École doctorale IAEM Lorraine

Methodology for an Augmented Business Process Management in IoT Environment

THÈSE

présentée et soutenue publiquement le 07 décembre 2022

pour l'obtention du

Doctorat de l'Université de Lorraine en cotutelle avec l'Université Mohammed V de Rabat

(mention informatique)

par

Abir Ismaili-Alaoui

Composition du jury

Président :	Kamel Smaïli	Professeur à l'Université de Lorraine, Nancy, France
Rapporteurs :	Walid Gaaloul	Professeur à Télécom SudParis, Inst Polytechnique de Paris, France
	Selmin Nurcan	MCF HDR, à l'Université Paris 1 - Panthéon - Sorbonne, France
	Nissrine Souissi	Professeur à l'Ecole Nationale Supérieure des Mines de Rabat, Maroc
Examinateur :	Hanan El Bakkali	Professeur à ENSIAS, Université Mohammed V de Rabat, Maroc
Invité :	Jamal Baïna	Expert Développement des Usages, à Moselle Fibre, Metz, France
Encadrants :	Karim Baïna	Professeur à ENSIAS, Université Mohammed V de Rabat, Maroc
	Khalid Benali	MCF, HDR, à l'Université de Lorraine, Nancy, France

Laboratoire Lorrain de Recherche en Informatique et ses Applications — UMR 7503

Mis en page avec la classe thesul.

Acknowledgments

No duty is more urgent than giving thanks.

James Allen

Knowing your 'WHY' and your own motivation factors is very important during a PhD. PhD is a challenging journey at different levels (intellectual, emotional, mental, etc.). At certain points, knowing your "why" and your strong interest and motivation for the research field could be the only catalyst that potentially helps you overcome and alleviate the pressures of the challenges you are facing. At the end of this journey, you realize that you have grown as a person, because a PhD is a gradual process influenced by an almost endless sequence of experiences and challenges. Fortunately, I was surrounded by many people who made this journey less solitary and less difficult, but most importantly more enriching, more valuable and more meaningful. The number of people I should thank for supporting me in achieving this goal is huge. Each of them, in their own way, has left their fingerprints on this journey. I am grateful for all those who have shared with me their knowledge, experience, passion, support, and laughter.

First and foremost, my genuine gratitude goes to my supervisors, Prof. Karim Baïna and Dr. Khalid Benali for their continuous support of my PhD work, their patience, dedication, communicative optimism and encouragement (especially amid the COVID19 pandemic). I would like to express my sincere gratitude to them for giving me this opportunity and providing invaluable and punctual guidance throughout this thesis. Our meetings were always excessively beneficial to my work.

I would like to express my sincere gratitude to my thesis examiners and the defense jury committee. Having the opportunity to present my thesis before you is a privilege. I appreciate the efforts of Prof. Walid Gaaloul, Dr. Selmin Nurcan and Prof. Nissrine Souissi for reviewing my dissertation. I truly appreciate their constructive feedback and time that they spare given their busy schedules. I also deliver my sincere gratitude to Prof. Hanan EL Bekkali for examining my work. I would like also to thank Prof. Kamel Smaïli, I really value your efforts, time, and constructive feedback during my PhD as my scientific referent.

A special thanks to all the COAST team members for their sympathy on a daily basis, and to our team leader Prof. François Charoy, thank you for organizing online meetings during the COVID time to check on each one of us, and thank you for the annual events like "Christmas Couscous" that allowed us to spend convivial moments with all the members of the team outside the working environment.

I also thank all the members of Rabat IT Center at ENSIAS, Mohammed V University, and all the "Innovation on Digital and Enterprise Architectures" team members, particularly Pr. Laila Kjiri.

I also thank the French Embassy in Morocco for their financial support during the first three years of my Ph.D.

A special thanks to Dr. Jamal Baïna and Pol Reitz. Thank you for having welcomed me into Angel Assistance and for providing me with the necessary knowledge and data to accomplish this work.

I thank my kind colleagues from Loria with whom I spent a lot of time during my PhD. I keep very pleasant memories about our lunches, coffee breaks, fascinating conversations, and fun 'Ciné-club' dinners. I will go in chronological order trying not to forget anyone. Phillippe Kalitine, Matthieu Nicolas, Quentin Laporte Chabasse, Victorien Elvinger, Kahina Bessai, Amina Ahmed Nacer, Riyadh Abdmeziem, Anis Ahmed Nacer, Chahrazed Labba, Siavash Atarodi, Hoai Le Nguyen, Vinh Dang Quang, Guillaume Rosinosky, Clément Parisot, Nouha Othman, Karima Abidi, Linda Ouchaou, Clélie Amiot, Alexandre Bourbeillon, Pierre-Antoine Rault, and Sami Djouadi. I probably forget some, but they will probably forgive me, knowing my goldfish memory that I acquired during this thesis.

I thank also the staff of the Inria Grand Est restaurant who welcomed us every day with a smile and a positive energy, without forgetting Isabelle good bakeries and Caro good coffee.

Particular heartfelt thanks go to Mario Lezoche and his "EcoleNonEcole" Martial Arts club. Although I have only spent a few months with you and the other members of the club, I have learned many things. I learned to make silence inside myself and my mind, especially at difficult moments of life, through meditation. Staying still for several minutes and focusing on my breathing, help me a lot to manage and reduce my stress and anxiety, and to unwind and detach from the outside world whenever I feel the need to.

I want to say the warmest thanks and express my deepest gratitude to my precious family, that I missed a lot here. Thank you for doing everything you could to make me happy. Nothing would have been possible without you, and words will never be enough to show my appreciation for you. I am entirely grateful to my mother Touria. Thank you for being my stronghold in my weakness times, for your immense and unconditional love, for your enormous support, for not letting me lose hope when I was ready to do it, and for believing in me when I was full of doubts. Thanks to my dear sister Saloua for being with me every second of my life even staying thousands of kilometres away. A warm thank you for my lovely and cute nephews Saif and Jad for every video call and every "Love you and miss you Tata Abir". Your love, words and encouragement have given me the strength to continue this journey.

I also thank myself for not giving up and for all the efforts, determination, and perseverance that I have shown during all these years.

I want also to give a sweet thank you to my cat Molly for being a stress relief and calming aid, whenever I cuddled her to procrastinate on writing my thesis.

Finally, I thank those people that I, unfortunately, forgot to mention.

Abir Ismaili-Alaoui
Nancy, 07/12/2022

Je dédie cette thèse
à ma mère.

Contents

List of Figures

List of Tables

Résumé

Les processus métiers peuvent être vus comme une collection d'activités exécutées et coordonnées afin de produire un résultat bien spécifique, répondant aux besoins d'un client (interne et/ou externe). La gestion des processus métiers (Business process management - BPM) est un domaine de recherche très actif dans la discipline des systèmes d'informations. Il a pour objectif l'amélioration continue de l'efficacité et la performance des processus d'une entreprise, par le biais des méthodes, des techniques et des outils qu'il offre afin d' appuyer la conception, la mise en œuvre, la gestion, l'analyse, et l'automatisation, dans la mesure du possible, des processus métier, et donc gagner en termes d'agilité, de flexibilité et de performance.

Même si plusieurs méthodes d'amélioration des processus métier (Business Process Improvement- BPI) sont disponibles dans la littérature, les organisations rencontrent toujours des difficultés pour les appliquer efficacement. Ces difficultés peuvent être justifiées par le fait que les méthodes BPI existantes ne répondent pas à toutes les exigences récentes des organisations et ne s'adaptent pas aux progrès réalisés, ces dernières années, dans plusieurs domaines tels que l'intelligence artificielle, les techniques d'analyse des données, l'apprentissage automatique, le process mining et le traitement des (flux) événements, etc.

En outre, avec cette nouvelle ère de digitalisation et l'essor de plusieurs nouvelles technologies telles que le Big Data, l'Internet des objets (IoT), le Cloud Computing, etc, les organisations sont confrontées à de nouveaux facteurs et défis redéfinissant le marché et qui génèrent de réels changements dans le BPM traditionnel. Parmi ces nouveaux défis on trouve la quantité de données et d'événements, provenant, avec une très grande vélocité, de différentes sources hétérogènes (des interactions internes ou externes de l'entreprise, IoT, etc). Ces données doivent être bien analysées et exploitées afin d'en extraire, des résultats à forte valeur ajoutée qui peuvent aider l'entreprise dans son processus de prise de décision. Cependant, les outils traditionnels proposés par la méthode du management des processus métiers présentent différentes limites concernant le traitement, la fouille et l'analyse des données et l'exploitation des résultats de ces analyses en temps réel.

La nature interdisciplinaire du BPM est un facteur clé qui favorise les perspectives d'amélioration dans ce domaine. L'objectif de ce travail de thèse est de proposer de nouvelles approches pour augmenter les processus métier, en s'appuyant principalement sur l'analyse des données, les algorithmes d'apprentissage automatique et le traitement des événements complexes, afin d'exploiter les données et évenements générés par l'exécution des processus métier et de trouver des moyens d'améliorer ces processus sous différents angles tels que l'ordonnancement des instances et la gestion des événements dans un environnement IoT.

L'IoT est en train de devenir une zone d'innovations technologiques et de promesses de développement économique pour de nombreuses industries et services. Ce nouveau changement de paradigme affecte toutes les couches de l'architecture d'entreprise, de l'infrastructure au métier. Le Business Process Management (BPM) est un domaine parmi d'autres qui est affecté par cette nouvelle technologie. Pour faire face à l'explosion des données et des événements résultant, entre autres, de l'IoT, les processus d'analyse de données combinés aux techniques de traitement des événements, examinent de grands ensembles de données pour découvrir des modèles cachés, des corrélations inconnues entre les événements collectés, soit à un niveau très technique (détection des incidents/anomalies, maintenance prédictive), soit au niveau métier (préférences des clients, tendances du marché, opportunités de revenus) pour fournir une meilleure efficacité opérationnelle, un meilleur service client et des avantages concurrentiels sur les organisations rivales. Afin de capitaliser la valeur commerciale des données et des événements générés par les capteurs de IoT, IoT, Data Analytics, Traitement des Évènements et BPM doivent se rencontrer au milieu.

Mots-clés: Gestion des Processus Métier, BPM augmentée, Internet des Objets, Traitement des événements Complexes, CEP, Sciences des données, Apprentissage automatique

Abstract

Business Processes (BP) can be seen as a collection of activities executed and coordinated in order to produce a specific result and to meet the needs of a customer (internal and/or external). Business process management (BPM) is a very active research area; its objective is to provide a comprehensive and insightful analysis of product flow and to identify inefficiencies and potential improvement areas in the process to achieve better decision making and results.

Although we can find several Business Process Improvement (BPI) methods in the literature, organizations are still facing some difficulties to apply these methods effectively. These difficulties can be justified by the fact that the existing BPI methods do not fit with all the recent requirements of the organizations and the progress that the world has achieved in the past few years, in several domains such as Artificial Intelligence, Data Analytics techniques, Machine Learning, Process Mining, Event (Stream) Processing, etc. Besides, with this new digitized era and the rise of several new technologies such as Big Data, Internet of things, Cloud computing, etc, organizations are faced with many factors and challenges that generate real changes in the traditional BPM. Among these challenges, we have a large amount of data and event data that are continuously gathered within the organization. These data represent a real engine of growth for organizations and must be adequately exploited to extract high added value that can assist the organization in its decision making process. Furthermore, enterprises are looking for advanced technologies that optimize time and resources and increase agility, productivity, and most importantly, proactivity. However, traditional BPM systems have different limits, as they do not facilitate the use of knowledge extracted from these data, by business processes, because they do not benefit from statistical functionalities and data analysis and manipulation techniques in real time.

The interdisciplinary nature of BPM is a key factor that fosters opportunities for improvement in this domain. The objective of this thesis work is to propose new approaches to augment business processes, by relying mainly on data analysis, machine learning algorithms, and complex event processing, to exploit the data generated by the execution of the business process (event data, event logs) and find ways to improve these processes from different perspectives, such as instance scheduling and event management in an IoT environment.

IoT is becoming a hot-spot area of technological innovations and economic development promises for many industries and services. This new paradigm shift affects all layers of enterprise architecture from infrastructure to business. Business Process Management (BPM) is a field among others that is affected by this new technology. To assist data and event explosion resulting, among others, from IoT, data analytic processes combined with event processing techniques, examine large data sets to uncover hidden patterns, unknown correlations between collected events, either at a very technical level (incident/anomaly detection, predictive maintenance) or at business level (customer preferences, market trends, revenue opportunities) to provide improved operational efficiency, better customer service, and competitive advantages over rival organizations. In order to capitalize the business value of data and events generated by IoT sensors, IoT, Data Analytics, Event Processing, and BPM need to meet in the middle.

Keywords: Business Process Management, BPM, Augmented BPM, Internet of Things, IoT, Complex Event Processing, CEP, Data Science, Machine learning

Chapter 1

Introduction and Problematic

> Intelligence is the ability to adapt to change.
>
> Stephen Hawking

Faced with a competitive and a continuous changing environment, organizations, now more than ever, tend to have recourse to agile methods and continuous improvement practices such as lean six sigma, Activity Based Costing (ABC) methods, and Business Process Management (BPM) in order to have a transversal view over their business, to identify opportunities for enhancement, to reduce costs and wastes, to improve customer satisfaction and overall profitability, and to overcome the problem of isolation and non-communication between the different hierarchical levels within the organization. Among all these existing approaches, Business Process Management (BPM) is still considered the most appropriate solution that helps organizations adapt to strategic, tactical, and operational changes, and also have more visibility and control over their business processes to be able to continuously improve and optimize them. This approach allows companies to gain in terms of agility, efficiency, and performance; it also enables effective communication and collaboration between their different stakeholders. Recently, with this new digitized era and the rise of several new technologies such as big data, the Internet of things, cloud computing, etc., organizations are faced with many factors and challenges that generate real changes in traditional BPM. Among these challenges, we have a large amount of data and event data that are continuously collected within the organization. These data represent a real engine of growth for organizations and must be adequately exploited to extract high added value that can assist the organization in its decision-making process. However, traditional BPM systems have different limits, as they do not facilitate the use of knowledge extracted from these data by business processes, because they do not benefit from statistical functionalities and data analysis and manipulation techniques. Over the past decade and until now, a lot of efforts have been done in both academic and industrial research environments, to improve all the aspects of Business Process and Business Process Management.

