an aggregated way towards achieving higher sustainability and energy efficiency goals for a manufacturing company*. This could be considered a suitable framework for industrial MG planning, modifying other traditional approaches [90]. A scheme of this innovative approach is presented in Fig. 2.

4. KDD-based approaches to industrial MG planning problems

MGs could definitely be considered a modern, small-scale version of centralized electricity systems. A MG is defined by CERTS as *clusters of generators, including heat recovery, storage, and loads, which are operated as single controllable entities*. P. Lilienthal highlights [91] different criteria for MG classification such as: types of

Table 1
KDD applications to power systems in review papers since 1997.

Authors	Citation number	Year	Referenced papers	KDD-based applications to power systems (as described in each paper)
S.Madam, W.Son and K. Bollinger	[78]	1997	9	Prediction (Forecasting), system description (Modeling)
H.Mori	[79]	2006	50	Security assesment, economic feasibility, fault detection, load profiling, power system control, economic load dispatch, data debbuging, network-expansion planning, fault detection, state estimation, load forecasting
Z. Vale et al.	[80]	2009	20	Demand response (DM-based), strategic bidding in electricity market (ML-based)
M. Kazerooni et al.	[81]	2014	36	Geographic visualization of frequency, 3D visualization of power system data, geographic data views, visualization of power system contingencies, power system visualization based on common information model, data-driven visualization, cable layout design for wind farms, load pocket identification, probabilistic evaluation of total transfer capability, load forecasting, event detection, data debugging and bad data detection, non technical loss detection, secure economic dispatch, accuracy versus interpretability trade-off, locating series compensators

energy generation, voltage level of distribution system, peak load, generation capacity, energy production, number of customers served, other grids interconnection, load management and metering. Regardless of their classification, the operation of a MG is closely tied to energy economics. This includes both the financials of interacting with the main grid and the cost of self-generation [92] and transmission. Moreover, MGs can operate in islanded mode and sustain the power supply in the event of a grid outage. Benefits from MGs should not only be economical. *MGs can also be viewed as a means of creating zero net-energy communities and meeting other environmental goals established by states or regulatory agencies***.** The establishment of a MG can make possible to achieve specific goals such as carbon emission reduction, diversification of energy sources, increase of reliability and cost reduction.

MG planning problems, as the energy community systems it is, must be approached from a cost minimization point of view [93]. But in a real microgrid, other goals should be considered, such as total environmental impact, power quality and reliability [94]. Regarding technical literature, some common problems could be defined for a MG planning process, even though each process has its own constraints and specific goals [90]. There can be considered five main stages such as *power generation and storage technology selection, sizing*, siting, *scheduling and pricing* . Among these planning problems, an additional critical stage can be also considered in order to test the final design of both architecture and control strategies for a microgrid. This additional stage is called sensitivity analysis and consists on determining how different values of an independent variable will impact a particular dependent variable, under a given set of assumptions. It is addressed towards predicting the outcome of a decision, if a situation turns out to be different compared to the key prediction [95]. The application of this analysis usually results on more robust planning results.

But MG planning does not finish with de definition of an optimal scheduling for the equipment. A MG, as the power systems it is, must be supervised and able to respond quickly and coherently to unexpected operational conditions. The control and supervision layer is implemented in a management system called EMS, such as it is for manufacturing processes. In fact, this system is the main source of synergies between manufacturing processes and MGs. But an EMS for a MG also incorporates a plan against unexpected events, which can be defined as contingency planning. KDD techniques are able to make a significant contribution to supervision and contingency planning for power systems. Some references of KDD-based techniques to power systems have been introduced in Table 2.

A growing number of new advances are progressively developing the conventional electrical energy supply, and among them, the application of KDD techniques is expected to make life easier to power companies but also to consumers. But up to now, these advances have generated *a great deal of enthusiasm, as well as a considerable amount of confusion* [75]. Grid intelligence for MG relies on data infrastructure, data management and data analytics layers. Since most manufacturing processes have already implemented data infrastructure and data management layers, only data analytics layer is left towards developing smart industrial microgrids. The quick expansion of manufacturing management systems such as MESs, EMSs and ERPs allows the increase of available historical data about manufacturing processes (such as energy consumption, environmental impact, production planning, quality keeping, risk prevention, resources planning, financial planning, etc.). In this context KDD techniques are ready not only to analyze, but also to capitalize these data [33]. Hence, *analytics can combine data management with the knowledge of energy efficiency and production process experts to uncover hidden saving potential*, thereby contributing effectively in making better business decisions. Indeed, the full potential of energy saving for a manufacturing process can only be defined through a holistic and integrated analysis of the complete value chain of the plant. Thus, MGs can solve the problem on energy alternative and compatible use, integrating system data to optimize operation and management.

Following this approach, a MG can be considered as a suitable energy supply system for a manufacturing process. The core of this system would be, without any doubt, the EMS. It is expected to manage the systems, gathering and analyzing information about how systems work in a day-by-day (even hour-by hour) basis. In fact, the use of KDD-based techniques on industrial MG planning involves a different approach to the manufacturing process planning, on the basis that a large amount of raw data about it is available. These data will be used to develop aggregated models for the industrial process, which can be addressed to effective decision-making. As a result, *it can be asserted that KDD techniques are ready to help energy planners to define economically feasible MG architectures for industrial processes*.

Among future capacities for these industrial MGs will be, without any doubt, the identification of optimal energy efficiency and sustainability scenarios for every state of the manufacturing process. In addition, EMS should help energy planners to identify the most efficient working parameters for the manufacturing process, even suggest them different strategies to exploit different

Table 2

Category	Objective	Method	Paper
State stimation	Dynamic security assessment	Decision tree	[116,125]
	Data debugging	Clustering	[128]
	Condition assessment	Correlation analysis	[127]
	Non technical losses detection	Classification	[137]
	Sensitivities of some indices at a target location to prescribed credible events	Patterns classification	[131,132]
Supervision and forecasting	Predicting-aided state estimation including bad data	Clustering	[117]
			[118]
		Correlation analysis	[118]
	Disturbance classification	Pattern recognition	[122]
	Power quality assessment	Clustering	[129,130]
			[130]
		Classification	[130]
	On-line prediction of power system transient stability	Decision tree	[133]
Fault detection and classification	Fault detection	Pattern learning	[119,134]
		Correlation analysis	[120]
		Decision tree	[123,136]
		Classification	[124]
Protection scheme configuration	Protection scheme configuration	Decision tree	[121,128,135]
		Rough set theory	[121,126]
		Classification	[134]

energy efficiency improvement opportunities, which could also be developed under supervision, supported by ML techniques.

Innovative energy efficiency and sustainability approaches to industrial microgrid planning allowed by the application of KDD techniques, are described below. In fact, techniques described below could be applied by a single planning engineer, but the real challenge is developing software applications and EMSs able to develop these approaches under DM and ML techniques (Fig. 3).

4.1. Sizing approaches

Typical goals of the sizing stage in a MG planning process are cost efficiency (low investment and operational costs), energy efficiency (low power losses and high renewable power sources penetration) and high reliability. Energy planners have usually sized a microgrid considering peak and base demand of the whole process, and oversizing power capacities in order to achieve reliability goals. Traditionally, the point of departure of a microgrid planning process has been the demand curve (on an hourly or quarter-hourly basis). A typical demand curves for a manufacturing process is shown in Fig. 4. Some characteristic points can be identified in them, such as peak demand, base demand. A power capacity histogram (as the one showed in Fig. 5) is also a valuable source of information to consider towards defining most common power capacity levels of a system. In addition, demand oscillations can be related with specific events in the manufacturing process, but the identification of these relationships could not be a simple task depending on the process and the way it is operated. KDD techniques are can nowadays address these and other potential tasks described below:

- New bottom-up approach to sizing: the existence of field data allows energy planners to adopt more complex strategies. Different models can be generated and validated using raw data, not only for the whole manufacturing process but also for different sub-processes or ancillary services. These models can be applied in order to define the optimal architecture and size for the MG. The analysis of raw data [83] makes it possible to define different power demand intervals and levels regarding different power capacity usage scenarios for manufacturing equipment. Once every power capacity level is characterized for every manufacturing unit and ancillary service, optimal technologies, sizes and power regulation capacities for generation equipment can be accurately defined.
- Capacity charge saving opportunities: an optimally sized MG can considerably minimize capacity charges from utilities. In order to achieve these savings, an optimal mix of power technologies and power capacities must be defined for MG. These power capacity reduction savings are based again into power consumption data analysis (from the manufacturing process), but also into additional information from electrical company such as capacity costs, peak and off peak periods schedule, or even additional signals based on-demand side management or demand response strategies. A detailed cost-benefit analysis on power capacity should be done in order to achieve a long-term economic viability for a grid-connected MG.
- Energy storage capacity definition: some MG aspects such as quality keeping and renewable energy resources penetration are highly depend on energy storage capacity and scheduling. Real field data can help in the task of developing accurate models for the MG in order to optimize the technology selection, size (and