1.1 Background and Motivation

Business Process Management (BPM) is a well-established discipline in both academia and industry. It is considered as a powerful solution that helps organizations adapt to strategic, tactical, and operational changes and gain more visibility and control over their business processes, so that they can continuously improve and optimize their activities and resources. Organizations use Business Process Management systems as an activity-based workflow manager that allow them to track the optimized functioning of their activities in order to gain in terms of agility, efficiency, and performance. The evolution of business process management has gone through four waves *"since its post-industrial revolution inception"* [12]. The first wave of business process improvement lasted from the 1970s to the 1980s. Quality management and task efficiency were the core interests of this wave. Total Quality Management was among the most commonly used process improvement methods during this wave [13]. Since the 1990s, the world has witnessed the second wave of process improvement. This wave was known as the process re-engineering wave, and it focused basically on process innovation, and process "best practices" to get better, faster and cheaper process execution [12]. Among the technologies that are the most used during this wave, we have Enterprise Architecture, ERP (Enterprise Resource Planning), CRM (Customer Relationship Management), and supply chain. The most important process improvement approaches discovered during this second wave are Six Sigma [14], Process Re-design/ Re-engineering Methods [15] [16], and Enterprise Resource Planning [17]. The third wave, known as the Business Process Management wave, started around 2000 [12] [18], and focuses on agility, flexibility, continuous transformation, and distribution in business processes [12]. Among the technologies commonly used during this wave, we have Service-Oriented Architecture (SOA) [19], Balanced Scorecard [20], Outsourcing, Co-Sourcing, In-sourcing [21], and BPM Methods and BPM Systems. Recently, a fourth wave of business process management (Augmented Business Process Management) is triggered by emerging technologies such as IoT, Big data, Artificial Intelligence, Data analytics advancement, Blockchain, etc. Among the prominent results of this new wave, we have process mining [9], Robotic Process Automation (RBA) [22], Augmented Process Automation [23], Prescriptive process monitoring [24] just to name a few. Although current business process approaches, which are based on all these new technologies, have reached a certain level, there is still a significant gap between what is proposed in these approaches and what actually happens within companies.

In most cases, business processes are isolated from each other [25] or from the external ecosystem of the organization. Thus, they do not benefit from the different added values that could be created from sensor data, for example, and the useful knowledge that could be extracted from event logs and historical data from previous executions. Furthermore, business processes operate in a reactive way [26] which is not sufficient when faced with new radical or incremental changes. Early anticipation is crucial to avoid the occurrence of the problem or respond to it quickly and efficiently. This lack of proactivity and predictability is remarkable in the three main steps of the BPM life cycle [27] [28]: design and redesign step, implementation step, and execution step. To address this problem, proactive-oriented concepts are starting to be used in the BPM glossary such as proactive business process management [26] [29], process forecasting or future-oriented BPM [28] and context-aware business processes [30]. Therefore, switching from reactive to proactive and adaptable business processes becomes mandatory for every organization. With this new digitization of industrial processes, comes also the age of assistance, which means that companies should be focused on customers, in order to offer personalized and adaptable services, and even predict their needs in almost real time. Dealing with such a continuous changing environment requires intelligent, adaptive, and flexible business processes. That is why organizations today find that traditional BPM and also BPI approaches have several limitations [31]. In the literature, different approaches have been proposed to improve business processes and keep them up-to-date with new technologies. Among the proposed techniques and technologies, we have recommended systems [32] [33], Ontologies [34] [35] [36], data analysis, data mining, and process mining [37] [38] [39], complex event processing [26] [40], Ubiquitous Computing [41] [42], Internet of Things - IoT [43] [44], just to name a few.

1.2 Research Objective

When dealing with incident management case studies, processes are executed within application systems belonging to the real world, where humans, cooperative computer systems, and even physical objects are involved. In fact, connected objects are becoming progressively more prominent in the business process execution environment, where we can have a set of sensors connected to a BPM platform for incident management and surveillance to have an instant reaction.

The Internet of Things (IoT) represents the internetworking of physical objects [45] [46] (also referred to as "things", "connected devices", "smart devices", "ubiquitous devices"), vehicles, and other items embedded with sensors, electronics, actuators, and network connectivity that enable these "things" to collect and exchange data when interacting and sensing their environment. Smart objects have been swept into our life to facilitate it in so many ways and in different domains such as transportation, health care, hospitalization, civil protection, smart home, smart cities, emergency, and individual automation. From smart phones to new smart objects that interact not only with people, but also with other machines (machine-to-machine communication). The concept of "Internet of Things" first emerged in a presentation by Keven Ashton, an expert on digital innovation, on the connection of Radio Frequency Identification (RFID) to the supply chain of Procter and Gamble in 1999. Since then, the IoT has been exploding and invading our daily lives in different aspects (smart phones, smart door locks, self-driving smart cars, smart cameras, smart medical devices, etc.). This new technology started to thrive right after the development of the Web in the 1990s and the mobile Internet in the 2010s. The Internet of Things (IoT) is growing in leaps and bounds. IoT infrastructures can range from connected, instrumented devices providing data to intelligent, standalone systems. The IoT enables "a world where things can automatically communicate with computers and each other, providing services for the benefit of humanity" [47].

Connected devices collect data by sensing their environment and exchange data with other devices and humans. All devices (things) act locally within their environment. Whereas, the IoT allows them to be remotely monitored via existing network infrastructures, including the Internet [1]. The IoT contributes to continuously feeding "big" data to every node [48]. Furthermore, the evolution of the IoT invokes significant opportunities for private data exchange, enabling new business models across heterogeneous networks [49]. However, it cannot generate value. That is why, it is necessary to couple it with other technologies to transform this huge amount of data into useful knowledge, in order to make meaningful decisions.

From a business process perspective, the widespread use of these IoT devices led to new business process interactions within organizations such as Person-to-Thing, Thing-to-person, and Thing-to-Thing due to the advent of Internet of Things technologies [50]. These new process interactions complete existing ones, such as Person-to-Person (P2P), Person-to-Application (P2A), and Application-to-Application (A2A) processes [27] [51] [52]. Thus, as we can see, the omnipresence of these "things" is both a challenge and an opportunity for process improvement.

In this thesis, we focus on incident management business processes, where process instances are launched by events generated from IoT devices. These processes could be seen as an Event-Driven Business Process. In fact, Event-Driven Business Process Management is an approach that combines two different fields, namely Business Process Management and Complex Event Processing (CEP), which aims to handle the requirements of high-volume event integration encountered in the BPM field. The combination of these two fields is not recent; however, it has become increasingly sought after in information systems that need to handle event-driven processes and real-time data sources such as IoT devices [53]. Internet-of-Things objects have become a part of several complex business processes. However, handling IoT technology in process-based information systems is still constrained by several challenges that hinder communication between these two worlds. The integration of IoT-BPM is still in its infancy. Most of the current research work on BPM-IoT integration proposes new approaches that target a specific aspect of the BPM life cycle. For example, updating business process models by enriching business process model and notation (BPMN) with new elements that correspond to the IoT domain and that can explicitly define IoT devices within a business process [54] [55], improving resource optimization and monitoring and task execution through an IoT context-specific knowledge provisioning [1], improving business process execution via an IoT-aware business process execution that exploits IoT for BPM by providing IoT data in a process-aware way [56], or proposing an architecture for IoT-BPM integration to cope with the issues and limitations raised by recent case studies in both industry and academia [1], [57].

This integration brings several capabilities into current business processes, such as:

— Reducing delays in responding to emergencies.
— Facilitate cases or instances escalating.
— A more responsive access to information with a customer-based focus.

We will discuss these impacts in more detail in the next subsection.

1.2.1 How IoT can impact and improve business processes?

The alignment of IoT and BPM is the focus of several research works. However, this alignment comes with various challenges that need to be tackled. Both technologies will affect each other and, of course, benefit from each other [1] [58] [59]. However, we focus only on the influence / benefit of IoT on BPM.

The integration of IoT can provide several benefits for BPM. Besides, it can also address some BP-related limitations. Among these limitations, we have the following:

— **Physical surrounding:** Business processes have no access (or limited access) to a physical surrounding, as they operate in a cyber surrounding. Taking into account the physical surrounding in the modeling phase can lead to correct resolution and execution of business processes [55] [59] [60].

— **Context-insensitivity:** Business processes are insensitive to context; they are considered blind and stateless, which means that in each business process execution we do not take into account neither the results from the last process instances nor the context (a context-aware business process execution). Therefore, they need to know the conditions and situations in which IoT operates, given that IoT is by default context sensitive through devices/sensors [59] [56] [61] [62] [58].

— **Data-input:** Business processes have no direct access to data generated by different devices and sensors. These data could be exploited in business process execution to progress via taking actions (e.g., IoT-based trigger events/alerts, IoT-based decisions, etc.). [58] [45] [56] [59].

— **Models complexity:** Integrating IoT technology can reduce the complexity of process models (for example, replace elements or patterns, etc.). It can also extend and enrich process models. As a result, we have more precise process definitions that accurately reflect the operational reality [58].

— **Transparency:** Integrating IoT technology improves the monitoring of business processes by increasing the transparency of business processes through the data provided by IoT sensors [58].

— **Latency:** The incorporation of IoT technology can ultimately lead to activity run time reduction and significant latency that can result in an overall performance enhancement [58].

— **Event-logs quality:** IoT sensors produce a large amount of data that enrich process event logs. Given that event logs are the fuel of process mining technology, enriched event logs provide enriched process models [58] [63].

Other limitations encountered concern the fact that business processes are supposed to be smoothly executed under different business situations and context. This constantly changing environment requires having business processes that can easily be adapted to the appropriate action taken. However, without being coupled with other technologies (IoT in this research work), business processes are still deficient regarding the critical ability to provide assistance to their users [67] due to a lack of two important aspects:

— **Context-awareness** : The emergence or even omnipresence of IoT solutions in different businesses forces organizations to adapt their processes to a high level of connectivity. Context-awareness is a fundamental characteristic of ubiquitous computing [64], and it is the key to benefit from the raw data collected by the sensors, as it allows to store contextual information related to these raw data and to decide which data should be processed, in order to facilitate interpretation [65] especially at the level of business processes.

— **Knowledge feedback loop**: traditional BPM systems have different limits, as they do not facilitate the use of knowledge extracted/generated from data by busi-

BPM Limitations	BPM Life-cycle phases				References
	Process modeling	Process analysis	Process execution	Process monitoring	
Physical surrounding	X		X		[55] [59] [60]
Context-insensitivity			X		[59] [56] [61] [62] [58]
Data-input			X		[58] [45] [56] [59]
Models complexity	X		X		[58]
Transparency				X	[58]
Latency			X		[58]
Event-logs quality		X			[58] [63]
Context-awareness	X	X	X	X	[64] [65]
Knowledge feedback loop	X	X	X	X	[57] [66]

Table 1.1 – Business process related limitations, addressed via IoT integration

ness processes after their execution. As a result, a tremendous amount of data and event data that are constantly collected within the organization are not exploited to improve business processes. As a matter of fact, these data represent for enterprises a real engine of growth. However, a large amount of raw data is not valuable; data must go through a whole process to extract value from it. The analysis of large data helps organizations extract information and then knowledge, because the real value is in how organizations will use those data and turn their organization into an information-centric company that relies on the insights derived from data analyses for their decision-making [68]. The integration of IoT technology can foster a feedback loop that feeds forward insight to adjust either Fog/Edge or device algorithms. This can also lead to better adjustment and improvement in the process life cycle [57] [66].

Table (1.1) summarizes and classifies these limitations related to the business process according to the different phases of the BPM cycle lifecycle.

1.2.2 Problems encountered in IoT/BPM integration

The integration of IoT with BPM certainly contributes to the improvement of business processes. However, this integration comes with several issues and challenges. In [69], authors present the challenges that need to be addressed in Business Process Management Systems (BPMS) to achieve an efficient integration of IoT, such as the absence of direct interaction between the business layer and the edge network, or the problem of complex and inflexible business process models due to a lack of standardization and interoperability when modeling IoT elements and components in BPM. The unexploited potential of the extracted data from sensing environment represents a challenging issue when integrating IoT with BPM [70]. Security and data privacy represent another level of IoT/BPM integration concerns [1]. Many other challenges exist and must be tackled when dealing with IoT and BPM integration and alignment; these challenges have been summarized by C.Janiesch et al. in their manifesto [1] (see Figure 1.1). In the following, we briefly outline these 16 challenges.

- Placing IoT sensors in a process-aware way and linking these sensors to running processes.
- Visualization support and monitoring for manually executed physical processes.
- Connection of analytical processes with IoT.
- Integrating the IoT with process correctness checks.
- Dealing with unstructured environments.
- Managing the links between micro processes.
- Breaking down end-to-end processes.
- Detecting new processes from data.
- Specifying the level of autonomy of IoT things.
- Specifying the "social" roles of agents.
- Concretizing abstract process models.
- Dealing with new situations.
- Bridging the gap between event-based and process-based systems.
- Improving Online Conformance Checking.
- Improving and optimizing resource utilization.
- Improving resource monitoring and quality of task execution.

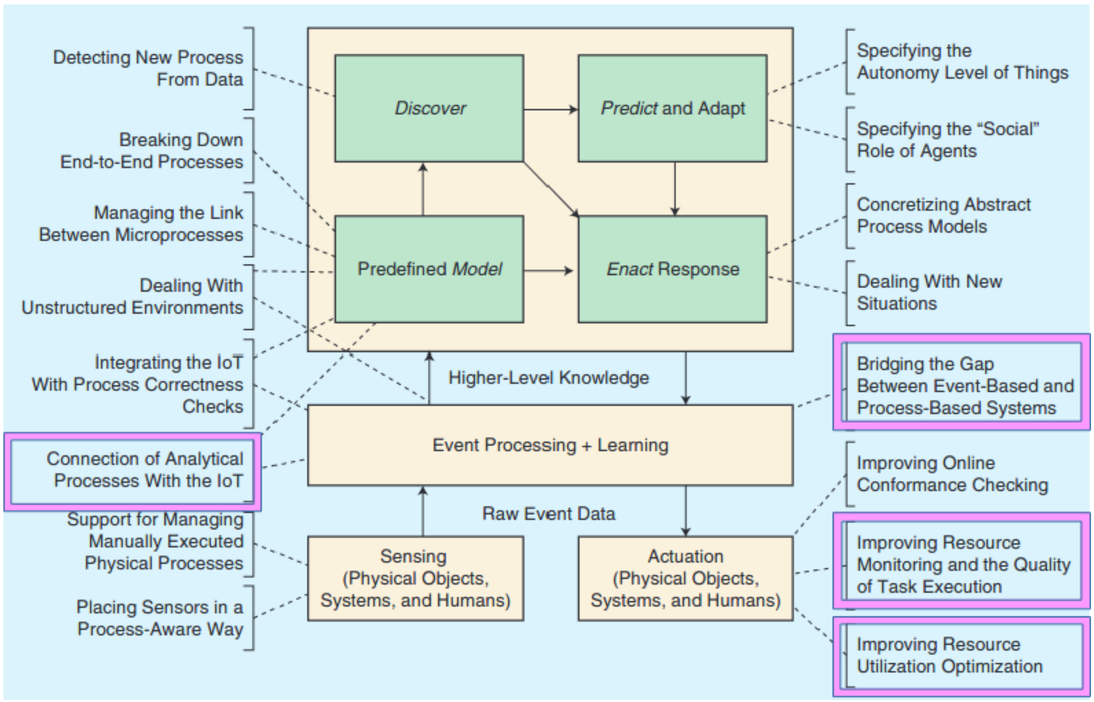

Figure 1.1 – A high-level overview of the interaction between IoT and BPM - Challenges [1]

In this thesis, we focus on the challenges related to event processing and also the learning aspect within an IoT-BPM architecture (see Figure 1.1).

Learning from data and event data that are gathered from IoT sensors and past process execution is an effective approach to improve the performance of business processes, especially those that perform repetitive tasks and activities. In fact, the insights obtained from these data represent valuable support for business process improvement and decision making. These raw data must be processed using data science techniques and event-based systems in order to transform these raw data and raw event data into high-level knowledge. This high-level knowledge can be exploited thereafter, for business process improvement purposes.

To go into more detail and find solutions to the challenges we have discussed in this section, it is appropriate to ask some questions in order to define our problem in a more concrete way.

— **Q01** : In any research work, the first step is to have a clear idea of the state of the art of the problem to be addressed. That is why our first question is as follows: What is the state-of-the-art on this problem? and how can we benefit from them?

Since our research work spans several independent domains, as described in the previous sections, the integration and/or communication between these domains prompts us to ask the following questions.

— **Q02** : What are the latest advancements in the field of BPI (Business Process Improvement)?
— **Q03** : To what extent can proactivity and prediction be integrated to BPM?
— **Q04** : Are there effective predictive and/or proactive approaches or methods for managing business process instances?
— **Q05** : How to ensure proactive execution of business process instances using machine learning algorithms?
— **Q06** : What is the state-of-the-art regarding the integration/communication between IoT and BPM?

1.3 Thesis Contributions

This section gives a general overview of our contributions. The main objective of this thesis is to propose a data and event-based end-to-end methodology to augment business processes in the IoT environment. This contribution will attempt to provide solutions to the following research questions.

— **RQ01** : How can we exploit the events, data, and event data collected from the execution of a business process to improve the execution phase of a business process?
— **RQ02** : How can solutions based on data analysis and machine learning techniques help achieve this IoT/BPM integration?
— **RQ03** : What methodology can we follow in order to conceive a IoT-BPM architecture towards an augmented BPM?
— **RQ04** : Can CEP technology facilitate communication between IoT and BPM?

In this work, we propose a methodology that aims to improve the execution aspect of business process management. Especially, supervision and incident management processes. The instances of these processes are, in some cases, launched by IoT generated events (Big Data Push paradigm: i.e., data are triggering processes).

Modern problems require new solutions. In the same perspective, new problems arise with new technologies and require new solutions. The ubiquity of all these new technologies, such as big data, fast data, cloud computing, the Internet of Things (IoT), etc., implies new business process problems and challenges that require new solutions to ensure continuous improvement of business process management.

The incorporation of two heavy paradigms such as IoT and BPM generates, without a doubt, a modeling methodology issue. There are three types of modeling methodologies: - The Top-down approach, - The Bottom-up approach, - and the Meet-in the-middle approach. We propose a meet-in-the-middle approach to facilitate the integration of IoT and BPM. In fact, the meet-in-the-middle approach is considered as a method of refinement going alternately from top to bottom to bottom to top. The combination of deductive and inductive iterative sprints in this approach allows both reuse/mutualization and disruptive thinking. In this thesis, we propose the integration of IoT and BPM via a methodology that aims to provide a meet-in-the-middle environment capable of capturing data and event data from IoT sensors, when they are sensing their environment, create actionable and useful knowledge, and allow this knowledge to be used in the business layer through business processes.

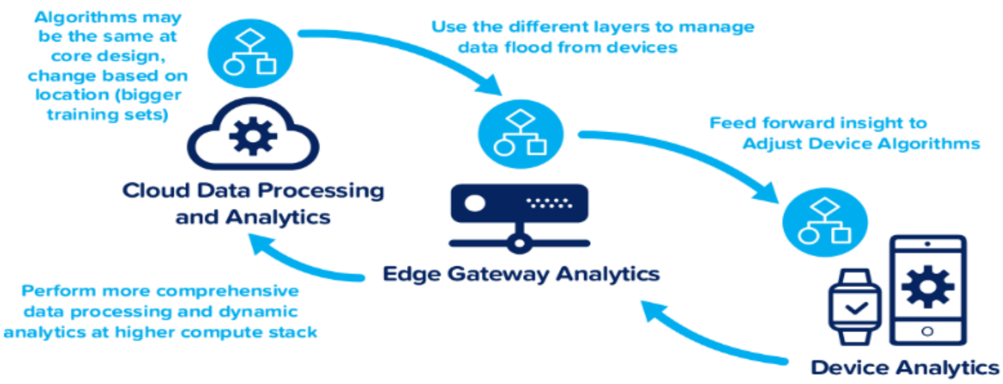

Figure 1.2 – Cascading Analytics Haze Architectural Pattern [2]

Our IoT-BPM integration methodology follows Haze Architecture and Cascading Analytics incarnated by a DIKW (Data Information Knowledge Wisdom) discovery pattern crossing the architecture from the device then Fog/Edge to the cloud, and a learning feedback loop that feeds forward insight to adjust either Fog/Edge or device algorithms as described in Figure 1.2 [2]. This methodology consists of the following steps:

- Priority-based business process scheduling problem modeling as linear programming system
- Process instance priority determination through Machine Learning
- Human resource allocation resolution through Genetic Algorithm
- Formulation of Business process uncertainty problem through Fuzzy Logic
- Business process instance (event priority level) inference through Fuzzy Logic

1.4 Structure of the dissertation

To facilitate the reading of this dissertation and allow the reader to easily access the different parts of this thesis, we propose a reading guide (Figure 1.3) that represents an overview of the structure of this thesis.

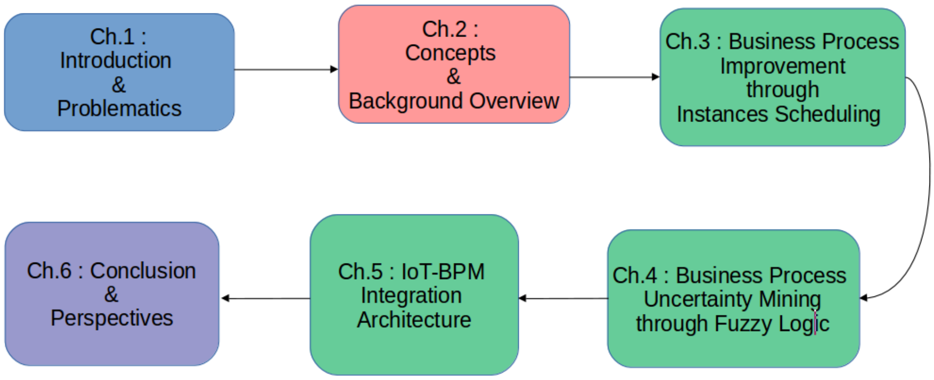

Figure 1.3 – Outline of the thesis

The thesis manuscript is made up of six chapters, including the introduction and conclusion.

Chapter 1 : This chapter presents the context of our work, the motivation behind our research, and provides a clear understanding of our thesis contribution.

Chapter 2 : The first section of this chapter provides basic concepts related to the core domain of this thesis, which is business process management. The second section is devoted to present a comprehensive overview of existing business process improvement approaches and methodologies. As my thesis contribution spans over a wide range of different domains, it would not be possible to give a single state-of-the-art. Therefore, we opt for a customized state-of-the-art in each contribution chapter.

Chapter 3 : This chapter represents the first part of our contribution. It provides all the details of the priority determination approach of the process instance through machine learning. And also, Human resource allocation resolution through genetic algorithm.

Chapter 4 : This chapter represents the second part of our contribution, which concerns the resolution of the uncertainty of process instances priority through fuzzy logic.

Chapter 5 : This chapter represents our end-to-end IoT BPM architecture for incident management, via an integration approach of four technologies : IoT - CEP - ML - BPM.

Chapter 6 : The last chapter gives a critical look at our various contributions and discusses the prospects and future work to be carried out.

Chapter 2

Concepts and Background overview

> Thoughts without content are empty,
> intuitions without concepts are blind.
>
> Emmanuel Kant

The purpose of this thesis is to propose methods to improve business process management. In this chapter, we present a state-of-the-art that focuses on work done in a context similar to ours and whose goal is to improve business process management. In the first section of this chapter, we will present some basic concepts related to the field of business process management. In the second section, we will present a comprehensive literature review of the most prominent business process improvement methodologies and approaches according to the BPM lifecycle.

2.1 Concepts

The scope of this thesis mainly focuses on business process management and business process improvement fields. This chapter therefore provides an elaborated list of the basic concepts and definitions that will be referred to in the following chapters.

2.1.1 Business Process

The Process approach has been increasingly adopted by companies since the 1980s, leading to a new organizational model and a new way of operating for organizations. Faced with a changing and competitive environment, traditional approaches that treat the company as a closed environment are no longer appropriate. Indeed, the process approach is a systemic approach that aims to transform the hierarchical and vertical structure of any organization into a transverse structure in order to satisfy the needs of external and internal customers. It is a method of analysis and modeling intended to ensure collaborative work in order to control and improve the efficiency and smooth running of the company.

This method is mainly based on the concept of business processes. A business process is the structure of activities and actions as they occur in the real world. It defines all possible paths in the real process and the rules that determine the path to follow and the actions to take [71]. The International Standards Organization (ISO) defines processes as a set of interrelated or interacting activities that transform inputs into output elements. For Harrington in [72], he presents business processes as an activity or set of activities that uses an input, adds some value to it, and delivers it as an output to an internal or external customer. Dumas et al. [3] represent a business process as a collection of interrelated events, activities, and decision points that involve a number of actors and objects, which collectively lead to an outcome that is of value to a customer. All these definitions and many others present business processes as a set of activities and tasks that exploit the different resources (human and/or machine) of the organization to achieve one or more objectives previously defined, in order to satisfy an internal or external customer. Each business process is attached to a single enterprise, but in some cases it may interact with other business processes belonging either to the same organization or other organizations [73]. To achieve its objectives and ensure efficient performance, an organization must subject its business processes to a continuous improvement mechanism that aims to optimize and automate these processes as much as possible.

Therefore, business processes occupy a very important place in the field of information systems, because they play a core role within every enterprise or organization. Moreover, the performance level of any organization is indirectly linked to the efficiency of its processes and the quality of their models. In fact, proper management of business processes within an organization can have a very positive impact on the efficiency and smooth running of its activities, as it allows this organization to have a clear vision of its objectives in order to better meet the requirements of competitiveness that are constantly increasing.

2.1.2 Business Process Management - BPM

Given the continuous evolution that organizations are experiencing today, they feel the importance of having a good information system, because it plays a crucial role in maintaining the performance level and promoting more communication and collaboration between different services and actors within the company. One of the strategies increasingly adapted by companies and their information systems is Business Process Management (BPM). Business process management is seen as the perfect solution that helps organizations adapt to the strategic, organizational, and technical evolution. When an organization adapts a business process management approach, it is in order to have more visibility and control over its activities and interactions between these processes, to be able to model, manage, improve and optimize these processes continuously, and therefore gain in terms of agility, flexibility, and performance. Business process management is the art and science of overseeing how work is done in an organization to ensure consistent results, leverage opportunities for improvement, and increase efficiency and performance within the organization [3]. Moreover, BPM allows one to foster good communication and collaboration between the different actors and services through the methods, techniques, and tools it proposes to ensure the design, implementation, monitoring, analysis, and automation, if necessary, of business processes. BPM suggests a life cycle to continuously improve processes, this life cycle includes the following steps (see Figure 2.1) [3]:

— **Process identification** : This phase starts with a positioning of a specific business

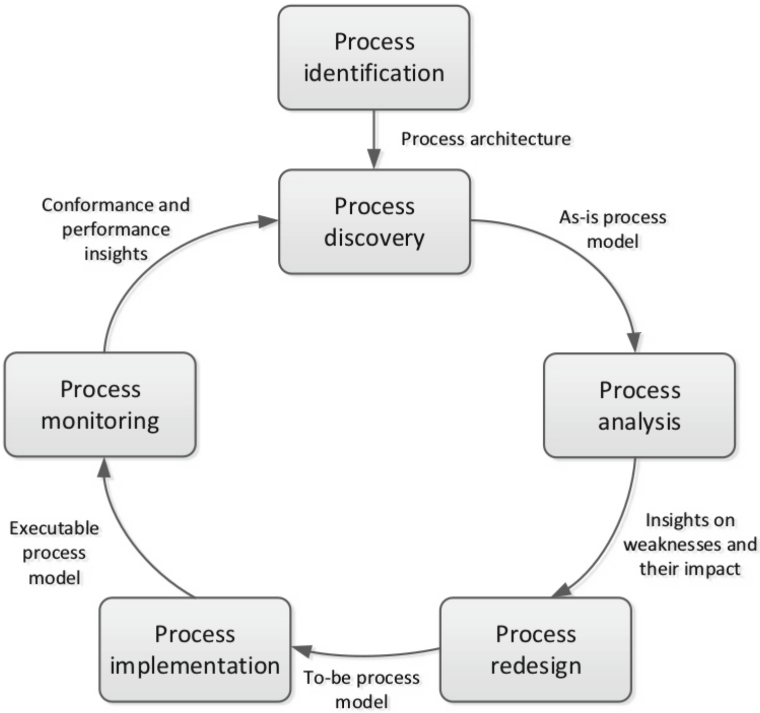

Figure 2.1 – The BPM lifecycle [3]

problem and ensures a better consistency between the real needs of the stakeholders and the business needs captured by the identified business process(es). The result of this identification phase is a new or updated process architecture, which gives an overall picture of processes in the enterprise.

— **Process discovery** : also called as-is process modeling. In this step, the current state of each identified relevant process is documented, usually in the form of one or more as-is process models.

— **Process analysis** : During this phase, problems associated with the as-is process are identified, documented, and, if possible, quantified using performance measures or metrics. The result of this phase is a structured set of problems. These are prioritized according to their potential impact and the estimated effort required to solve them.

— **Process redesign** : also called process improvement. The objective of this phase is to identify process adjustments that would address the issues identified in the previous phase and enable the organization to achieve its performance objectives. Process redesign/improvement and process analysis go hand in hand, when new change options are proposed, they are analyzed using process analysis techniques. The most promising changes are ultimately selected and combined into a redesigned process. The result of this phase is usually a new process model (to-be process).

— **Process implementation** : During this phase, the necessary changes to shift from the existing process (as-is process) to the new process (to-be process) are prepared and implemented. The implementation of the process covers two aspects:

1. **Organizational change management** : which refers to the set of activities necessary to change the way all participants involved in the process operate.
2. **Automation** : which refers to the development and deployment of information systems (or enhanced versions of existing information systems) that support the new process.

— **Process monitoring** : Once the redesigned process is running after improvement, relevant data are collected and analyzed to determine how well that process performs compared to its performance measures and goals. Bottlenecks, recurring errors, or deviations from expected behavior are identified, which will allow corrective measures to be taken later.