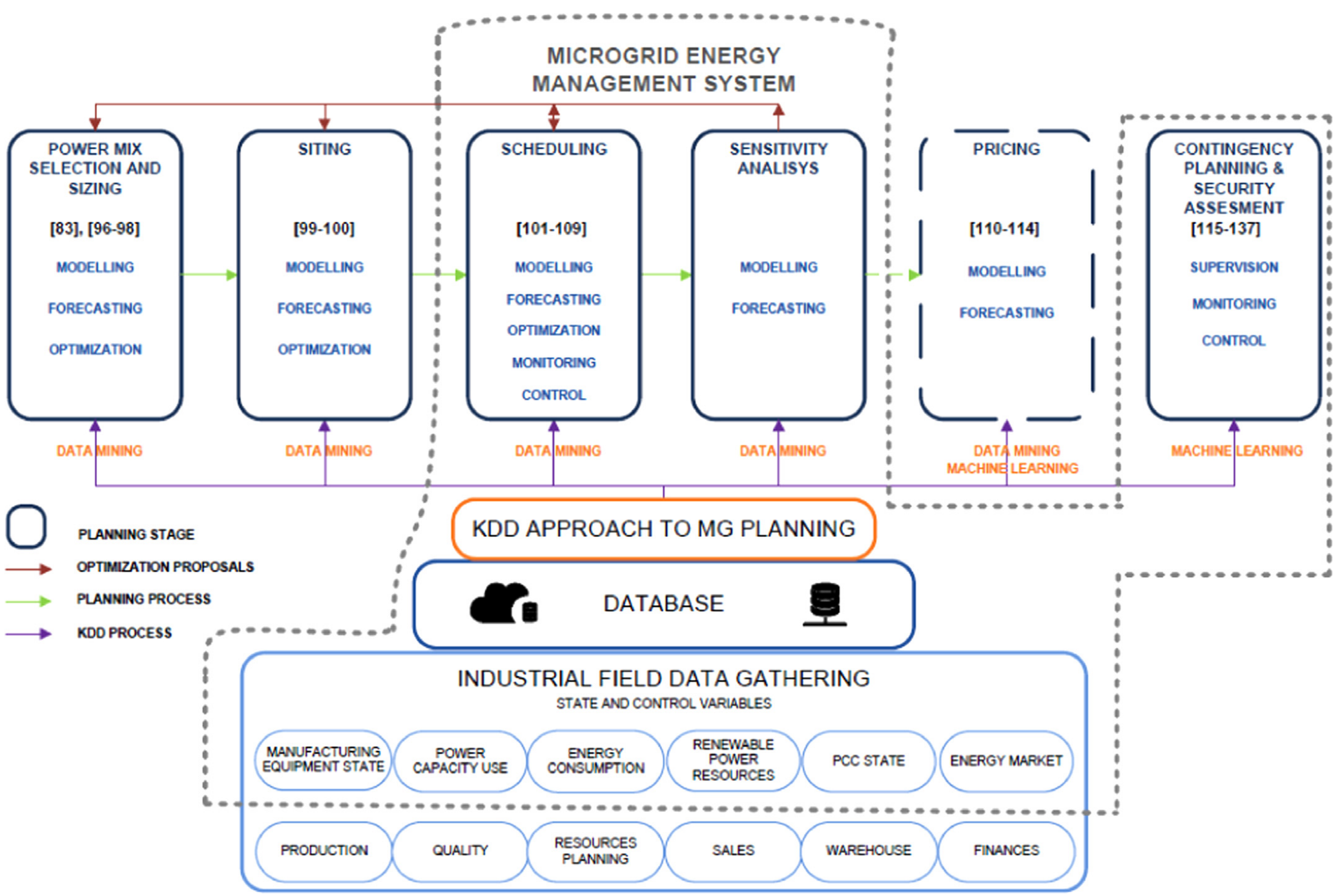

Fig. 3. A KDD-based approach to microgrid planning process scheme.

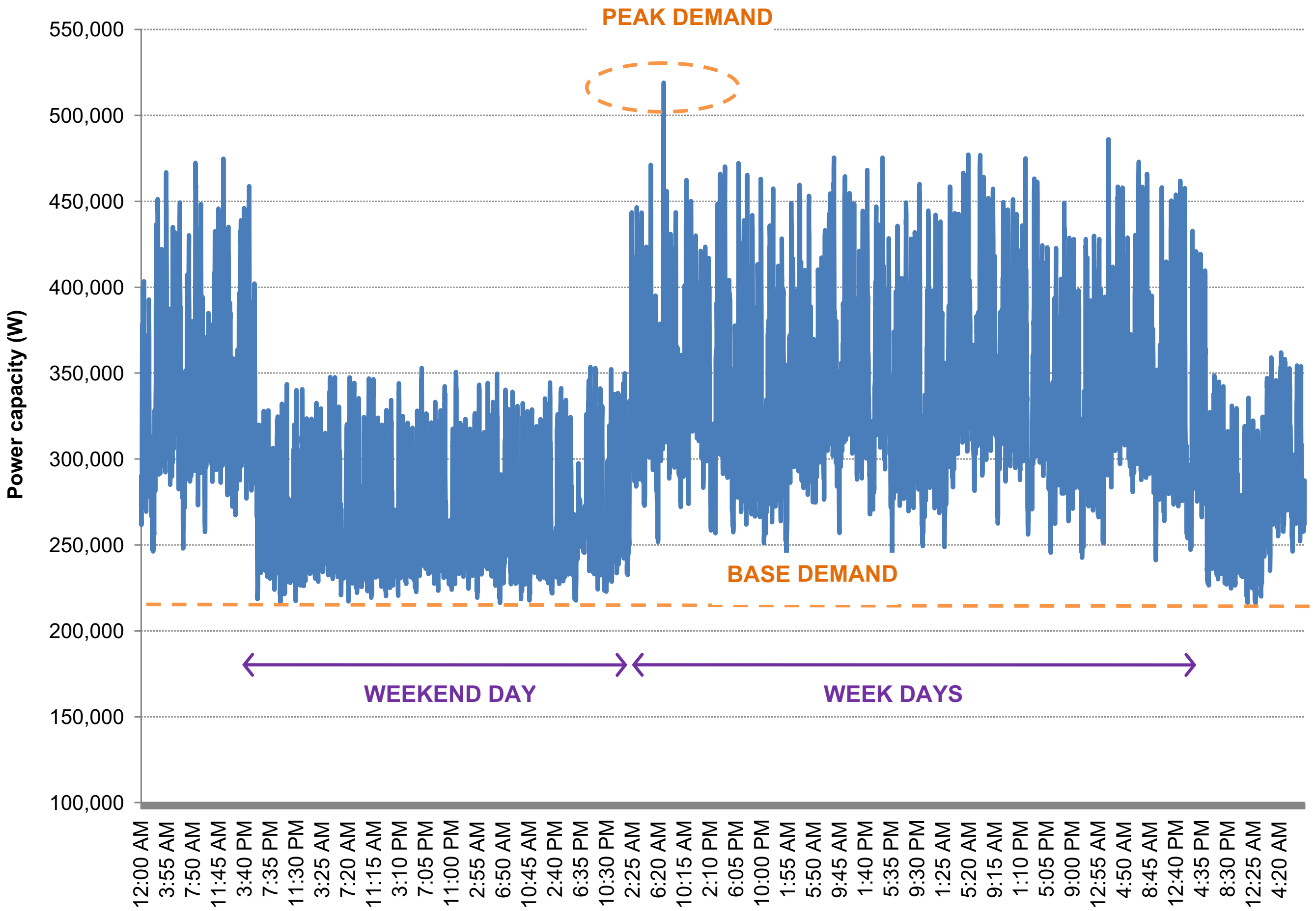

Fig. 4. Demand curve for a continuous manufacturing process.

allocation) of storage devices according to economic and environmental constraints. In fact, the economic viability of applications based on renewable power sources strongly depends on the required quantity and capacity of associated energy storage devices, due to their high initial and replacement costs.