This cycle is continuously repeated, each time new problems arise. At the end of each cycle, we have a new version of the process in question, increasingly adapted to the needs and requirements of the enterprise. This further promotes the vision of continuous improvement adapted by the company to reach a good level of maturity and performance.

BPM standards

Business process modeling is one of the means increasingly adapted by organizations that adopt the BPM approach, because it promotes good communication and collaboration between the various departments and actors within the organization. Indeed, business process modeling is an important step in the process-centric management approach. It provides graphical representations of the important workflows as they occur in the enterprise and captures the different tasks, activities, and interactions between stakeholders. This step helps organizations to have a clear overview of their different processes, so that they can detect bottlenecks, backlogs, or any other issues or dysfunctions. Process modeling takes into consideration different perspectives such as [74] :

— **Control-flow perspective**: This perspective allows us to describe the logical order that exists between the different activities and tasks of a business process.

— **Data-flow perspective**: This perspective allows one to describe information transmission, i.e. the data flow exchanged between the different activities and tasks within the process.

— **Resource perspective**: This perspective concerns the allocation of machine and/or human resources to the different tasks and activities of the process.

— **Time perspective**: This perspective concerns the representation of time aspect in a business process: duration, start time, end time, deadlines, different events, etc.

Nevertheless, business process design and modeling is not an ultimate goal for the organization, but rather a crucial step in order to be able to analyze and control the various processes, and thus have an overall mastery and control over the company's performance and efficiency. Process modeling languages are mainly based on two formalisms, namely, the graph-based formalism and the rule-based formalism. A graph-based modeling language is based on graph theory or its variants, while a rule-based modeling language is based on formal logic [75]. Several modeling languages and standards exist, which allows to frame the necessary semantics and syntax, in order to design and model the business process features. These standards are classified as follows [76]:

— **Graphical standards**: This allows users to represent business processes, their flows, and transitions in a diagrammatic format. For example, BPMN (Business Process Model and Notation) and YAWL (Yet Another Workflow Language).

— **Execution standards**: These standards allow the computerization of the deployment and automation of business processes. Among these standards, we find XPDL (XML Process Definition Language) and YAWL (Yet Another Workflow Language).

— **Interchange standards** : These standards facilitate the portability of data. Among these standards, we have XPDL and XML (Extensible Markup Language).

— **Diagnosis standards** : They provide administrative and process monitoring capabilities (such as run-time and post-modeling). They also allow to identify bottlenecks, to audit and to query in real time business processes. For example, BPRI (Business Process Runtime Interface) and BPQL (Business Process Query Language).

As we have seen in this section, there are several graphical modeling languages that allow to model business processes, such as BPMN (Business Process Model and Notation) [77], UML (Unified Modeling Language - Activity diagrams) [78], YAWL [79], EPC (Event-driven Process Chain) [80] [81], IDEF3 (Integrated DEFinition for Process Description Capture Method) [82], and Workflow Management Coalition (WfMC) standards [83]. Each of these modeling languages provides its own appropriate syntax and semantics for modeling business processes. Nevertheless, they all share the basic concepts in process modeling, i.e. activities, tasks, events, control flows, artifacts, resources, gateways, etc. Yet BPMN, currently in version 2.0, remains the most widely used modeling language in the BPM domain, and considered as the standard. BPMN was first introduced in 2004 by the Business Process Management Initiative (BPMI), which was a consortium of companies specialized in the process management product market, and was merged in 2005 with the Object Management Group (OMG). BPMN defines a business process diagram (BPD) which is based on a flowcharting technique suitable for creating graphical models of business process operations. Therefore, a business process model is a network of graphical objects, which are activities and control flows that determine their order of execution [84].

BPMN basic elements

BPMN provides a rich set of elements to model a business process taking into account different levels of detail. These elements can be categorized into a set of core graphical elements and an extended, more specialized set. The core set contains the essential basic elements to model simple business processes. Whereas the extended set is used for more complex process scenarios as it contains more specialized elements that add a level of detail that facilitates and simplifies process simulation, evaluation or even execution [85]. Figure 2.2 presents an overview of a set of BPMN elements that are related to control-flow specification . Among these elements we have *objects* that can be an event, activity, or gateway, *sequence flows*, these represent a link between two objects in a process model and this link denotes a control-flow relation which mean an ordering relation, and *message flows* which allow to capture interactions between processes [4].

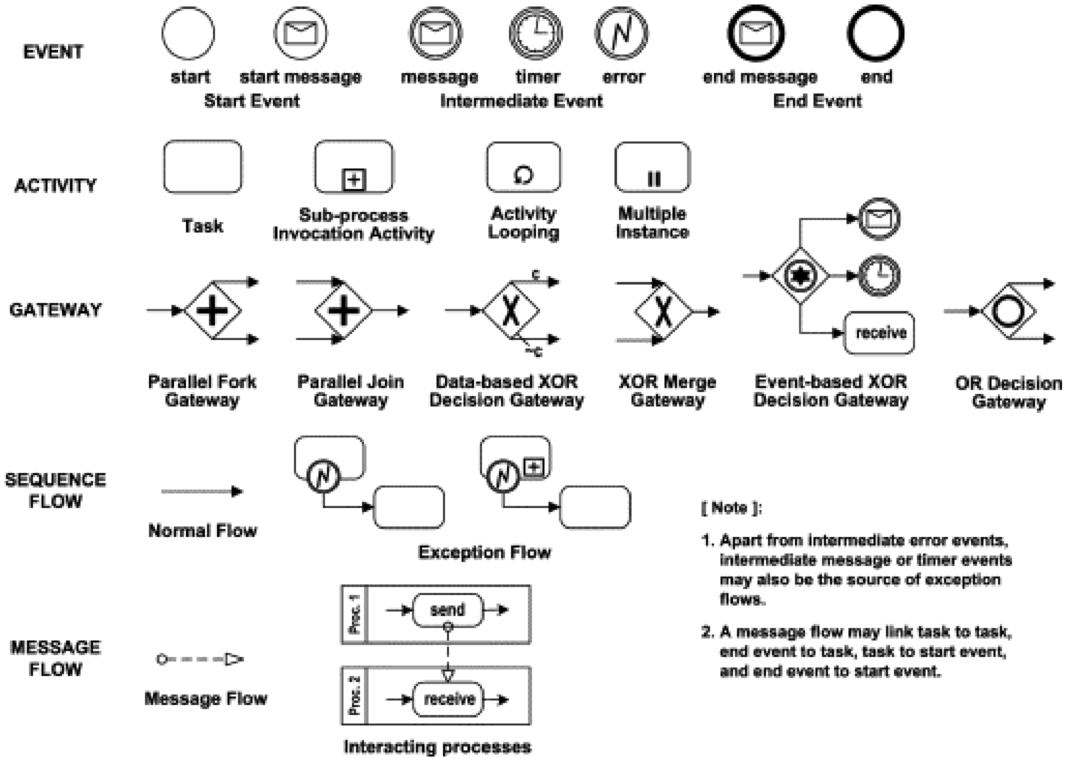

Figure 2.2 – Overview of BPMN elements - [4]

In the following, we add the definitions of the main BPMN elements:

- **Event**: an event can indicate the beginning of a process (*start event*) or the end of a process (*end event*), or can take place asynchronously during the execution of the process (*intermediate event*). Different situations may launch these events such as sending or receiving a message (*message event*), the end of an instance given time (*timer event*), or an error or exception that occurs during the process (*error event*), etc.
- **Activity**: It can be either a *task*, which represents an atomic activity, or a *sub-process* which is composed of a flow of other activities. A task represents a work to be done by either a human resource or a machine.
- **Gateways** : a gateway represents a routing element in BPMN. Shows the routing of a process flow. There are several types of gateway such as *parallel gateways* used to create concurrent flows in a process, for example AND-split, AND-join, *exclusive or inclusive gateways* used for exclusive or inclusive choices based on conditions, for example, XOR-split, XOR-join, and OR-split, etc.

BPMN contains other sets of elements that are not related to control-flow specification. These include artifacts, lanes, groups, pool and associations as shown in Figure 2.3. These elements are used to model *resource* and *data* perspectives in the process [5]:

- **Swimlanes**: Used to organize activities that are executed by a specific role within a process model. Swimlanes include *Pools* and *lanes*.

— **Artificats** : These elements are used to add more information into the process to have a more readable process model. These elements include *data objects*, *text annotations* and *group*.
— **Association** : this element is considered as a *connecting object* that allow to connect a data object element to an activity for example.

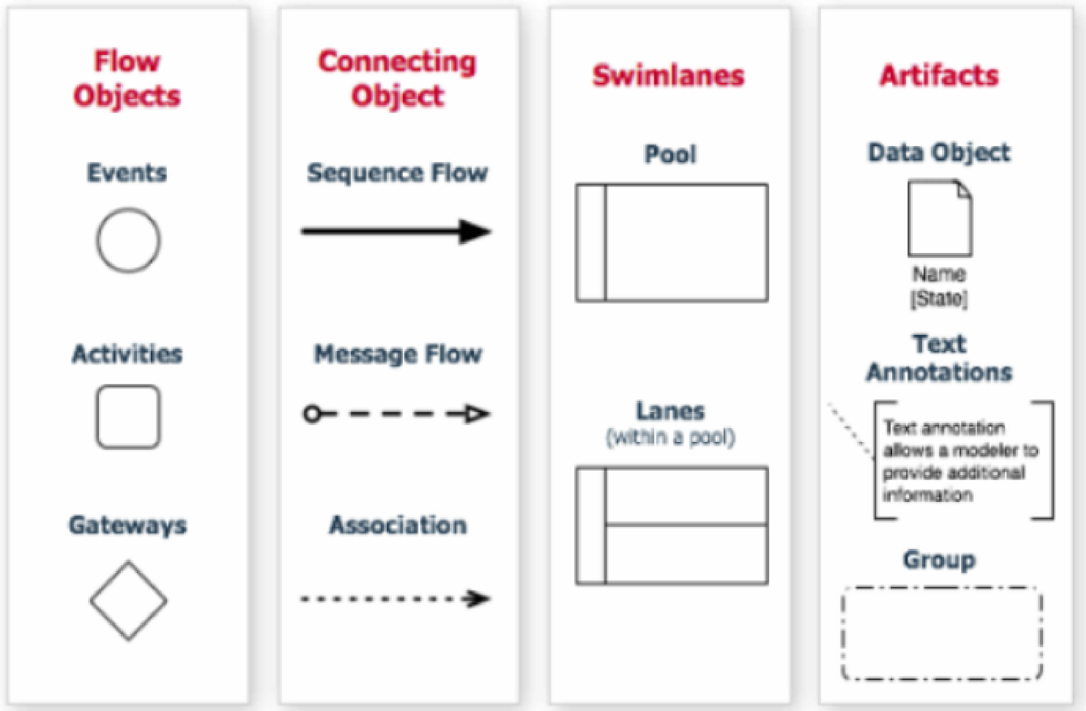

Figure 2.3 – Categories of BPMN object elements - [5]

In this section, we have introduced the concept of business processes and business process management (BPM) and Business Process Model and Notation (BPMN). We have also introduced the basic standards in this domain. In this thesis, we have used BPMN to represent our business process models. Our choice is justified by the fact that it is simple to use and to understand, and also its richness with all the elements that it proposes (Flow Objects : activities, tasks, events, gateways, complex gateways. Connecting Objects : Sequence Flow, Message Flow, Association. Swimlanes : Pool, Lane. Artifacts: Data Object, Group, Annotation.). Figure 2.4 illustrates a simplified business process model designed using BPMN. This model is made up of a pool *call center agent* that includes three activities: *search for an available agent*, *make the customer wait*, and *handle the call*, one gateway, and one data object *call history* linked to an activity by an association object.

Throughout the following section, we aim to introduce the core axe of this thesis, which is Business Process Improvement (BPI). We will propose a BPI overview to better understand the different enhancements proposed in the literature.

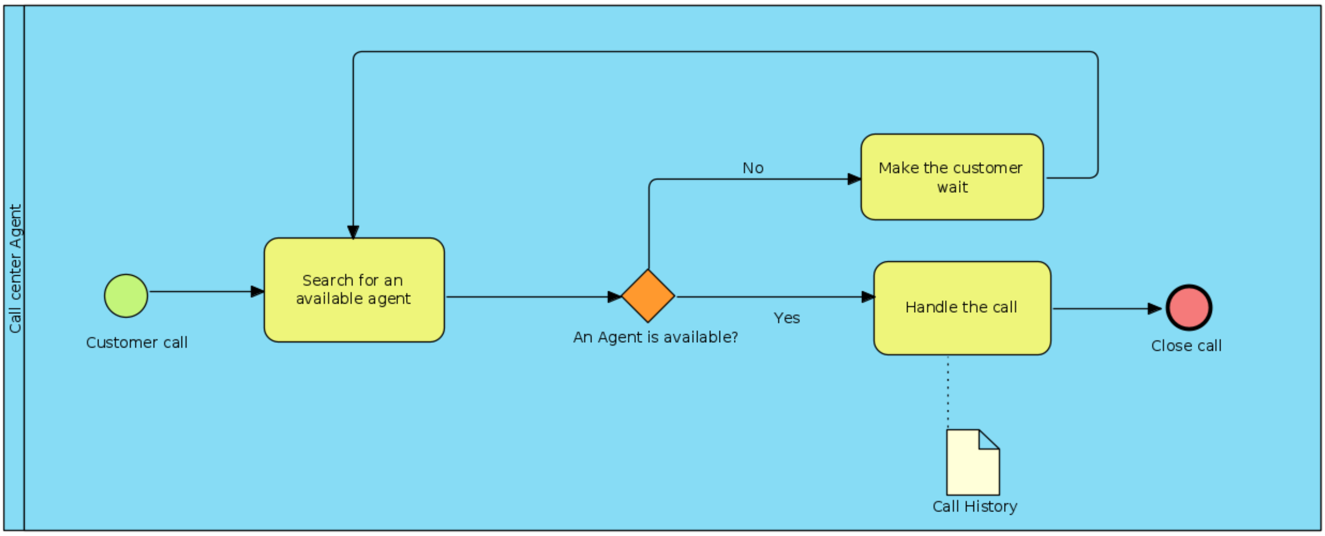

Figure 2.4 – Call center request treatment process

2.2 Business Process Improvement (BPI)

Business processes reflect how work is done and how business is operated within an organization. They are considered as the key to maintain competitiveness. That is why it is mandatory to apply a continuous improvement strategy to keep processes sharp and up-to-date within the organization.

Business process improvement and optimization is an important step in the business process life cycle. It is a crucial step for any organization that adopts a business process approach. In fact, engaging in a business process improvement journey helps the organization maintain competitive advantage and foster the creation of business value.

Improvement means "the organized creation of beneficial change; the achievement of unprecedented levels of performance." [86]. Performance is the achievement of an objective previously defined by the organization and that depends on the context of the study. Thus, Business Process Improvement represents any positive change that can be done in any organization to enhance the performance of their business processes. It is in fact a systematic approach that enables organizations, once adopted, to continuously enhance their existing Business Processes. It is considered an intrinsic aspect of effective Business Process Management [87], and as an incremental bottom-up enhancement of existing processes within functional boundaries [88].