- Other advanced modeling capabilities: the availability of raw data about a real process is a clear advantage in order to design, even to redesign an industrial power supply system. For example, an empirical analysis using real energy use data for smart grid planning, customer education and demand response strategies design is presented in [96]. In [97] the trade-off between the accuracy and the transparency of data mining-based models in the context of catastrophe predictors for power grid response-based remedial action schemes are studied. Regarding manufacturing processes, not every process keep design conditions along its entire lifecycle. Manufacturing facilities are usually modified in order to process new materials of even to manufacture new products. Hence, EMSs at industrial MGs should at least be able to identify optimal opportunities for power systems redesign, towards continue fulfilling the same (or even new) objectives about renewable power sources penetration, energy consumption, net-metering capacities and combined heat and power integration. Depending on the type of manufacturing process and its constraints, some demand side management and demand response strategies could be successfully applied.

4.2. Siting approaches

As it has been cited in the previous section, power quality and reliability are the main goals of siting problems in MG planning. The previous existence of electricity consumption and quality data about the manufacturing process is a clear advantage. Power quality problems will be, for sure, well defined since they cannot be only modeled, but also these models can be validated using real field data. In other words, the capacity and allocation requirements both for storage and power generation systems can be defined using optimization techniques on previously developed models.

Moreover, siting is also a critical problem in MG expansion. Considering a multi-plant supply or a competitive power market environment, consumption data that includes spatial and temporal characteristics is quite useful for power networks development and marketing strategies planning. The proposed spatial modeling approach is an exploratory data analysis, trying to discover useful patterns in spatial data that are not obvious to the data user and are useful in the spatial load forecast. For example, a modified form of the mountain method is adopted for cluster estimation [98]. In addition, other approaches to siting problems are solved using KDD techniques. Locational Marginal Pricing (LMP) is a mechanism for using market-based prices for managing transmission congestion. LMP resulting from bidding competition represents electrical and economical values at nodes or in areas that may provide economical indicator signals to the market agents. In [99] a data mining-based methodology that helps characterizing zonal prices in real power transmission networks is proposed. A two-step and k-means clustering algorithms are used in order to extract knowledge to support the investment and network-expansion planning.

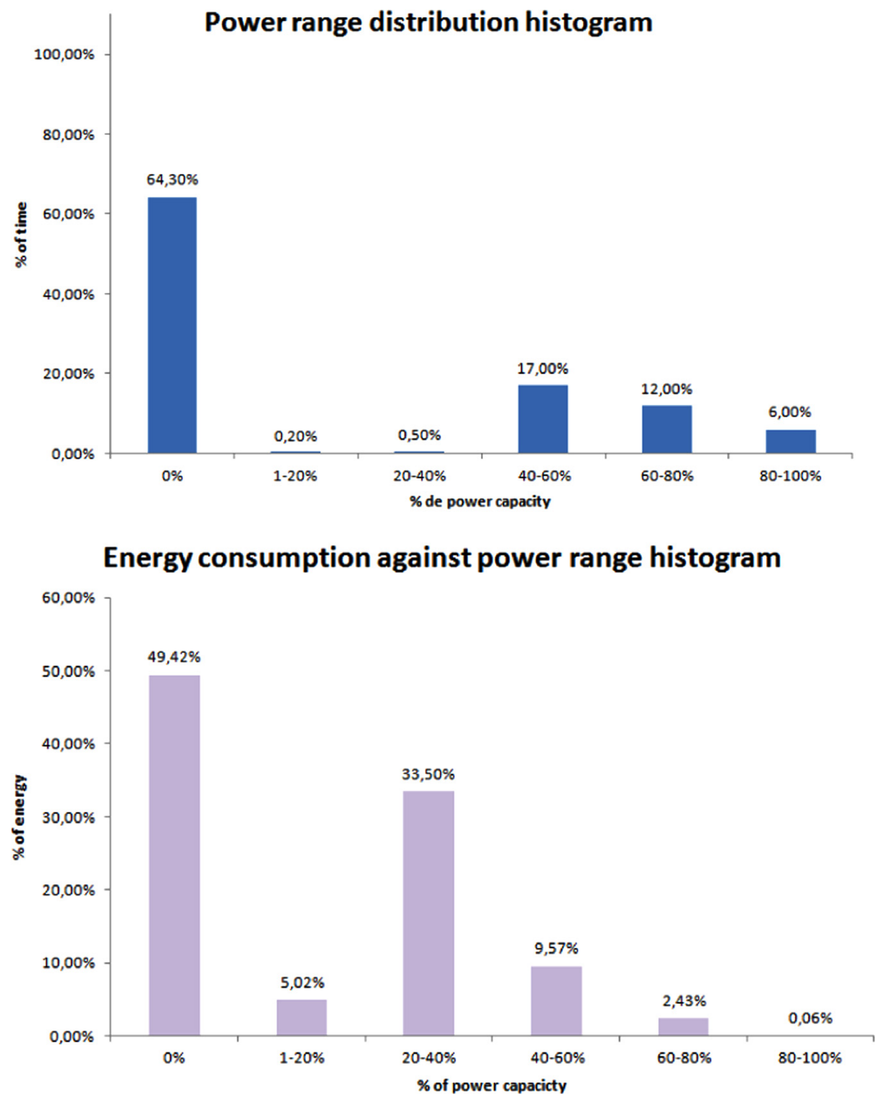

Fig. 5. Frequency histograms of a welding area in a heat exchanger manufacturing process.

4.3. *Scheduling approaches*

Perhaps the most powerful source of cost and environmental emissions savings, regarding MG planning and following a KDD approach, is resources scheduling. Real data from a manufacturing process is the basis of a robust manufacturing process model development. Once this model is developed, forecasting capabilities have proven to be very useful for scheduling tasks and savings achievement. Simulation and sensitive analysis could be used in order to find optimal work conditions both for power supply and manufacturing equipment. As it has been cited in the previous sections, the future of systems such as ERPs and EMSs is in integration and analytics. A combined analysis tool both for energy supply and manufacturing processes will allow the company to identify new performance improvement scenarios. Some of these future capacities for EMSs are:

- On-line monitoring and scheduling is also a desirable capability to integrate in these systems. Total time required to manufacture an item includes some specific intervals such as order preparation time, queue time, setup time, run time, move time, inspection time, and put-away time. Some of these times can be considered stand-by times from energy consumption point of view. Some stand-by periods could be considered wasted energy and should be avoided or minimized, if possible. Hence, start and stops for manufacturing and power equipment must be scheduled, turn off and turn on protocols must be followed and monitored. An example of unsupervised stops is shown in Fig. 6. It can be highlighted that usually power capacity during stand-by periods is never the same or even similar. In this line, a novel approach to the identification of the devices present in an electrical installation based on the measurement of current at the incoming supply point is presented in [100]. This new approach performs a black-box analysis to determine what devices are possibly present, just by taking measurements from the electrical wiring that leads to the black-box. Several neural network-based models were developed and tested for signature identification (pattern recognition) of electrical devices based on the current harmonics, even under noisy conditions. Once main consumptions are identified, power supply requirements can be modeled together with production process requirements. So, the whole system could be scheduled and results forecasted. For example, a real-time energy management problem is faced up in [101] and [102]. A research platform driven by an existing campus MG for developing (DM-based) predictive analytics for real-time energy management is presented in [102]. Meanwhile Wang et al. present in [101] an online algorithm for optimal real-time energy distribution. Four data mining approaches for wind turbine power curve monitoring are

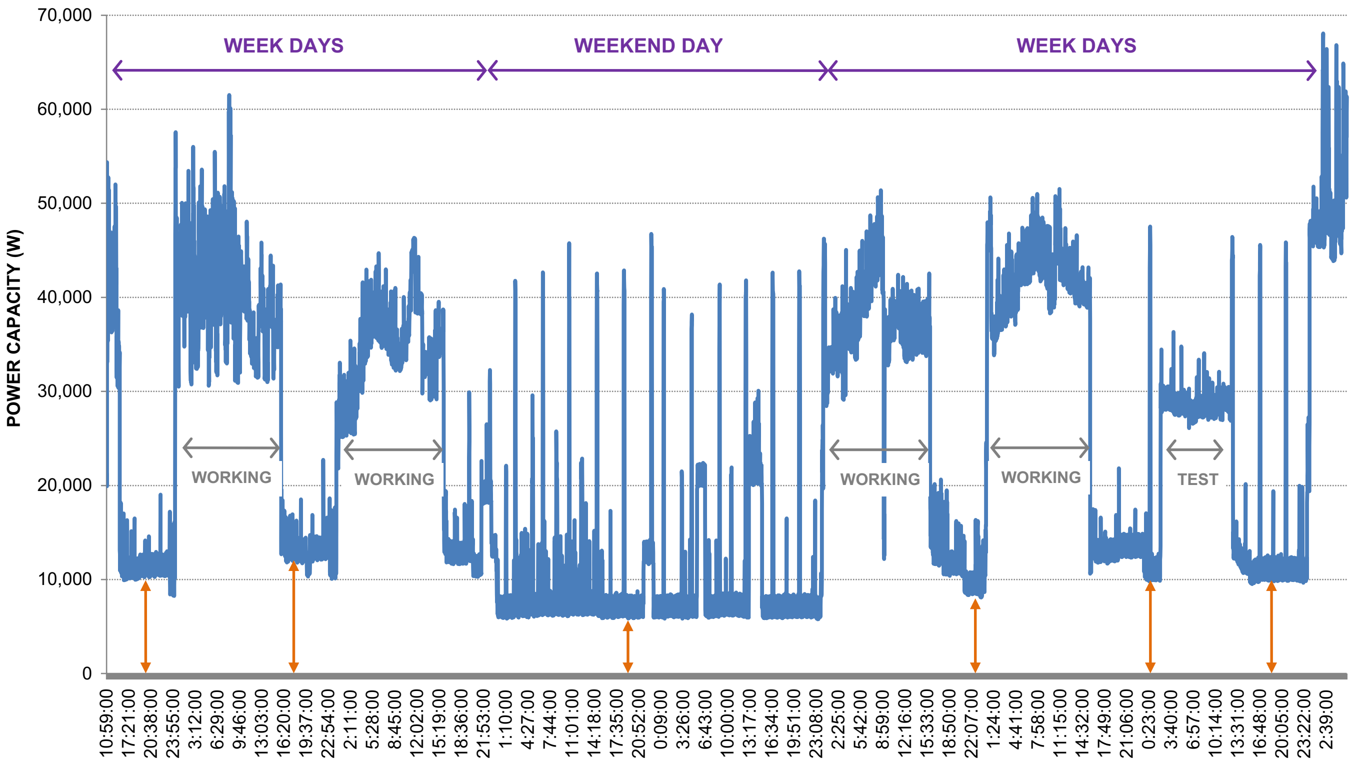

Fig. 6. Power capacity/energy consumption during stop and stand-by states in a manufacturing process.

compared in [103]. An optimal voltage control is proposed in for a MG [104] using data mining techniques: regression rules to estimate the optimum reactive power of the wind farms and classification trees to estimate the optimum transformer taps.

- Optimal supervision and control parameters: perhaps the most valuable knowledge that can be extracted from process data is partial and full load ratios regarding energy consumption. These real ratios can be calculated from raw data, but in some occasions some experiments must be forced in order to complete the scope of every sub-process. In a manufacturing process it is critical important to know performance ratios at the most common load percentages for every sub-process and ancillary service, including power generation equipment. Further relationship with other variables such as quality parameters should be studied in order to reject working parameters that do not fulfill process constraints. This data would allow energy planners to develop more effective scheduling. High percentages of energy savings could be reached, especially in oversized manufacturing and power equipment if additional quality and production constraints are fulfilled in time. Also equipment must allow load regulation. Following these guidelines, a supervision system could be designed in order to verify that some critical variables and index remains in a range of efficiency values and different constraints are fulfilled. If the system is working under inefficient scenarios, automatic (M2M) or semi-automatic (H2M) control strategies are suitable to be developed [105]. In [106] a formulation for energy management of a microgrid is proposed using ML techniques jointly with linear-programming-based multi-objective optimization, aiming to minimize the operation cost and the environmental impact of a microgrid. An artificial neural network ensemble is developed with demand and generation forecasting purposes, based on short term load forecasting and using a self-supervised adaptive neural network [107]. In [108] four time series models for different prediction horizons for a wind farm are built by data mining algorithms.

Among all the articles reviewed in order to write this paper, only Overturf et al. [109] combine DER (energy supply) an manufacturing process (production) planning for industrial companies. They develop a continuous improvement approach (which they call Sufficiency Kaizen) and conclude that taking control over the allocation and production of energy, corporations can obviate energy cost impact and risk, with relief immediate, and absolute impact over time.

4.4. Pricing approaches

Forecasting not only power generation, but also electricity prices plays a significant role in making optimal scheduling decisions in competitive electricity markets. Predominantly, price forecasting looks for the exact values of future prices. However, in some applications, such as demand-side management, operation decisions are based on certain price thresholds. Thus, it can be focused as an electricity price classification problem [110]. Energy price spike forecasting is studied in [111]. Support vector machine and probability classifier algorithms are chosen to be the spike occurrence predictors and realistic market data are used to test the proposed model. In [112] the application of classical techniques towards forecasting energy prices in combination with data mining neural networks yields a more accurate and realistic performance than conventional forecasting techniques.

Instead of reacting to changing market conditions system managers can adopt a fix price or a time of use tariff. But also they can be proactive in their approach to energy efficiency and management adopting demand side management and demand response strategies [113,114]. Advanced energy efficiency strategies such as net-metering could also be studied in order to generate profits for the company. As it has been mentioned in the previous section, some of these opportunities can derive on automatic or semi-automatic control strategies in order to reduce energy costs. But a manufacturing company must verify carefully the application of pricing strategies. *Production planners look for a* *stable environment for their process, and that is the reason why dynamic pricing and demand response strategies are not usually attractive for them, unless cost savings are high*. Additionally some energy efficiency scenarios can cause other cost rising, e.g. work at night will increase salaries in spite of decreasing energy costs.

4.5. Contingency planning and security assessment

KDD techniques can help not only to detect contingence situations but also to manage a system in an unexpected environment. ML techniques applied to power systems security assessment are described in [115]. Indeed security assessment is a very important task in MGs, and there exist several KDD-based applications in this field, summarized and referenced in Table 2.

5. Conclusions and future trends

Although manufacturing processes are among the most monitored and supervised, IoT development is pushing this tendency nowadays. Data gathered by manufacturing companies is strongly increasing, but, on the other side, software management tools are not usually able to make a deep analysis of this data. In other words, KDD techniques are not widely extended, and companies are not exploiting the full potential from their gathered data yet, especially regarding knowledge acquisition. As it has been addressed by Cios and Kurgan [25], the future of KDD techniques is in *achieving overall integration into other popular industrial standards*. Another very important issue is to *provide interoperability and compatibility between different software systems and platforms, which also concerns KDD-based applications. Such systems would serve end-users in automating, or more realistically semi-automating, work with the development of innovative DM and ML-based applications* [105]. Recent advances in innovative architectures and techniques, such as IoT and DM, make it possible for manufacturing management systems to evolve quickly towards forecasting, modeling and optimizing capabilities. These capabilities are expected to take part of manufacturing management systems software, such as MES, EMS and ERPs, soon.

From an energy efficiency point of view, it could be asserted that the full potential of energy saving for a manufacturing process can only be defined through a holistic and integrated analysis of the complete value chain of the plant: not only from a generation and distribution point of view, but also from consumption and manufacturing process requirements. Following this approach, KDD techniques and specially DM techniques can exploit data gathered by MES, EMS and ERPs. *The discovered information can be combined with the knowledge of industrial experts to uncover hidden saving potential, thereby contributing effectively in making better business decisions***.**

*At the same time, MGs are proving to be reliable and sustainable alternatives to traditional power systems***.** They are also called to be the evolution of the present power systems due to their scalability and their potential for smart capabilities integration. In this context, MG architectures must be considered in order to provide more reliable and cleaner energy to manufacturing processes, including the most energy-intensive ones. Manufacturing processes and microgrid systems share a common layer, the EMS. EMSs are expected to be under a strong development in future years due to the integration with KDD and artificial intelligence techniques and technologies.

Actually, manufacturing processes and MGs have many goals in common such as quality keeping, environmental impact minimization, cost minimization and energy efficiency. Under these circumstances, *the integration of MG architectures with manufacturing processes under a KDD approach has not been deeply analyzed***.** Many synergies are present**,** especially regarding the integration of both EMSs based on IoT and KDD techniques. This innovative approach is expected, at least, to reveal innovative cost-saving and environmental impact minimization opportunities for the whole system. At the same time, traditional planning process for an industrial microgrid could be modified regarding innovative KDD-based techniques, as it is proposed in Fig. 3.