Business Process Improvement is not a new approach; in fact, it was used for the first time by H.James Harrington in 1991 [89], where he provides a guide for organizations to start the journey of process improvement based on the quality improvement techniques used basically in the manufacturing sector. He described BPI in 1997 [90] as a "methodology that is designed to bring about step-function improvements in administrative and support processes using approaches such as process benchmarking, process redesign, and process re-engineering". The ultimate purpose of BPI according to J.Harrington is to make business processes more efficient, more effective, and more adaptable. Parallel to the concept of BPI, we have Business Process Re-engineering (BPR) that was introduced by Hammer in 1990 [15], it is considered as a radical methodology to redesign business processes from scratch, unlike BPI which aims to enhance the quality and productivity of the existing ones.

Other different methodologies have been used to describe the improvement and enhancement in the BPM field, albeit under different titles. We have tried to present these methodologies in ascending chronological order:

— **Process mapping:** Process mapping is a management tool proposed by General Electric as part of their work strategy to improve and enhance their bottom-line business performance [91]. However, this method finds its roots in the presentation "Process Charts—First Steps in Finding the One Best Way" by Frank Gilberth to members of The American Society of Mechanical Engineers (ASME) in 1921, at their annual Meeting [92].
A single graphic image can express the essence of a complex idea or concept that would require many pages of text to describe. As the Russian writer Ivan Turgenev once said “A picture shows me at a glance what it takes dozens of pages of a book to expound” [93]. Process mapping is founded on this principle. In fact, process maps represent a visual aid for analyzing work processes and aligning system elements in the same direction to allow process improvements [94], [95].

— **Lean:** The concept of Lean was originally pioneered by Toyota, a Japanese automotive company, in the late 1940s, in order to reduce the cost and to improve productivity by eliminating wastes or non-value added activities. During the 1980s, this approach witnessed a keen interest from the western manufacturers because of growing Japanese imports [96]. But it was until 1990, exactly after the oil crises in the early of 1990s, that the term "Lean" was intensively used in a published book named "The Machine that Changed the World" by James P. Womack and Daniel T. Jones [97]. Then it has been spread across countries and industries due to its overall efficiency in terms of quality, cost, flexibility, and quick responsiveness [98]. Lean represents a manufacturing strategy based on the idea of eliminating waste and increasing the activities value, by eliminating non-value adding steps and increasing value adding steps in the processes of the entire organization ([99], [100], [101], [102]).

— **Total Quality Management (TQM):** The concept of Total Quality Management was first espoused by William Edwards Deming in the late 1950s [103]. His ideas were not endorsed by American industry, but were eagerly embraced by Japan upon its recovery from World War II ([104], [105], [106], [107]). The foundations of TQM are rooted in the teachings of Dr. W. E. Deming, known as Deming 14 points [108]. The 14 points are intended to create a strong management commitment to quality, to process design and control through statistical tools, to continuous research and correction of quality problems, and to a purchasing policy that prioritizes quality over cost [109].

— **Kaizen:** Kaizen is the Japanese word for "continuous improvement" or "change for the better". According to Masaaki Imai, founder of the Kaizen Institute and author of "Kaizen: The key to Japan's Competitive Success" and "Gemba Kaiezn: A Commonsense Approach to a Continuous Improvement Strategy", Kaizen is not only about continuous improvement, but also an improvement every day, every-where and for everybody [110]. The term Kaizen started spreading in the 1950s as a part of the Toyota production system [111].

— **Benchmarking:** The core of benchmarking is the continuous process of comparision of a company's strategy, products, and processes with those of the world's leaders and best-in-class organizations [112]. The concept of benchmarking for business improvement was proposed by Rank Xerox in 1970. The formal definition of benchmarking used by Rank Xerox is: "a continuous systematic process of evaluating companies recognized as industry leaders, to determine business and work processes that represent best practices and establish rational performance goals" [113]. The idea behind benchmarking is to find and identify which companies are the best performers in an activity and then modify, adapt and implement their best practices according to the circumstances and operational requirements of the company, as no process is unique ([114], [115], [116], [117]).

— **Six Segma:** When properly applied, Six Sigma is seen as a recognisable evolution of Total Quality Management [118]. This approach was first developed in the late 1980s by Motorola, and it is considered as a disciplined project and data-driven approach and methodology designed to help organizations eliminate defects and flaws in a process - from manufacturing to transactions and from product to service [119], and enhancing customer value and efficiency [120], [121]. Using Six Sigma for process improvement involves five steps:

1. Identify and define the problem;
2. Measure the current process performance and assess defect levels;
3. Analyze the data and conduct an analysis of root causes;
4. Improve the process by selecting a potential solution for the problem;
5. Control the improved process to ensure that the improvements are sustainable and that the customer requirements or objectives are met.

— **Theory of Constraints (TOC):** Theory of Constraints is a management philosophy that focuses on the weakest ring(s) or part of the chain or the process that represents bottlenecks, to improve the performance of systems [122]. This approach was set by Eliyahu M. Goldratt in his book entitled "The Goal: A Process of Ongoing Improvement", in 1984 [123]. It focuses on the concept of constraints. According to Eliyahu Goldratt, a constraint represents " any element or factor that limits the system from doing more of what it was designed to accomplish" [123]. Therefore, TOC helps organizations maximize their profit by identifying and eliminating the process constraints (i.e., throughput) ([124], [125], [122], [126], [127], [128]).

— **ISO 9000:** ISO 9000 is now a well-known label to many organizations. It has dominated the field of quality management since 1987. The International Organization for Standardization (ISO) took the initiative to establish an international standard for quality systems with the help of several countries [129]. ISO focuses basically on quality control systems "from the process of product design to process design and from production process through to service after sale" [130]. The ISO 9000 family includes the fundamental concepts and principles of quality management that are universally applicable to organizations seeking sustained process improvement ([131], [132], [133], [134], [135])

— **Business Process Re-engeneering (BPR):** Reengineering is a term coined by Michel Hammer in 1990 [15]. The core idea behind BPR is a radical process re-design and rethink to ensure large-scale improvement and enhancement in business performance ([136], [137], [138]). Reengineering is different from almost all other process improvement approaches because "it does not focus on what is, but rather on what should be" [136]. So, the new process is actually designed from scratch and not from the existing processes.

In the literature, we find several others terms to design process improvement such as Business Process Change [139], Business Restructuring [140], Process Innovation [141] and Business Process Redesign [142]. The last one is mostly related to business process Reengineering [143], [144]. Each of these methods has its strengths and weaknesses. Companies are sometimes forced to test several methods to find the right one to help them improve the performance of their business. Choosing the most appropriate method depends on several criteria such as business area of expertise, resources, objectives, etc. In the following table (2.1) we present an overview of the key success factors and critical negatives aspects of each process improvement methodology.

As we have seen in the first part of this section, BPI affects the entire enterprise, including the employees who do the work, the information systems that support the processes, and the measures established to evaluate the effectiveness, efficiency, and adaptability of the process. As a result, the use of this methodology within an organization is becoming essential to evolve its performance and achieve its business objective.

In the following subsection, we will present an overview of methods and approaches used in the literature to improve the first step in the business process lifecycle, which is the modeling and design step.

Improvement method	Proposed by	Year	Key Advantages	Critical limitations
Process mapping	General Electric -	1921	Enables end-to-end process visibility across the organization [136] Enables knowledge transfer and exchange Process mapping is more than just a diagram of process activities, in fact it enables business process analysis by capturing meta-data at each step of the process	Accuracy of process data collection is mandatory to succeed in this approach Failure to define the boundaries of the process (the beginning and end of the process) [145] Not always obvious to link the goal of process improvement to the organization competitive priorities [136]
Lean	Toyota - Japan	1940s	Minimizing waste by limiting activities that do not add value to the process. Maximizing profits by reducing costs. Enhancing customer interaction, since that lean focuses on the needs of loyal customers.	Not simple to implement. High cost of implementation. Not easily accepted by employees.

TQM	William Edwards Deming/Japan	1949	Contributes to lower costs throughout the organization and its business infrastructure. The emphasis on engagement. TQM is known as "People's success", so the quality is the responsibility of all stakeholders, not the responsibility of production layer only. reducing non-productive activities.	The concepts of TQM meet especially the needs of large organizations and multinationals, than those of small and medium organizations [146], [147]. High cost of implementation. Generates resistance to change among workers because they feel that their jobs within the organization are at risk when a TQM program has been started. Creativity limitation because TQM focuses more on task standardization.
Kaizen	Toyota - Japan	1950s	Improving teamwork by encouraging the involvement of all stakeholders in the business process. Improving efficiency, quality of services and time management. Reducing waste in business processes. Avoids employee resistance, by introducing a softer approach to change.	The hole kaizen project may fail if the company keeps its closed communication method. Implementing kaizen in an existing system is difficult, and once implemented, it is very difficult to switch to the old management system. Because Kaizen strongly alters the current one.

Benchmarking	Xerox - USA	1970s	Encourage communication and knowledge sharing within the organization. Facing up to the competition by analysing the competitors business and strategy. The use of standardized business processes improves the overall quality and efficiency of the organization and minimizes the risk of errors.	Copying the ideas and practices of other companies can be a deterrent to innovation and original ideas. Sometimes it is difficult to obtain processes and information about leaders firms in the industry, this may limit and hinder the benchmark approach. Following the market leaders and copying their strategies increases dependency. Benchmarking implementation is very costly.
Six sigma	Motorola - USA	1980s	Optimizes processes, especially supply chain processes, and boosts customer satisfaction. Focuses on the whole production process and not only on the final results. Six Sigma is considered a proactive rather than reactive methodology because it helps the organization identify potential problems and proactively propose improvements before deficiencies are encountered.	No direct impact on customers, since it fails to involve, at the same time, both customers and suppliers. Limited results if the process on which six sigma is applied needs advanced improvement or redesign. Because this approach works better on a predefined process that needs only improvement [136].

TOC	Eliyahu Goldratt - Isreal	1984	Potential for significant productivity production capacity increase with minor changes in operations. One of the simplest techniques to communicate and apply. Ideal for promoting teamwork and also for initiating improvement actions, since it offers immediate and very noticeable benefits [148].	Sometimes difficult to apply if the process constraints are constantly changing.
ISO 9000	ISO - Europe	1987	Helps organizations to achieve better results in their business and increases productivity. Given that ISO standards are broadly recognized and widely regarded, compliance with ISO 9000 standards may enhance a company's attractiveness to prospective business partners. The standardization reduces audits. Facilitates continuous improvement and process monitoring.	Registration and assessment of ISO standards are costly and take time (sometimes it can extend beyond a year). Not suitable for small businesses or enterprises with limited resources.

BPR	Michael Hammer - USA	1990	Minimizing organizational complexity by eliminating unnecessary activities It is a "thinking out of the box" approach as starting from scratch when redesigning a process encourage new and innovative ideas BPR establishes strong communication between the different stakeholders and also promotes full commitment of the leaders Improving overall quality and reducing cost and time	Can be costly at both human and financial levels Human resources in this approach are considered as a cost to be reduced rather than as a resource to be developed Radical change can be risky sometimes

Table 2.1 – Advantages and limitations of the most eminent business process improvement methodologies

Quality improvement of business process models

As we discussed in the previous section, business process improvement goes necessarily through the process model discovery and design phase. The quality of a business process model is crucial for the successful completion of all phases of its life cycle. It is easier to implement, execute, and evaluate a high-quality business process model than a low-quality one [149]. The literature presents an assortment of methodologies and approaches to improve the quality of business process models, such as ontology ([150], [151], [152]), modeling guidelines ([153], [154], [155]), transformation rules ([156], [157]), refactoring technique ([158], [159], [160] [161]), etc.

The importance given to business process modeling in an organization is not recent. In fact, its origins are found in several already established approaches that were mainly interested in measuring and improving the quality of conceptual models. To our knowledge, the first approach to defining the concept of model quality was proposed by Batini et al. in 1992 [162]. Several approaches have been proposed to improve the quality of the models. These approaches can be classified into two categories [163]:

1. The first category deals with understanding and identifying quality.
2. The second category deals with quality measurement and evaluation.

Several researchers focused, thereafter, mainly on the first category, because it allows evaluating the quality of the models according to 3 aspects, namely [164] :

— **Syntactic quality** : The model should respect the rules and constraints of the modeling language used.

— **Semantic quality** : the model must correspond exactly to the domain being presented. This approach is based on understanding and mastering domain knowledge.

— **Pragmatic quality** : the model must be easily and, above all, correctly understood by the users. Understanding the model can be affected by the size of the model, the complexity of the domain, and the context described by the model, etc.

Among these three types of quality, semantic quality has recently been the subject of several research works in the field of business process improvement. Semantic quality is mainly based on domain knowledge, since this knowledge is crucial for understanding process models and evaluating the quality of process models [165]. Ontology is widely used in the domain knowledge-based approach to improve the semantic quality of the business process model [150], [151], [152]. Ontologies have been used since the 1990s mainly in the fields of knowledge engineering and artificial intelligence (AI), to facilitate the representation and modeling of knowledge [166]. "An ontology is an explicit specification of a conceptualization. The term is borrowed from philosophy, where an ontology is a systematic account of existence. For knowledge-based systems, what "exists" is exactly what can be represented"[167].

In software engineering and programming, refactoring represents “the process of changing a software system in such a way that it does not alter the external behavior of the code yet improves the internal structure” [168], [169]. The same concept can be applied to business process models. In fact, refactoring is a technique applied to enhance the internal quality of a BP model without altering its semantic and external behavior [160]. However, these techniques do not solve modeling problems or add new functionality to the process [149]. 11 refactoring techniques have been proposed by Weber and Reichert in 2008 [160], and are classified into three categories: i) refactoring for process model trees, ii) refactoring for process model variants, and iii) refactoring for model evolution. Fernandez-Ropero et al. proposed nine other refactoring techniques in 2012 [158]. These techniques aim to improve the quality of business process models by improving the comprehensibility and modifiability of BPMN models. In spite of the efficiency proven by these refactoring techniques to enhance the quality of existing business process models, they are still considered as static techniques that do not take into account the particularities of each BP model. Also, they cannot be applied during the modeling phase [149].

Transformation rules is another approach used to improve business process models. A transformation rule is an operation that adds and / or deletes elements (nodes, edges) to/from the BP model, provided that certain preconditions are met [156]. 28 transformation rules related to behavioral and organizational perspectives have been proposed by Wiem Khlif et al. [170], in order to improve business process quality by reducing the complexity of business process models. The behavioral rules aim to reduce the process models complexity by reducing the number of gateways and, therefore, the number of sequence flows. While the organizational rules reduce complexity by reducing the number of lanes in a pool and also the number of activities. To complete their approach, Wiem Khlif et al. in 2017 proposed a method called EVARES [157]. Their method is based on the set of 22 transformation rules that cover several perspectives of a business process model, including behavioral, organizational, informational, and functional perspectives. A heuristic algorithm was also proposed to determine the optimal order of these rules application, according to the designers preferences. Another set of four transformation rules has been proposed by Ivanchikj et al. in 2017 [156] to help designers improve the understandability of their process models. These rules aim to generate a new business process model, which is, indeed, syntactically different but semantically equivalent to the original model. In fact, these two models generate the same traces or execution logs, which proves that they are semantically equivalent. Their method is based on the trace equivalence notion proposed by van der Aalst et al. in 2006 [171]. Admittedly, all these transformation rules ensure a better enhancement of process models quality. However, there are so many rules in the literature, and the designers do not find an adequate tool system that can help them choose the most suitable rule to apply according to the particularities of each business process model.