Industrial processes have been selected for this approach due to they are among the most data-intensive and energy-intensive applications, and they are usually controlled by a single entity (usually a company). This same approach could be applied to other MG potential applications less intensive in data gathering or energy consumption, such as military and university campus, airports and hospitals. The results of viability studies will be conditioned by social, environmental and legal factors but specially by those regarding initial investment on data gathering, energy costs and energy consumption (such as type of facility, size, working hours in a year and location). *The search of suitable, sustainable and profitable scenarios for this integration, and specially the development of advanced EMSs (friendly and integrated software tools with smart capabilities) are some of the following stages for this line of research***.**

References

[1] Short J, Bohn R, Baru C. How much information? 2010: Report on enterprise server information. UCSD Glob Inf Ind Cent 2011. p. 1–38.

[2] Mattern F, Floerkemeier C. From the internet of computers to the internet of things. LNCS Lect Notes Comput Sci (Including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) 2010;6462:242–59. http://dx.doi.org/10.1007/978-3-642-17226-7_15.

[3] Akyildiz I, Jornet J. The Internet of nano-things. IEEE Wirel Commun 2010;17:58–63. http://dx.doi.org/10.1109/MWC.2010.5675779.

[4] Perera C, Liu CH, Jayawardena S, Chen M. A survey on internet of things from industrial market perspective. IEEE Access 2014;2:1660–79. http://dx.doi.org/10.1109/ACCESS.2015.2389854.

[5] Hilbert M, López P. The world's technological capacity to store, communicate, and compute information. Science 2011;332:60–5. http://dx.doi.org/10.1126/science.1200970.

[6] Bernstein A, Provost F, Hill S. Toward intelligent assistance for a data mining process: an ontology-based approach for cost-sensitive classification. IEEE Trans Knowl Data Eng 2005;17:503–18. http://dx.doi.org/10.1109/TKDE.2005.67.

[7] Matheus CJ, Chan PK, Piatetsky-Shapiro G. Systems for knowledge discovery in databases. IEEE Trans Knowl Data Eng 1993;5:903–13. http://dx.doi.org/10.1109/69.250073.

[8] Tsai CW, Lai CF, Chiang MC, Yang LT. Data mining for internet of things: a survey. IEEE Commun Surv Tutor 2014;16:77–97. http://dx.doi.org/10.1109/SURV.2013.103013.00206.

[9] Yoon JP, Kerschberg L. Framework for knowledge discovery and evolution in databases. IEEE Trans Knowl Data Eng 1993;5:973–9. http://dx.doi.org/10.1109/69.250080.

[10] Fayyad U, Piatetsky-Shapiro G, Smyth P. From data mining to knowledge discovery in databases. AI Mag 1996;17:37–53. http://dx.doi.org/10.1609/aimag.v17i3.1230.

[11] Last M, Klein Y, Kandel A. Knowledge discovery in time series databases. IEEE Trans Syst Man Cybern Part B Cybern 2001;31:160–9. http://dx.doi.org/10.1109/3477.907576.

[12] Fu TC. A review on time series data mining. Eng Appl Artif Intell 2011;24:164–81. http://dx.doi.org/10.1016/j.engappai.2010.09.007.

[13] Kurgan L a, Musilek P. A survey of knowledge discovery and data mining process models. Knowl Eng Rev 2006;21:1. http://dx.doi.org/10.1017/S0269888906000737.

[14] Gandomi A, Haider M. Beyond the hype: big data concepts, methods, and analytics. Int J Inf Manag 2015;35:137–44. http://dx.doi.org/10.1016/j.ijinfomgt.2014.10.007.

[15] Xiang Z, Schwartz Z, Gerdes JH, Uysal M. What can big data and text analytics tell us about hotel guest experience and satisfaction? Int J Hosp Manag 2015;44:120–30. http://dx.doi.org/10.1016/j.ijhm.2014.10.013.

[16] Apte C. Data mining: an industrial research perspective. IEEE Comput Sci Eng 1997;4:6–9. http://dx.doi.org/10.1109/99.609825.

[17] Li J. On optimal rule discovery. IEEE Trans Knowl Data Eng 2006;18:460–71. http://dx.doi.org/10.1109/TKDE.2006.1599385.

[18] Espejo PG, Ventura S, Herrera F. A survey on the application of genetic programming to classification. IEEE Trans Syst Man Cybern Part C Appl Rev 2010;40:121–44. http://dx.doi.org/10.1109/TSMCC.2009.2033566.

[19] An A, Chan C, Shan H, Cercone N, Ziarko W. Applying knowledge discovery to predict water-supply consumption. IEEE Expert Syst Their Appl 1997;12:72–8. http://dx.doi.org/10.1109/64.608199.
[20] Shin CK, Yun UT, Kim HK, Park SC. A hybrid approach of neural network and memory-based learning to data mining. IEEE Trans Neural Netw 2000;11:637–46. http://dx.doi.org/10.1109/CDC.2014.7039892.
[21] De La Iglesia B, Reynolds A. The use of meta-heuristic algorithms for data mining. In: Proc 1st Int Conf Inf Commun Technol ICICT 2005: 2005. p. 34–44. http://dx.doi.org/10.1109/ICICT.2005.1598541.
[22] Kulkarni RV, Member S, Förster A, Venayagamoorthy GK. Computational intelligence in wireless sensor networks : a survey. Commun Surv Tutor IEEE 2011;13:68–96. http://dx.doi.org/10.1109/SURV.2011.040310.00002.
[23] Mukhopadhyay A, Maulik U, Bandyopadhyay S, Coello CAC. A survey of multiobjective evolutionary algorithms for data mining: part i. IEEE Trans Evolut Comput 2014;18:4–19. http://dx.doi.org/10.1109/TEVC.2013.2290086.
[24] Mukhopadhyay A, Maulik U, Bandyopadhyay S, Coello CAC. Survey of multiobjective evolutionary algorithms for data mining: part II. IEEE Trans Evolut Comput 2014;18:25–35. http://dx.doi.org/10.1109/TEVC.2013.2290082.
[25] Cios KJ, Kurgan LA. Advanced techniques in knowledge discovery and data mining. http://dx.doi.org/10.1007/1-84628-183-0.
[26] European Commision. Factories of the Future. Web Page n.d. 〈http://ec.europa.eu/research/industrial_technologies/factories-of-the-future_en.html〉 [accessed 1.02.15].
[27] Bi Z, Xu L, Da, Wang C. Internet of things for enterprise systems of modern manufacturing. IEEE Trans Ind Inform 2014;10:1537–46. http://dx.doi.org/10.1109/TII.2014.2300338.
[28] Said O, Masud M. Towards internet of things: survey and future vision. Int J Comput Netw (IJCN) 2013;5:1–17.
[29] Fukatsu T, Nanseki T. Monitoring system for farming operations with wearable devices utilized sensor networks. Sensors (Basel) 2009;9:6171–84. http://dx.doi.org/10.3390/s90806171.
[30] Xu L, He W, Li S. Internet of things in industries: a survey. IEEE Trans Ind Inform 2014:1–11. http://dx.doi.org/10.1109/TII.2014.2300753.
[31] Dustdar S, Schulte S, Schuller RSSA. D. Plug-and-play Virtual factories. IEEE Comput Soc 2012. http://dx.doi.org/10.1109/1089-7801/12.
[32] Dutta D, Bose I. Managing a big data project: the case of Ramco Cements limited. Int J Prod Econ 2014. http://dx.doi.org/10.1016/j.ijpe.2014.12.032.
[33] Tseng TL, Huang CC. Capitalizing on knowledge: a novel approach to crucial-knowledge determination. IEEE Trans Syst Man Cybern Part A Syst Hum 2005;35:919–31. http://dx.doi.org/10.1109/TSMCA.2005.851294.
[34] Frawley WJ, Piatetsky-shapiro G, Matheus CJ. Knowledge discovery in databases : an overview. AI Mag 1992;13:57–70. http://dx.doi.org/10.1609/aimag.v13i3.1011.
[35] Pan SJ, Yang Q. A survey on transfer learning. IEEE Trans Knowl Data Eng 2010;22:1345–59. http://dx.doi.org/10.1109/TKDE.2009.191.
[36] Meidan Y, Lerner B, Rabinowitz G, Hassoun M. Cycle-time key factor identification and prediction in semiconductor manufacturing using machine learning and data mining. IEEE Trans Semicond Manuf 2011;24:237–48. http://dx.doi.org/10.1109/TSM.2011.2118775.
[37] Choudhary a K, Harding J a, Tiwari MK. Data mining in manufacturing: a review based on the kind of knowledge. J Intell Manuf 2009;20:501–21. http://dx.doi.org/10.1007/s10845-008-0145-x.
[38] Harding JA, Shahbaz M, Srinivas, Kusiak A. Data mining in manufacturing: a review. J Manuf Sci Eng 2006;128:969. http://dx.doi.org/10.1115/1.2194554.
[39] Wang KWK, Tong STS, Eynard B, Roucoules L, Matta N. Review on application of data mining in product design and manufacturing. In: Proceedings of the fourth int conf fuzzy syst knowl discov (FSKD 2007): 2007;4. p. 613–8. http://dx.doi.org/10.1109/FSKD.2007.482.
[40] Liao SH, Chu PH, Hsiao PY. Data mining techniques and applications – A decade review from 2000 to 2011. Expert Syst Appl 2012;39:11303–11. http://dx.doi.org/10.1016/j.eswa.2012.02.063.
[41] He T. Data mining for the optimization in manufacturing process. In: Proc - 2010 2nd int work intell syst appl ISA 2010: 2010: p. 5–8. http://dx.doi.org/10.1109/IWISA.2010.5473459.
[42] Casali A, Ernst C. Discovering correlated parameters in semiconductor manufacturing processes: a data mining approach. IEEE Trans Semicond Manuf 2012;25:118–27. http://dx.doi.org/10.1109/TSM.2011.2171375.
[43] Shukla SK, Tiwari MK. GA guided cluster based fuzzy decision tree for reactive ion etching modeling: a data mining approach. IEEE Trans Semicond Manuf 2012;25:45–56. http://dx.doi.org/10.1109/TSM.2011.2173372.
[44] Kwak C, Yih Y. Data-mining approach to production control in the computer-integrated testing cell. IEEE Trans Robot Autom 2004;20:107–16. http://dx.doi.org/10.1109/TRA.2003.819595.
[45] Tsuda H, Shirai H. Improvement of photolithography process by second generation data mining. IEEE Trans Semicond Manuf 2007;20:239–44. http://dx.doi.org/10.1109/TSM.2007.901839.
[46] Evans R, Dore J, Van Voorthuysen E, Zhu J, Green MA. Data mining photovoltaic cell manufacturing data. In: Proceedings of the IEEE 40th photovolt. spec. conf.: 2014. p. 2699–704. http://dx.doi.org/10.1109/PVSC.2014.6925486.
[47] Chien C, Chuang S. A framework for root cause detection of sub-batch processing system for semiconductor manufacturing big data analytics. IEEE Trans Semicond Manuf 2014;27:475–88. http://dx.doi.org/10.1109/TSM.2014.2356555.
[48] Braha D, Shmilovici A. On the use of decision tree induction for discovery of interactions in a photolithographic process. IEEE Trans Semicond Manuf 2003;16:644–52. http://dx.doi.org/10.1109/TSM.2003.818959.
[49] Köksal G, Batmaz I, Testik MC. A review of data mining applications for quality improvement in manufacturing industry. Expert Syst Appl 2011;38:13448–67. http://dx.doi.org/10.1016/j.eswa.2011.04.063.
[50] Cai-yan LCL, You-fa SYS. Application of data mining in production quality management. In: Proceedings of the third int symp intell inf technol appl: 2009; 2. p. 285–8. http://dx.doi.org/10.1109/IITA.2009.81.
[51] Chang CW, Chao TM, Horng JT, Lu CF, Yeh RH. Development pattern recognition model for the classification of circuit probe wafer maps on semiconductors. IEEE Trans Compon Packag Manuf Technol 2012;2:2089–97. http://dx.doi.org/10.1109/TCPMT.2012.2215327.
[52] Gao L-HGL-H, Luo S-XLS-X, Qiu J-QQJ-Q, Li F-CLF-C, Ran H-CRH-C. Data mining on a kind of complex industrial process. In: Proc 2003 int conf mach learn cybern (IEEE Cat No03EX693): 2003;1. p. 105–7. 〈http://dx.doi.org/10.1109/ICMLC.2003.1264451〉.
[53] Hessinger U, Chan WK, Schafman BT. Data mining for significance in yield-defect correlation analysis. IEEE Trans Semicond Manuf 2014;27:347–56. http://dx.doi.org/10.1109/TSM.2014.2337251.
[54] Huang CCHCC, Fan YNFYN, Tseng TLTTL, Lee CHLCH, Chuang HFCHF. A hybrid data mining approach to quality assurance of manufacturing process. In: Proc IEEE int conf fuzzy syst (IEEE World Congr Comput Intell): 2008. p. 818–25. http://dx.doi.org/10.1109/FUZZY.2008.4630465.
[55] Kamal AMM. A data mining approach for improving manufacturing processes quality control. In: Proc 2nd int conf, next gener inf technol (ICNIT): 2011; p. 207–12.
[56] Karim MA, Russ G, Islam A. Detection of faulty products using data mining. In: Proc 11th int conf comput inf technol: 2008;00. p. 101–7. http://dx.doi.org/10.1109/ICCITECHN.2008.4803116.
[57] Krömer P, Platos J, Snášel V. Mining multi-class industrial data with evolutionary fuzzy rules. In: Proc IEEE int conf cybern CYBCONF: 2013. p. 191–6. http://dx.doi.org/10.1109/CYBConf.2013.6617453.
[58] Kusiak a. Decomposition in data mining: an industrial case study. IEEE Trans Electron Packag Manuf 2000;23:345–53. http://dx.doi.org/10.1109/6104.895081.
[59] Li Z, Baseman RJ, Zhu Y, Tipu F a, Slonim N, Shpigelman L. A unified framework for outlier detection in trace data analysis. IEEE Trans Semicond Manuf 2014;27:95–103. http://dx.doi.org/10.1109/TSM.2013.2267937.
[60] Nelson PC, Tirpak TM, Lane S a. An intelligent data mining system for drop test analysis of electronic products. IEEE Trans Electron Packag Manuf 2001;24:222–31. http://dx.doi.org/10.1109/6104.956808.
[61] Ooi MPL, Joo EKJ, Kuang YC, Demidenko S, Kleeman L, Chan CWK. Getting more from the semiconductor test: Data mining with defect-cluster extraction. IEEE Trans Instrum Meas 2011;60:3300–17. http://dx.doi.org/10.1109/TIM.2011.2122430.
[62] Sandborn P a, Mauro F, Knox R. A data mining based approach to electronic part obsolescence forecasting. IEEE Trans Components Packag Technol 2007;30:397–401. http://dx.doi.org/10.1109/TCAPT.2007.900058.
[63] Zhang F, Luk T. A data mining algorithm for monitoring PCB assembly quality. IEEE Trans Electron Packag Manuf 2007;30:299–305. http://dx.doi.org/10.1109/TEPM.2007.907576.
[64] Huang CC, Liang WY. Explication and sharing of design knowledge through a novel product design approach. IEEE Trans Syst Man Cybern Part C Appl Rev 2006;36:426–38. http://dx.doi.org/10.1016/j.jvs.2005.12.031.
[65] Zhuo CZC, Bingtu YBY, Yinglong WYW, Zefeng SZS. A new model for the industrial process control based on Data Mining. In: Proceedings of the chinese control decis conf: 2008. p. 1368–70. http://dx.doi.org/10.1109/CCDC.2008.4597541.
[66] Kusiak A, Song Z. Combustion efficiency optimization and virtual testing: a data-mining approach. IEEE Trans Ind Inform 2006;2:176–84. http://dx.doi.org/10.1109/TII.2006.873598.
[67] Song Z, Kusiak A. Constraint-based control of boiler efficiency: a data-mining approach. IEEE Trans Ind Inform 2007;3:73–83. http://dx.doi.org/10.1109/TII.2006.890530.
[68] Kusiak a, Zhe Song. Clustering-based performance optimization of the boiler–turbine system. IEEE Trans Energy Convers 2008;23:651–8. http://dx.doi.org/10.1109/TEC.2007.914183.
[69] Ma Y, Wu H, Wang L, Huang B, Ranjan R, Zomaya A. Remote sensing big data computing : challenges and opportunities. Futur Gener Comput Syst 2014. http://dx.doi.org/10.1016/j.future.2014.10.029.
[70] ISO. The ISO Survey of Management System Standard Certifications – 2014 Executive summary.
[71] Maricar NM. Industrial energy audit web application using data mining model. In: Proc IEEE, GCC Conf (GCC): 2006. p. 1–6. http://dx.doi.org/10.1109/IEEEGCC.2006.5686199.
[72] Dettmer R. Unlocking the power of the grid. IEE Rev 2002;48:9. http://dx.doi.org/10.1049/ir:20020301.
[73] Bui N, Castellani AP, Casari P, Zorzi M. The internet of energy: a web-enabled smart grid system. IEEE Netw 2012;26:39–45. http://dx.doi.org/10.1109/MNET.2012.6246751.
[74] Simmhan Y, Aman S, Kumbhare A, Liu R, Stevens S, Zhou Q. Cloud-based software platform for data-driven smart grid management. Comput Sci Eng 2013;15:1–11.
[75] GTMResearch. The soft grid 2013–2020: Big data & utility analytics for smart grid: 2013.