To overcome the complexity of the business process modeling phase, researchers have proposed some good practices to help designers produce high quality business process models. These good practices are known as guidelines for business process modeling. Several frameworks for quality assurance of models have already been presented. Some frameworks focus on the quality of data models [162], [172]. Other focus on some specific modeling or conceptual requirements ([173], [174], [164]). A guidelines of modeling (GoM) framework have been proposed by Becker et al. in 2000 [153]. This framework is composed of six guidelines of correctness, relevance, economic efficiency, clarity, comparability and systematic design. These guidelines help to enhance the business process models and also the modeling process. Seven other modeling guidelines have been proposed in 2010 by

Mendling et al. [154]. These guidelines help organizations to either build a new business process model or improve the quality of an existing one. However, designers are sometimes overwhelmed by the abundance of these guidelines, especially in some cases where different guidelines can be applied to improve the quality of business process models. To overcome this problem, Mendling et al. propose in their approach a method to prioritize the different guidelines, based on the opinions of the experts in the organization. But it does not take into consideration the particularity of each business process model. Five others guidelines or recommendations have been proposed by Leopold et al. in 2016 [155], to overcome the three main problems that hinder the quality of business process models, which are: i) structural problems, ii) layout problems, and iii) labeling problems. In general, all these proposed guidelines offer a well-structured framework to improve the quality of business process models. However, they are not easily applied by designers because of their general nature, as they do not take into consideration the particularities of each business process.

As discussed in this section, business process modeling helps organizations understand the key mechanisms of their business [152]. Hence, the importance of having high-quality business process models that are easily understood and interpreted by both stakeholders and machines. The quality of a model is defined by how well it represents the needs of the business in a reliable and readable way. Therefore, we note that improving the quality of business process models represents a topic that is still relevant today. Several researches on this field aim at achieving a suitable alignment between business process models and domain needs and business requirements, in order to improve the performance of every service in the organization, and facilitate the work of business process designers.

Companies are becoming aware of the undeniable impact that improved understanding and modeling of business processes can have on the efficiency, consistency, and transparency of their business. Nevertheless, business processes do not have only a conceptual aspect, they also have other specific aspects related to the fact that they must be executed. These aspects need to be improved as well.

Performance improvement of business processes : Metrics, measures, and objectives

Business process models are a direct input into the software development process [152]. Subsequently, they have a great impact on the efficiency of the organization.

As we have seen before, a business process is a set of activities linked to one another by control flows and whose main purpose is to describe how a company achieves its objectives, following a managerial and transverse approach. Among all types of business processes, we distinguish three main types of business processes that are extremely important to the company, namely:

- **Management processes** : They ensure the steering and overall management of all processes in the organization.
- **Operational processes** : Also known as production processes. These processes are directly linked to the manufacturing in order to supply the clients with the product or service, and are therefore processes that affect the core business of the organization.
- **Support processes** : These are the processes that provide other processes in the

company with the necessary resources for their proper functioning. They contribute in an indirect way to the manufacturing of the product or service.

In order to meet customer demands, companies must achieve a good level of performance, especially on these three types of business processes. Achieving this level of performance comes down to the interest given not only to the quality of the business process models, but also to the performance of these business processes. Performance is basically linked to four core process aspects: process cost, process flow time, process flexibility, and process quality [175]. Therefore, in addition to improving the quality of these business process models, discussed in the first section, the organization must also address the performance aspect of its business processes. In addition, efficient business processes allow companies to gain a lot in terms of agility and adaptability, which makes the company capable of dealing with a changing business environment in terms of both challenges and opportunities.

Performance is an abstract notion. It can be defined or measured differently from one domain to another. Business process performance is measured using metrics known as Key Performance Indicators (KPI). Key performance indicators are defined as quantifiable or qualitative measures that allow organizations to gauge their effectiveness in achieving their strategic and operational objectives [176]. Key performance indicators are a means of measuring progress towards an expected result. They represent a measurable expression of the achievement of a desired level of results in an area relevant to the activity of the entity being evaluated. KPIs provide managers and executives with a means to track and assess their business performance and verify whether or not they are committed to their strategic plan. Key performance indicators differ and are based on industry-specific standards [176].

In the field of business process management, improving a business process means improving the intrinsic characteristics of this process. Indeed, business processes are characterized by specific performance measures and metrics. A. Shtub et al. [177] introduced what they call the four dimensions of competitiveness: Flexibility, Quality, Cost, and Time, those four cornerstones of competitiveness, are leading to the survival of an organization and to its success. F.Forster [88] used these same four measurement indicators of improvement : Time, Quality, Cost, and Flexibility, to develop business process improvement patterns. These four indicators are considered, in fact, as the most important measurement for evaluating the business activities in any organization. In the same perspective, these four main dimensions are widely used in the BPR literature [178], [179], [180], and they are part of what are known as the BPR best practices [181], [182], [183]. They are also known as the "Devil's Quadrangle". This concept was introduced in 1995 by Brand and Van der kolk [184]. In [179] SW. Cranenbroek explains each dimension as follows:

— **Quality**: which is divided into two types (internal quality and external quality). Internal quality refers to the social and psychological factors related to work. For example : variety in tasks, hierarchical differences, etc. External quality consists of the degree of how the products meet the requirements of the customers. For example: the default time in which the product can be used by the customer.

— **Time**: which can be represented in three main categories of time dimension according to Brand and van der Kolk : service time, queue time and wait time.

— **Flexibility**: which represents the degree to which organizations can adapt to changes in their (internal and / or external) environment.
— **Cost**: which can change based on the type of process and the resources (human and machine) used in this process. It includes different types of costs, such as production costs, transport costs.

For some other researchers, such as A. Yousfi [185], the purpose of Business Process Improvement is to improve those metrics by enhancing the quality of a business process, reducing its response time and its cost. That is why, they limit their work on business process improvement to only these three main performance metrics, quality (which refers to the disparity level between the projected outcome of the process (product or service) and its actual outcome), cost (which implies the expense of running the business process) and time (which represents the duration it takes the process to run from start to end). Whereas, G. Masiuk in [186] added three other metrics, when he introduced six key performance indicators for processes: Quality, Cost, Time, Quantity, Satisfaction and value. G.Masiuk classifies these six keys on four categories of process measures: first we have the input measures which represent the assessment of products, services, or information that feed into a process (quality, on time delivery). Second, we have process efficiency measures which represent the assessment of how well the process is functioning (cycle time, task time, accuracy, completeness, cost). In third, there are the output result measures which represent the assessment of the outputs of the process (quality, meeting employee/team/stakeholder requirements, quantity/volume produced). And finally the outcome measures which means the assessment of the outcome (impact) of outputs on the business and stakeholders (stakeholder satisfaction, business value, productivity). But according to G.Masiuk, acting only on these metrics do not generate a big change in process improvement; that is why an organization must implement and monitor a focused plan to reach a good level of process improvement.

To contribute to process improvement, S. Page in [187] chooses a different path by focusing on three main objectives of BPI:

— **Effectiveness**: which refers to the capacity of the business process to meet and satisfy the needs of the customers.
— **Efficiency**: which means that the process must be easily used and understood by the employees and also that it must minimize the consumption of resources.
— **Adaptability**: which represents the degree of flexibility of the business process with respect to changes in business requirements.

So to obtain effective, efficient and adaptable business processes, organizations have to follow a 10 steps Road-map to Business Process Improvement [187]: - Develop the Process Inventory, - Establish the Foundation, - Draw the Process Map, - Estimate Time and Cost, - Verify the Process Map, - Apply Improvement Techniques, - Create Internal Controls, Tools, and Metrics, Test and Rework, - Implement the Change, and finally - Drive Continuous Improvement.

As we can see, to specify the intrinsic characteristics of a business process, some researchers choose to work with business process metrics, others choose business process measurements, and others go with business process objectives. To enable organizations to choose the adequate business process characteristics in the BPI journey, we should first distinguish between a metric, a measure, and an objective. In fact, according to C. Kahraman et al. in [188] there is a difference between metric and measurement. A metric can be a measurement, but a measurement is not necessarily a metric. A measurement is simply some dimension, quantity, or determination of capacity; a metric is a standard of measurement by which the efficiency, performance, progress, or quality of a plan, a process, or a product can be assessed [189]. Whereas, an objective generally reflects the end goals based on the mission of a function. Based on this comparison, we noticed that business process improvement requires most of the time a combination of some metrics, measurements, and objectives to reach a scalable, flexible, and intelligent business process.

Nowadays, Improving business processes is no longer limited only to the quality of process models or the business process performance (by improving the predefined metrics and measurements). Because recently, organizations are becoming increasingly interested in achieving other aspects in their business processes, such as adaptability, intelligence, and proactivity. In the literature we found several approaches that aim to improve the intelligent and proactivity aspects of business process based on different techniques such as: ubiquitous computing [1], [185], machine learning [190], [57], [29], [191], "Complex" Event processing [26], [192], Process mining [37] [193], [194], Robotic Process Automation [195], [196], etc. Improving business process is a very open issue that still attracts more attention from the research community; most of these contributions focus on adding new improvement techniques in order to manage correctly and successfully their business processes, and to make them more adaptable, proactive, and smart. However , all these techniques have data as an important ingredient to achieve process improvement, that is why harnessing the power of these data can have a huge impact on business process improvement.

Data science as a tool for process improvement

Nowadays, in several companies, data are considered as a catalyst of innovation, development, and continuous improvement of their performance, their services, and their customer relationships. Data and events are continuously generated due to the omnipresence of sensors and networked machines in all major industries. Data ubiquity, as well as advances in machine learning algorithms, data analytic, data mining techniques, and artificial intelligence, increasingly encourage companies to exploit these new technologies to extract business value and insights from this data.

A proper functioning value chain is the backbone of achieving optimal customer service, competitive advantage, and efficiency for any organization. Data must go through a whole process to extract value and knowledge from them, which analysts and business users can translate into tangible business value. Although Data, information, and knowledge have slightly different meanings, they are frequently used interchangeably. Schreiber et al. [197] proposed the following distinct definitions and tried to build the definition of each concept based on the previous ones:

— **Data:** is uninterpreted as signals or symbols that are brought to our senses every minute in huge amount, such as strings of numbers, characters, integers, etc.
— **Information:** represents data equipped with a meaning that allows its interpre-

tation by another entity (human being, machine, etc.). As an illustration, the triggering of the fire alarm is not just an audible signal, rather, it is interpreted as indication of fire in the building, for example.

— **Knowledge:** is a set of assimilated data and information used to carry out tasks and create new information. According to Schreiber et al. [197], knowledge adds two other distinct aspects. The first one is a *sense of purpose*, given that knowledge is the "intellectual machinery" employed to achieve a goal. The second one is a *generative capability*, given that knowledge can be used to produce and generate new information, as it is one of the major functions of knowledge. That is why knowledge is heralded as a new factor of production.

The new digitized era and the rise of several new technologies such as big data, fast data, cloud computing, Internet of Things (IoT), etc., implies new business process problems and challenges linked basically to the tremendous amount of data and event data that are constantly collected within the organization. These data represent for enterprises a real engine of growth. However, a large amount of raw data is not valuable. Analyzing the huge amount of data generated by the zillions helps organizations extract useful information and then knowledge. In fact, the real value lies in how to use data and turn an organization into an information-centric company that relies on insights/knowledge derived from data analysis for their decision-making [198]. Hence integrating data science in each step of business process life cycle is very crucial for process improvement.

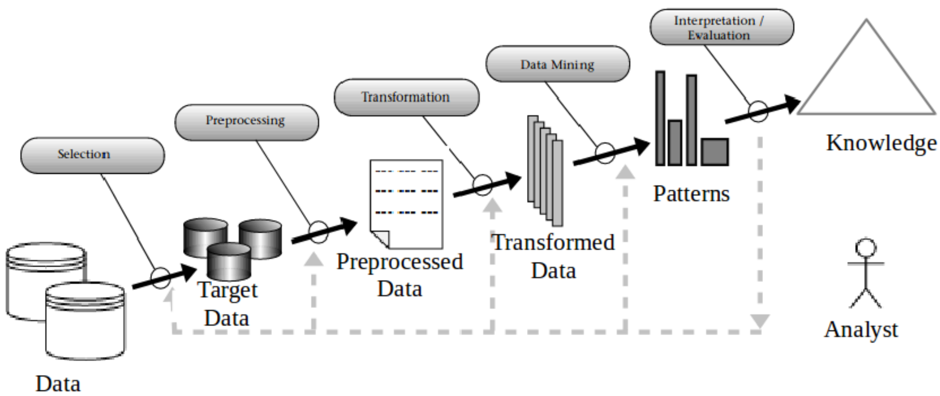

Figure 2.5 – Steps of the process of Knowledge Discovery from Databases (KDD) [6]

As we have seen previously, knowledge can be generated from data and information. This process of generating new knowledge is called Knowledge Discovery from Databases (KDD). KDD is an iterative and interactive data mining process that was developed in 1996 by Frawley et al. [199] and defined as " the nontrivial extraction of implicit, previously unknown, and potentially useful information from data". In 1996, Fayyad et al. [6] proposed a five-step process for KDD (see Figure 2.5). The first step is *selecting* a data set from the available data to begin the KDD process. This selection, on which discovery is to be performed, takes into consideration the domain knowledge, the purpose of this discovery or analysis, and also the objectives of the analyst. After that, the selected data is *preprocessed*, in order to have a tidy data, with no missing or noisy data. At this stage, the data can be *transformed* in order to have an accepted format that can be used by different mining algorithms, for example, Clustering, Classification, regression, or recommendation. Choosing one algorithm among others depends on the purpose of the analysis (see Table 3.2 in Chapter 3). The output of the mining step represents patterns that need to be *interpreted* and *evaluated* in order to validate their accuracy and to see whether they can be useful or not for the analysis.

Karl Broman once said that *"When physicists do mathematics, they don't say they're doing "number science". They are doing math. If you are analyzing data, you are doing statistics. You can call it data science or informatics or analytics or whatever, but it is still statistics."*[1]. However, statistics focus basically on generative modeling and theoretical results instead of real-world analysis that required prediction, or on real-world challenges related to data volume and data quality [9].