[76] Simmhan Y, Aman S, Kumbhare A, Liu R, Stevens S, Zhou Q, et al. Cloud-based software platform for big data analytics in smart grids. Comput Sci Eng 2013;15:38–47. http://dx.doi.org/10.1109/MCSE.2013.39.
[77] Yun M, Yuxin B. Research on the architecture and key technology of internet of things (IoT) applied on smart grid. In: Proc int conf adv energy eng (ICAEE): 2010. p. 69–72. http://dx.doi.org/10.1109/ICAEE.2010.5557611.
[78] Madan S, Son W-KSW-K, Bollinger KE. Applications of data mining for power systems. In: Proceedings of the CCECE '97 can conf electr comput eng, eng innov voyag discov conf: 1997;2. p. 403–6. http://dx.doi.org/10.1109/CCECE.1997.608243.
[79] Mori H. State-of-the-art overview on data mining in power systems. In: Proc Power Syst Conf Expo 2006 PSCE '06 2006 IEEE PES: 2006. p. 33–4. http://dx.doi.org/10.1109/PSCE.2006.296245.
[80] Vale ZA, Ramos C, Ramos S, Pinto T. Data mining applications in power systems: case-studies and future trends. In: Proc 2009 Transm Distrib Conf Expo Asia Pacific 2009. p. 1–4. http://dx.doi.org/10.1109/TD-ASIA.2009.5356830.
[81] Kazerooni M, Zhu H, Overbye TJ. Literature review on the applications of data mining in power systems. In: Proc Power Energy Conf Illinois: 2014. p. 1–8. http://dx.doi.org/10.1109/PECI.2014.6804567.
[82] Rana MM, Li L. Microgrid state estimation and control for smart grid and internet of things communication network. Electron Lett 2015;51:149–51. http://dx.doi.org/10.1049/el.2014.3635.
[83] Silva D, De, Yu X. A data mining framework for electricity consumption analysis from meter data. IEEE Trans Ind Inform 2011;7:399–407. http://dx.doi.org/10.1109/TII.2011.2158844.
[84] IBM. Managing big data for smart grids and smart meters: 2012. http://dx.doi.org/IMW14628-USEN-00.
[85] Pipattanasomporn M, Willingham M, Rahman S. Implications of on-site distributed generation for commercial/industrial facilities. IEEE Trans Power Syst 2005;20:206–12. http://dx.doi.org/10.1109/TPWRS.2004.841233.
[86] Leif Hanrahan B, Lightbody G, Staudt L, Leahy PG. A powerful visualization technique for electricity supply and demand at industrial sites with combined heat and power and wind generation. Renew Sustain Energy Rev 2014;31:860–9. http://dx.doi.org/10.1016/j.rser.2013.12.016.
[87] Bracco S, Delfino F, Pampararo F, Robba M, Rossi M. A mathematical model for the optimal operation of the university of genoa smart polygeneration microgrid: evaluation of technical, economic and environmental performance indicators. Energy 2014;64:912–22. http://dx.doi.org/10.1016/j.energy.2013.10.039.
[88] Soshinskaya M, Crijns-graus WHJ, Van Der Meer J, Guerrero JM. Application of a microgrid with renewables for a water treatment plant. Appl Energy 2014;134:20–34. http://dx.doi.org/10.1016/j.apenergy.2014.07.097.
[89] Pinceti P, Vanti M, Brocca C, Carnesecchi M, Macera GP. Design criteria for a power management system for microgrids with renewable sources. Electr Power Syst Res 2015;122:168–79. http://dx.doi.org/10.1016/j.epsr.2015.01.010.
[90] Gamarra C, Guerrero JM. Computational optimization techniques applied to microgrids planning: a review. Renew Sustain Energy Rev 2015;48:413–24. http://dx.doi.org/10.1016/j.rser.2015.04.025.
[91] Lilienthal P. How to classify microgrids: setting the stage for a distributed generation energy future. 〈Http://blog.homerenergy.com/how-to-Classify-Microgrids-Setting-the-Stage-for-a-Distributed-Generation-Energy-Future/n.d〉 Last time accessed: Jan. 2016.
[92] Farzan F, Lahiri S, Kleinberg M, Gharieh K, Farzan F, Jafari M. Microgrids for fun and profit: the economics of installation investments and operations. IEEE Power Energy Mag 2013;11:52–8. http://dx.doi.org/10.1109/MPE.2013.2258282.
[93] Asano H, Bando S. Economic evaluation of microgrids. In: Proc IEEE power energy soc 2008 gen meet convers deliv electr energy 21st Century, PES: 2008. p. 1–6. http://dx.doi.org/10.1109/PES.2008.4596603.
[94] Morris GY, Abbey C, Wong S, Joos G. Evaluation of the costs and benefits of microgrids with consideration of services beyond energy supply. In: Proc IEEE power energy soc gen meet: 2012. p. 1–9. doi:10.1109/PESGM.2012.6345380.
[95] Zhang D, Samsatli NJ, Hawkes AD, Brett DJL, Shah N, Papageorgiou LG. Fair electricity transfer price and unit capacity selection for microgrids. Energy Econ 2013;36:581–93. http://dx.doi.org/10.1016/j.eneco.2012.11.005.
[96] Aman S, Member S, Simmhan Y, Member S. Holistic measures for evaluating prediction models in smart grids. IEEE Trans Knowl Data Eng 2015;27:1–15. http://dx.doi.org/10.1109/TKDE.2014.2327022.
[97] Kamwa I, Samantaray SR, Joós G. On the accuracy versus transparency trade-off of data-mining models for fast-response PMU-based catastrophe predictors. IEEE Trans Smart Grid 2012;3:152–61. http://dx.doi.org/10.1109/TSG.2011.2164948.
[98] Wu Hung-Chih, Lu Chan-Nan. A data mining approach for spatial modeling in small area load forecast. IEEE Trans Power Syst 2002;17:516–21. http://dx.doi.org/10.1109/TPWRS.2002.1007927.
[99] Ferreira J, Ramos S, Vale Z, Soares J. A data-mining-based methodology for transmission expansion planning. IEEE Intell Syst 2011;26:28–37. http://dx.doi.org/10.1109/MIS.2011.4.
[100] Srinivasan D, Ng WS, Liew a C. Neural-network-based signature recognition for harmonic source identification. IEEE Trans Power Deliv 2006;21:398–405. http://dx.doi.org/10.1109/TPWRD.2005.852370.
[101] Wang Y, Mao S, Nelms RM. Online algorithm for optimal real-time energy distribution in the smart grid. IEEE Trans Emerg Top Comput 2013;1:10–21. http://dx.doi.org/10.1109/TETC.2013.2273218.
[102] Balac N, Sipes T, Wolter N, Nunes K, Sinkovits B, Karimabadi H. Large Scale predictive analytics for real-time energy management. In: Proc IEEE int conf big data, big data 2013: 2013. p. 657–64. http://dx.doi.org/10.1109/BigData.2013.6691635.
[103] Schlechtingen M, Santos IF, Achiche S. Using data-mining approaches for wind turbine power curve monitoring: a comparative study. IEEE Trans Sustain Energy 2013;4:671–9. http://dx.doi.org/10.1109/TSTE.2013.2241797.
[104] Egido I, Sáiz-Marín E, Lobato E. Optimal voltage control by wind farms using data mining techniques. IET Renew Power Gener 2013:141–50. http://dx.doi.org/10.1049/iet-rpg.2013.0025.
[105] Chan PPK, Chen WC, Ng WWY, Yeung DS. Multiple classifier system for short term load forecast of Microgrid. In: Proc int conf mach learn cybern: 2011;3. p. 1268–73. http://dx.doi.org/10.1109/ICMLC.2011.6016936.
[106] Chaouachi A, Kamel RM, Andoulsi R, Nagasaka K. Multiobjective intelligent energy management for a microgrid. IEEE Trans Ind Electron 2013;60:1688–99. http://dx.doi.org/10.1109/TIE.2012.2188873.
[107] Yoo H, Pimmel RL. Short term load forecasting using a self-supervised adaptive neural network. IEEE Trans Power Syst 1999;14:779–84. http://dx.doi.org/10.1109/59.761912.
[108] Kusiak A, Zheng H, Song Z. Short-term prediction of wind farm power: a data mining approach. IEEE Trans Energy Convers 2009;24:125–36. http://dx.doi.org/10.1109/TEC.2008.2006552.
[109] Overturf MC, Chrosny WM, Stine R. Energy sufficiency kaizen achieving zero energy cost and variance with sufficiency-inclusive private energy portfolios. Distrib Gener Altern Energy J 2011;26:36–56. http://dx.doi.org/10.1080/21563306.2011.10462203.
[110] Huang D, Zareipour H, Rosehart WD, Amjady N. Data mining for electricity price classification and the application to demand-side management. IEEE Trans Smart Grid 2012;3:808–17. http://dx.doi.org/10.1109/TSG.2011.2177870.
[111] Zhao JH, Dong ZY, Li X, Wong KP. A framework for electricity price spike analysis with advanced data mining methods. IEEE Trans Power Syst 2007;22:376–85. http://dx.doi.org/10.1109/TPWRS.2006.889139.
[112] Garcia MP, Kirschen DS. Forecasting system imbalance volumes in competitive electricity markets. IEEE Trans Power Syst 2006;21:240–8. http://dx.doi.org/10.1109/TPWRS.2005.860924.
[113] Mazidi M, Zakariazadeh A, Jadid S, Siano P. Integrated scheduling of renewable generation and demand response programs in a microgrid. Energy Convers Manag 2014;86:1118–27. http://dx.doi.org/10.1016/j.enconman.2014.06.078.
[114] Zakariazadeh A, Jadid S, Siano P. Smart microgrid energy and reserve scheduling with demand response using stochastic optimization. Int J Electr Power Energy Syst 2014;63:523–33. http://dx.doi.org/10.1016/j.ijepes.2014.06.037.
[115] Wehenkel L. Machine learning approaches to power-system security assessment. IEEE Expert 1997:12. http://dx.doi.org/10.1109/64.621229.
[116] He M, Zhang J, Vittal V. Robust online dynamic security assessment using adaptive ensemble decision-tree learning. IEEE Trans Power Syst 2013;28:4089–98. http://dx.doi.org/10.1109/TPWRS.2013.2266617.
[117] Huang SJ, Lin JM. Enhancement of anomalous data mining in power system predicting-aided state estimation. IEEE Trans Power Syst 2004;19:610–9. http://dx.doi.org/10.1109/TPWRS.2003.818726.
[118] Kaci A, Kamwa I, Dessaint L a, Guillon S. Synchrophasor data baselining and mining for online monitoring of dynamic security limits. IEEE Trans Power Syst 2014;29:2681–95. http://dx.doi.org/10.1109/TPWRS.2014.2312418.
[119] Pan S, Morris T, Adhikari U. Developing a hybrid intrusion detection system using data mining for power systems. IEEE Trans Smart Grid 2015. http://dx.doi.org/10.1109/TSG.2015.2409775.
[120] Tso SK, Lin JK, Ho HK, Mak CM, Yung KM, Ho YK. Data mining for detection of sensitive buses and influential buses in a power system subjected to disturbances. IEEE Trans Power Syst 2004;19:563–8. http://dx.doi.org/10.1109/TPWRS.2003.821479.
[121] Othman ML, Aris I, Abdullah SM, Ali ML, Othman MR. Knowledge discovery in distance relay event report: a comparative data-mining strategy of rough set theory with decision tree. IEEE Trans Power Deliv 2010;25:2264–87. http://dx.doi.org/10.1109/TPWRD.2010.2055587.
[122] Gaouda a M, Kanoun SH, Salama MM a, Chikhani a Y. Pattern recognition applications for power system disturbance classification. IEEE Trans Power Deliv 2002;17:677–83. http://dx.doi.org/10.1109/TPWRD.2002.1022786.
[123] Zeineldin HH, Kan'an NH, Casagrande E, Woon WL. Data mining approach to fault detection for isolated inverter-based microgrids. IET Gener Transm Distrib 2013;7:745–54. http://dx.doi.org/10.1049/iet-gtd.2012.0518.
[124] Xu LX Le, Chow M-YCM-Y, Taylor LS. Using the data mining based fuzzy classification algorithm for power distribution fault cause identification with imbalanced data. In: Proc IEEE PES power syst conf expo: 2006;22. p. 164–171. http://dx.doi.org/10.1109/PSCE.2006.296482.
[125] Bernabeu EE, Thorp JS, Centeno V. Methodology for a security/dependability adaptive protection scheme based on data mining. IEEE Trans Power Deliv 2012;27:104–11. http://dx.doi.org/10.1109/TPWRD.2011.2168831.
[126] Hor CL, Crossley P a. Unsupervised event extraction within substations using rough classification. IEEE Trans Power Deliv 2006;21:1809–16. http://dx.doi.org/10.1109/TPWRD.2006.874670.
[127] Strachan SM, McArthur SDJ, Stephen B, McDonald JR, Campbell A. Providing decision support for the condition-based maintenance of circuit breakers through data mining of trip coil current signatures. IEEE Trans Power Deliv 2007;22:178–86. http://dx.doi.org/10.1109/TPWRD.2006.883001.
[128] Huang SJ, Lin JM. Enhancement of power system data debugging using gap statistic algorithm-based data mining technique. IEEE Power Eng Rev 2002. http://dx.doi.org/10.1109/MPER.2002.4311742.
[129] So MK, Lau WH, Norman CFT, Tsang KF, Leung MC. Locating major harmonic current contributors by layered knowledge discovery in database. In: Proc