Data science is an interdisciplinary field aiming to turn data into real value. Data may be structured or unstructured, big or small, static or streaming. Value may be provided in the form of predictions, automated decisions, models learned from data, or any type of data visualization delivering insights. Data science includes data extraction, data preparation, data exploration, data transformation, storage and retrieval, computing infrastructures, various types of mining and learning, presentation of explanations and predictions, and exploitation of results taking into account ethical, social, legal, and business aspects. When analyzing this definition, we realize that data science is not just applying data mining techniques and statistics to liberate value from raw data. Data science is an *"an amalgamation of different partially overlapping (sub)disciplines"* [9], as shown in Figure 2.9. The combination of two or more data science ingredients allows to answer different data-driven questions. These questions have been grouped into four categories, by W. Van der Aalst [200]:

— What happened? (Reporting).
— Why did it happen? (Diagnosis).
— What will happen? (Prediction).
— What is the best that can happen? (Recommendation).

Enterprises that have shifted toward a BPM enhancement journey, in order to continuously improve their business processes, experience the need to deal with data and master the workflow of added value extraction.

It is noteworthy that data has a plethora of facets that can be exploited to help organizations enhance their performance. In a survey based on 325 responses [201], 70 percent of organizations consider (Big) data as "an opportunity to discover new facts about their customers, markets, partners, costs, and operations, and then they use that information for business advantage". But data is a double edge sword; it can either improve the quality of business and business process or deteriorate it, based on its quality and how we use it. In fact, there is a proportional relationship between the quality of the data and the quality of process, and subsequently the quality of decisions taken by the organization. This idea was detailed and clarified by Tamim et al. in [7] as illustrated in Figure 2.6.

1. https://kbroman.org/blog/2013/04/05/data-science-is-statistics/

Figure 2.6 – Impact of data quality - [7]

In the field of business process management, the data that are gathered after the execution of each business process instance is called event data or event logs. Figure 2.7 represents an example of an event log. The term event data refers to actions performed by entities.

Event_Source	Patient	Timestamp	Assessment Steps	Assessment Steps_id	Status	resource
9Chambre112	8	2017-01-05 04:56:11	IncidentRegistration	1	start	HR1
9Chambre112	8	2017-01-05 04:56:15	IncidentRegistration	1	Complete	HR1
9Chambre112	8	2017-01-05 04:57:05	Incident_transmission	2	Complete	HR1
18Chambre106	110	2017-07-26 02:51:04	Incident_handling	4	start	HR4
18Chambre106	110	2017-07-26 02:55:14	Incident_handling	4	Complete	HR4
18Chambre108	20	2018-04-25 23:04:09	Incident_closing	6	Complete	HR7
6Chambre204	103	2018-04-30 09:59:14	Incident_confirmation	3	Complete	HR3

Figure 2.7 – Example of incident process event log

Each event data must have at least three key pieces of information: action, timestamp, and state [202]. But for more accuracy, event data must have at least 6 different pieces of information [8] (see Figure 2.8):

- **Case identifier:** the identifier of the case to which the event belongs. Where a case represents an instance of the process that we are dealing with.
- **Activity identifier:** the identifier of the activity that the event refers to.
- **Activity instance identifier:** the identifier of the activity instance.

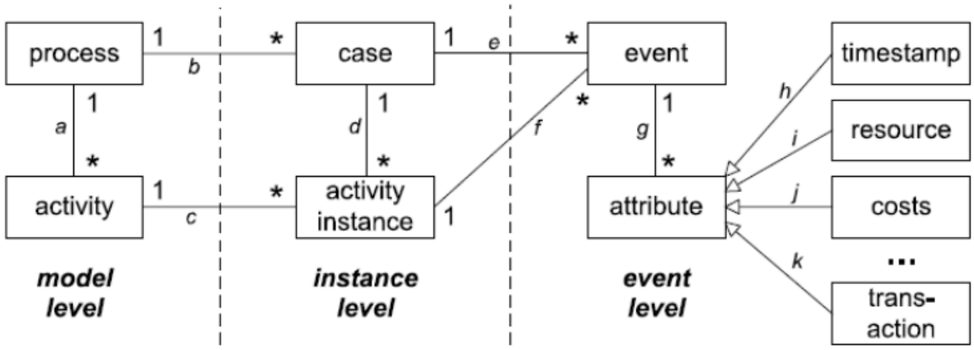

Figure 2.8 – Event data model [8]

— **Transactional life cycle stage:** the status in the transactional life cycle of the activity instance.
— **Resource identifier:** the identifier of the resource that execute the activities.
— **Timestamp:** the timestamp of the event.

Given the crucial role of data and event data, improving business processes mainly involves making BPM more data-driven and event-driven. Data science is a very broad field that includes several other fields (see Figure 2.9) such as machine learning, statistics, data mining, process mining, etc, and that provides several opportunities to improve business process management.

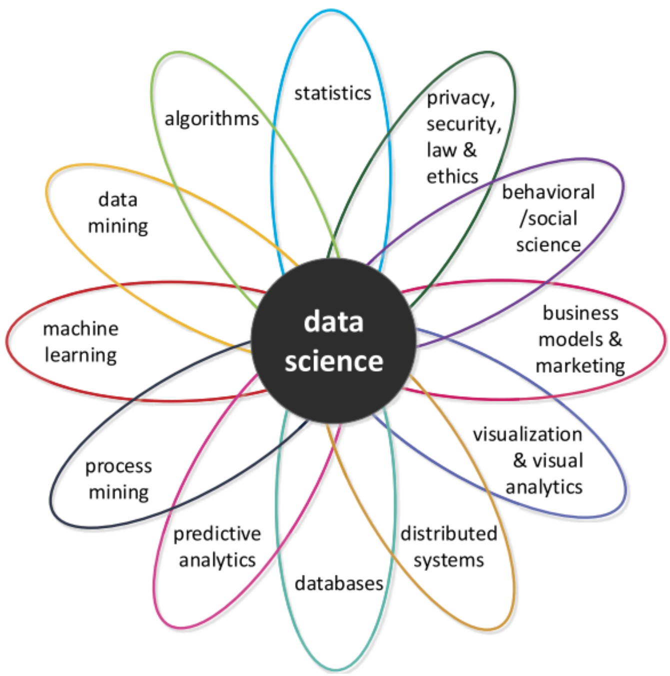

Figure 2.9 – Data Science ingredients [9]

In this thesis, we have explored the possibility of improving business process management using some machine learning algorithms and data mining techniques, and also some event processing techniques.

Process mining is also widely used in the field of BPM. It represents a set of techniques that are used to extract knowledge from event logs that are available in information systems, in order to discover, monitor, and improve real processes [37].

Improving business process is a very open issue that still attracts more attention from the research community, through the application of new improvement techniques, as we mentioned before, such as Robotic Process Automation (RPA) ([22], [203]), or Blockchain ([204], [205]), Advanced data analytics, Process mining, just to name a few, in order to correctly and successfully manage their business processes, and make them more adaptable, proactive, and intelligent.

2.3 Conclusion

In this chapter, we have introduced the main concepts of this thesis, namely business processes, business process management, and business process improvement. In fact, our contributions in this thesis can be considered as ways to enhance business processes especially for organizations that do not have big data processing capabilities.

After introducing the different concepts necessary for the understanding of our work, we have presented an overview of methods and approaches used to improve business process management from the first usage of BPI in the literature to the present day. The propositions discussed in this chapter made it clear that there are a lot of research work done in the field of business process management and business process improvement to help organizations enhance their business processes.

The next chapter focuses on our first contribution, which is enabling a priority-based instances scheduling within the frame of business process management.

Chapter 3

Business Process Improvement through instances scheduling

A plan is what, a schedule is when. It takes both a plan and a schedule to get things done.

Peter Turla

Business Process Management is concerned with continuously enhancing business processes. However, this cannot be achieved without effective resource allocation and priority-based scheduling.These are important steps toward optimizing time, cost and performance in business processes. Although there are several approaches and algorithms for scheduling and resource allocation problems, they do not take into consideration information gathered from past process executions, given the stateless aspect of business processes. Extracting useful knowledge from this information can help achieve effective instance scheduling decisions without compromising cost or quality of service. This chapter addresses the issue of enabling priority-based instance scheduling within the context of business processes. We pave the way for a combination approach, which is based on unsupervised machine learning algorithms for clustering and genetic algorithm (GA) to ensure the assignment of the most critical business process instance tasks, to the qualified human resource while respecting several constraints such as resource availability and reliability, and taking into consideration the priority of the events that launch the process instances. Our approach seeks to ensure that business process instances could be proactively scheduled based on priority determination. An overall example is proposed to illustrate our approach. It is inspired by the health care domain and the silver economy domain, which is a new industrial sector officially launched in 2013 in France [206], to create personalized services and new technologies that are expected to improve disability-free life expectancy or to help dependent elderly people and their caregivers on a daily basis. The risk of falls increases with age. In fact, loss of physical capacities due to age or some kind of accidents can lead to serious falls of elderly people and these falls can have adverse repercussions. Hence the need to propose an approach to ensure an effective scheduling in the case of critical tasks that must be executed by human resources.

3.1 Why Instances Scheduling ?

Business Process Management is about "continuous improvement and optimizing process to ensure high performance by achieving agility and flexibility as a tool to gain competitive advantages" [207]. Most of the existing studies in BPM focused on maintaining and enhancing the process business logical correctness, or improving the process performance at both levels: build-time and run-time, by focusing on the optimization of process modeling issues at build-time and process scheduling issues at run-time. Process scheduling is considered as a crucial step in the path to improve business process performance, since there is a strong link between effective resource allocation and business process improvement [208]. Scheduling is a fundamental aspect of business process management; it has attracted a lot of research efforts because it allows tasks and resources management and, therefore, performance improvement. In general, scheduling a process consists of respecting several constraints (task precedence constraints, resource availability constraints, data constraints, etc.) linked to process tasks and activities, and also in determining the (human/machine) resources to be allocated to these tasks. Business processes are different from other types of processes (e.g. scientific workflow), as they may contain automatic tasks and non-automatic tasks. Moreover, human resources are more difficult to manage, as a human resource can execute other tasks that do not belong to the main process, or they may be available for only a specific time slots. Additionally, several characteristics must be taken into account to choose the right human resource to carry out a critical task (especially in critical sectors such as healthcare or banking, etc.), such as availability [209], competence [210], seniority or reliability [211].

Resource allocation and task scheduling problems can be classified into two main categories [209]:

- **Off-line scheduling problems:** in this type of scheduling problem, the arrival dates of the tasks are known prior to resource allocation and task scheduling.
- **Online scheduling problems:** in this type of scheduling problem, the arrival dates of the tasks are not known in advance.

In this research work, we deal with business processes defined in an organization that cannot control the arrival of tasks (online scheduling), but at the same time it should maintain a balance between multiple constraints such as (priority, time, quality of service, lack of resources) to better manage resources and to minimize the overall execution time without compromising the quality of service.

Conceptual background

This section will describe some of the related researches that have been done to solve the problem of scheduling and (human) resource allocation in business processes.

Resources allocation and scheduling have been recognized as an important topic for business process execution, as they have an important influence on process performance, especially in cases that deal with minimizing the overall execution time of the process and also minimizing cost generated from resource allocation in processes scheduling operations, without compromising the proposed quality of service.

This issue has been intensively explored in the literature, and several important approaches have been proposed to optimize execution time and achieve an effective resource allocation and task assignment. Scheduling problems differ from one company to another based on different characteristics such as its domain of application, its objectives (performance, quality of service, fast scheduling time,...), the heterogeneous computing environments to where belongs the process in question (cloud computing or grid computing environments, ...), and the metrics or criteria to be optimized that are targeted by this organization. In fact, the chosen approach depends on the number of these criteria, and according to [212] the existing optimization approaches can be classified into two main categories based on the number of these criteria as we mentioned:

1. **Mono-criterion approaches :** this kind of approach is used to optimize (minimize or maximize) only one criterion. For example, cost optimization in [213] and load balancing optimization in [214].
2. **Multi-criteria approaches :** the optimization objective of these approaches includes several conflicting criteria, and the purpose is to ensure a balanced and optimized solution without compromising one criterion over another. For example, cost and deadline optimization in [215], cost, makespan and fault tolerance optimization in [216], and cost and makespan optimization in [217].

Proposing a scheduling and resources allocation approach consist in general on optimizing at least one of these criteria that we summarize in Table 3.1.

Resource management and scheduling problems in workflows and business processes generally belong to NP-complete or NP-hard complexity problems. For this type of problem, researchers tend to propose efficient heuristics instead of exact algorithms, because exact algorithms require an enormous calculation time that grows exponentially with the size of the input [233]. This issue has been extensively explored in the literature, and several important approaches have been proposed to optimize execution time and achieve effective resource allocation and task assignment. Researches propose different techniques and different algorithms to tackle this challenging problem of task assignment and resource management. To achieve efficient scheduling and high performance in parallel and distributed systems, several approaches have been proposed, such as the use of particle swarm optimization (PSO) for task-resource mapping to minimize the overall cost of execution and, at the same time, respecting constraints such as budget and load balancing [234]. Another particle swarm optimization (PSO) based heuristic for scheduling workflow applications in cloud computing environments was proposed by [235], this model was used for resource-task mapping in order to reduce execution cost (computation cost and data transmission cost). A comparison of the results obtained with the proposed heuristic with Best Resource Selection (BRS) heuristic has been done by the authors of this approach, and they found that PSO based task scheduling achieves three times cost savings and a better distribution of resource workload. The authors in [232] presented three workflow scheduling heuristics based on the Ant Colony System (ACO) to maximize the reliability of workflow execution and minimize violations of reliability and constraints. The simulation results show, according to the authors, that this method increases the feasibility ratio and ensures a high-reliability scheduling. In [236] an Artificial Bee Colony (ABC) algorithm was proposed, the authors in their approach considered

Optimization Criteria	Definition	References
Makespan	The time difference between the start and finish of a sequence of jobs or tasks.	[217] [218] [219] [216] [220] [221]
Response time	This is the time elapsed between the initial launch of a task execution and the first response.	[222] [223] [224] [225]
Waiting time	It corresponds to the average time spent in the system, by a given task, in order to be accomplished.	[218] [225]
Cost	An amount that has to be paid or spent to execute a given task	[213] [215] [226] [227] [216] [228] [217] [229] [230] [209] [222]
QoS	Is the measurement of the overall performance of a process execution	[231]
Efficiency	The ability to avoid wasting resources, cost, and time in executing a given task	[230]
Deadline	The latest time or date by which a given task should be completed.	[226] [215] [227]
Resource utilization	Resources used to execute a given task	[231] [222]
Fault tolerance	Is the ability to continue tasks execution where some resources are subject to failure.	[216]
Load balancing	Is the ability to efficiently distributing incoming tasks on available resources	[220] [214]
Reliability	Is the probability of a schedule successfully completing its execution,	[219] [228] [232] [29]
Availability	Is The times or shifts when a resource (human or machine) can execute a given task	[209] [208] [29]
Fairness	An equitable sharing of resources between the different instances of the process	[209] [224]

Table 3.1 – Scheduling optimization criteria

time as a QoS factor. This algorithm improves total task finishing time, load balancing time, and mean task finishing time when compared to the ant colony optimization algorithm. Besides several others Swarm intelligence heuristics proposed in the literature for resource management and scheduling problems, we also have Genetic Algorithms (GA). In [237] authors proposed a task scheduling algorithm using an improved Genetic Algorithm by merging Min-Min and Max-Min scheduling methods in standard Genetic Algorithm to achieve efficient tasks scheduling, so these tasks can be finished in a minimum time. The Greedy approach is another approach that has been widely used to solve different problems, such as scheduling and activity selection [238].