IEEE power energy soc gen meet: 2011. p. 1–5. http://dx.doi.org/10.1109/PES.2011.6039139.
[130] Asheibi A, Stirling D, Sutanto D. Analyzing harmonic monitoring data using supervised and unsupervised learning. IEEE Trans Power Deliv 2009;24:293–301. http://dx.doi.org/10.1109/TPWRD.2008.2002654.
[131] El-arroudi K, Joós G, Kamwa I. Intelligent-based approach to islanding detection in distributed generation. IEEE Trans Power Deliv 2007;22:828–35.
[132] Bitaraf H, Sheikholeslamzadeh M, Ranjbar AM, Mozafari B. A novel SVM approach of islanding detection in micro grid. In: Proc 2012 IEEE innov smart grid technol; Asia, ISGT Asia: 2012. p. 1–5. http://dx.doi.org/10.1109/ISGT-Asia.2012.6303335.
[133] Guo T, Milanovic JV. Probabilistic framework for assessing the accuracy of data mining tool for online prediction of transient stability. IEEE Trans Power Syst 2014;29:377–85. http://dx.doi.org/10.1109/TPWRS.2013.2281118.
[134] Casagrande E, Woon WL, Zeineldin HH, Svetinovic D. A differential sequence component protection scheme for microgrids with inverter-based distributed generators. IEEE Trans Smart Grid 2014;5:29–37. http://dx.doi.org/10.1109/TSG.2013.2251017.
[135] Kar S, Samantaray SR, Zadeh MD. Data-mining model based intelligent differential microgrid protection scheme. IEEE Syst J 2015:1–9. http://dx.doi.org/10.1109/JSYST.2014.2380432.
[136] Kar S, Samantaray SR. Combined S-transform and data-mining based intelligent micro-grid protection scheme. In: Proc students conf. eng. syst.: 2014, p. 1–5. http://dx.doi.org/10.1109/SCES.2014.6880053.
[137] Nizar AH, Dong ZY, Wang Y. Power utility nontechnical loss analysis with extreme learning machine method. IEEE Trans Power Syst 2008;23:946–55. http://dx.doi.org/10.1109/TPWRS.2008.926431.