A business process scheduling problem becomes more complicated when we have Human resources, because human resources are more difficult to manage as they can execute other tasks that do not belong to the main process or they may be available for only a specific time slots. Human Resource Allocation Problem (HRAP) is considered as a special case of assignment problem. S.Bouajaja et al. write a survey on human resource allocation problems [239] where they present the main approaches proposed in the literature to solve HRAP in different real-life applications. Among these approaches, we find exact methods [240] or metaheuristics [241]. However, to deal with the human resource allocation problem in the context of business processes and achieve efficient resource allocation and scheduling in business processes, several approaches have been proposed in the literature. In [242] authors focus on the integration of the priority aspect of human resource allocation in the business process based on preferences. This approach also provides a mechanism for ranking resources. Another approach to resource allocation in business processes has been proposed in [222], where the authors tackle the problem of resource scheduling for a number of process instances by proposing two approaches based on heuristic rules to achieve rational scheduling at build time and to take into account different dependencies that may exist between instances at run time.

To the best of our knowledge, only few of these works present an effective instances scheduling based on event priority determination in incident management business processes. Because, the concept of priority has not been sufficiently addressed in the literature, as a scheduling optimization criterion (see Table 3.1). In fact, Cabanillas et al. [242], tried to integrate this concept into resource allocation and assignment, by extending the modeling concepts, but their approach focuses only on the conceptual and modeling level.

Besides, these existing research works do not take into account the stateless aspect of these processes, as such a process does not distinguish between events, and it treats each event independently and without taking into consideration information that can be gathered from the previous executions. In the next section, we present in detail our approach, which is based on a genetic algorithm and a clustering algorithm.

3.2 Proposed approach for Instances Scheduling

Each company must submit its business processes to a continuous improvement mechanism respecting their life cycle. However, achieving a high level of enhancement cannot be done without systematically integrating business process instances priority determination with business process improvement approaches. In fact, an optimized resource allocation based on instances priority ensures a positive impact on business processes performance, as it addresses time constraints and cost requirements without compromising the output quality. Some work on resource allocation focuses more on changing and adapting the structures of the business process to better fit the resources available in the enterprise [208], others try to ensure an equitable sharing of resources between the different tasks or process instances [209]. Regardless of the adapted approach, managing efficiently resource allocation and time consumption could become a very important competitive advantage especially for organization where time and resources are crucial for their business improvement. Scheduling approaches in business process management take into consideration a lot of constraints related to instances of a business process, such as execution start time, finishing time, and dependencies between tasks, in order to determine their priority. Despite this, instances of the same business process can still be executed in first-in-first-out order, which hinders the efficiency of the service, especially when one of these instances is launched by a critical event. Besides, this situation becomes more complicated when most of the tasks in this business process are executed by human resources.

To proactively manage the process instances, we build our approach based on the concept of priority. This priority is determined based on the criticality level of the event that initially launched these instances.

In this contribution, we couple the machine learning algorithm with the genetic algorithm following the outline below:

1. We determine the priority of the business process instance, as a first step, based on the criticality of the event that launched this instance (see Sub-Section 3.2.3).
2. We propose a Genetic algorithm to solve our optimization problem which aims to achieve an effective assignment of the most critical process instance(result of the first step) to the most available human resource, while respecting several constraints such as human resource availability and reliability (see Sub-Section 3.2.4).

This proposed approach is summarized in the following figure (see Fig.3.1). This approach is characterized by two phases. In the first phase, the training or learning phase (see the first part of Figure 3.1), a clustering model is built using the data contained in the execution event log or data set. Priority clusters are determined using this clustering model. These clusters are used to determine the instances' priority levels. We also calculate the reliability score for each resource at this level. In the second phase, the execution or run-time phase (see the second part of Figure 3.1), the learned model is exploited in order to obtain an estimation of the priority level related to each instance candidate that will be executed by a human resource. The instance candidates are the result of the filtering step, where every instance which does not represent an incident or does not require the intervention of a human resource is eliminated. The most important step in this phase is Matching a process instance to a suitable human resource, based on the priority level of the instance and the reliability score of the resource.

Before tackling each step in more detail, we must understand some vocabularies related to the business process scheduling problem (see Sub-section 3.2.1) in order to represent the formulation of our priority-based business process scheduling problem in Sub-section 3.2.2.

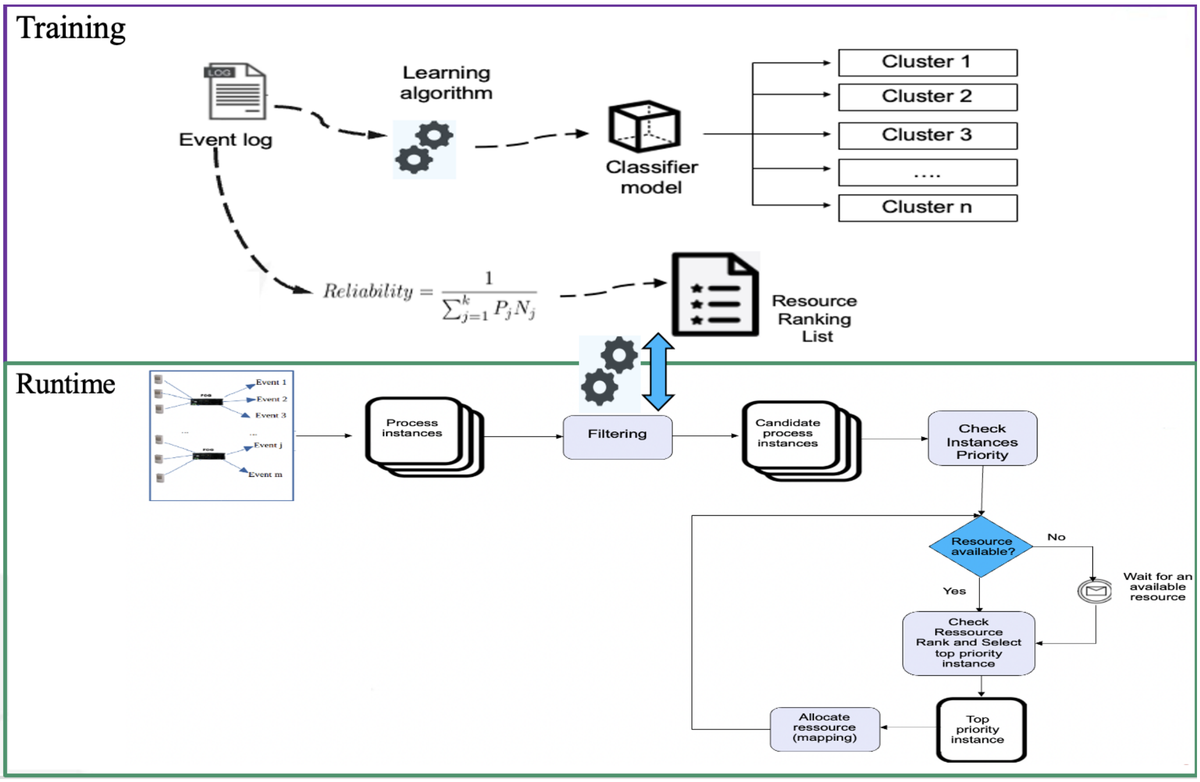

Figure 3.1 – An overview of the proposed approach

3.2.1 Definitions related to the business process scheduling problem

To understand the resource allocation problem in a business process that we discuss in this contribution, we need to define the following concepts:

— **Resource:** r represents a unit that can be a human or a machine used to execute tasks of a business process. A resource must fulfill several constraints such as availability, execution time and cost, in order to be suitable for a specific task. A is a set of agents (human resources), with n its cardinality.

— **Task:** t is a logical unit of work in a business process that can be executed by a set of human or machine resources, depending on whether this task is automated or not. T is a set of tasks, with m its cardinality.

— **Business process:** is a set of activities and tasks that exploit different resources to achieve one or more objectives.

— **Process instance:** is a specific execution of a business process that is characterized by the execution start time and the execution finishing time for each task in this instance.

— **Resource allocation:** is a match between a task t of a process instance and an appropriate resource r.

— **Constraint:** is a rule that controls execution tasks in a business process instance.

— **Priority:** : is a parameter used to choose between two or more tasks that need the same resource at the same time. The lowest priority task must wait for the resource occupied by the highest priority task.

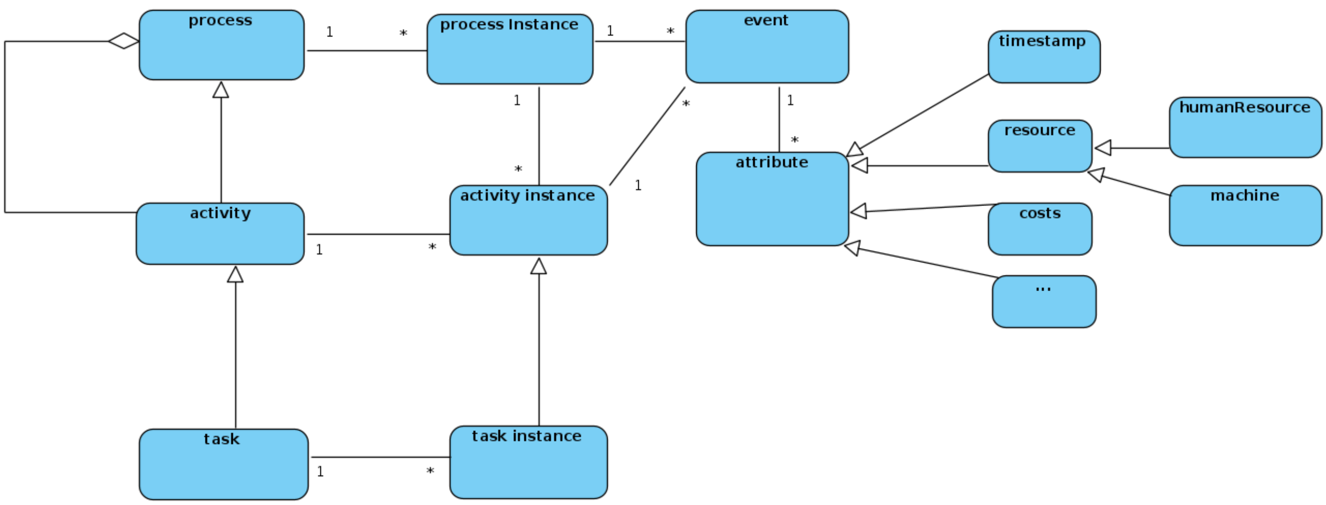

Figure 3.2 – Class diagram for process model basic concepts

Figure 3.2 represents a class diagram of a business process at three levels: model level, instance (or case) level, and event level. The core of this class diagram is inspired from the object model proposed by Van der Aalst in [9] (see Figure 2.8 in Chapter 2). A business process can have one or many process instances or what we call a case, and it is consisted of several tasks and activities. In fact, within the same process instance, we can have several instances of the same activity or task. It is worth pointing out that a task is considered as an activity, but not all activities are tasks, because a task is an atomic activity that is included within the process.

The class diagram illustrated in Figure 3.2 summarizes the following associations and cardinalities [9] :

- Each process can have zero or many process instances or cases. However, each process instance belongs to one process.
- Each process can have a different number of activities. However, each activity belongs to one process.
- Each activity may have a different number of activity instances. But each activity instance refers to one activity that belongs to a specific process instance.
- Each event is linked to a specific process instance.
- Each event is linked to one activity. But each activity instance can have several events.
- Atomic activity is considered as a task.
- Each event can have different number of attributes. But each attribute refers to one event.
- An event attribute has different subclasses, such as, timestamp (or the time of occurrence of this event),

3.2.2 Formulation of priority-based business process scheduling problem

The main objective of our approach is to ensure an effective and optimal human resource allocation and instance scheduling, while respecting the following constraints: 1) Priority of a process instance: the priority in our approach depends not only on the execution time interval, but also on the criticality of the event that triggers the instance. 2) Availability of human resources: in our approach, we have two types of availability: Initial availability, which is related to SLA (Service-level agreement) between the hired human resource and the company. And the availability at time t, which is related to whether a human resource is assigned to execute a task or not. To determine the time that a human resource will spend to execute the assigned tasks in order to determine his availability, existing approaches proposed several methods to estimate the available time slot of each resource based on the time that a specific task requires to be executed. However, to gain more flexibility and to ensure a real-time service, we propose in our approach, to manage the availability of each human resource, an online system that shows whether a specific human resource is available to receive a new task or he /she is not available (absent or allocated to an other task). 3) Reliability R_i of each human resource r_i: Since we are dealing with incident management business processes, the error rate must be very low, especially for critical tasks. That is why we include this metric which is calculated based on the number of errors that a specific human resource has made in a determined time interval.

$$Reliability = \frac{1}{\sum_{j=1}^{k} P_j N_j} \tag{3.1}$$

with P represents a weight that is proportional to the criticality level of the event. And N represents the total number of errors a human resource has committed while qualifying previous events, for each criticality level k, ($N \neq 0$ and $P \neq 0$).

The objective of our model is to minimize the total cost-reliability ratio for all available human resources (Equation 3.2). While respecting the constraints in order to ensure that a human resource can be assigned to one task at a time, we must also respect the human resource initial capacity and also his/her availability in order to assign to them only tasks that occur in their availability time slot.

$$\min_{(x_{1,1},\ldots x_{n,m})} \quad [F(x_{1,1},...,x_{n,m}) = \frac{C_{1,1}}{R_1}x_{1,1} + ... + \frac{C_{n,m}}{R_n}x_{n,m} = \sum_{i=1}^{n}\sum_{j=1}^{m}\frac{C_{i,j}}{R_i}x_{i,j}] \tag{3.2}$$

Subject to

$$\sum_{i=1}^{n} x_{i,j} = 1, j = 1, ..., m \tag{3.3}$$

$$\sum_{j=1}^{m} a_{i,j}x_{i,j} \leq Init\_Availability(r_i), i = 1, ..., n \tag{3.4}$$

The objective function represents the cost-reliability ratio, where c_{ij} represents the cost of allocation of human resources r_i to task t_j, and R_i refers to the reliability of each human resource r_i (Equation 3.1). x_{ij} in the first constraints, represented by the equation 3.3, represents the decision variable ($x_{ij} = 1$ if the human resource r_i is allocated to perform the task t_j; 0 otherwise). This constraint means that each task is assigned to only one human resource. In equation 3.4 a_{ij} represents the total time used by the human resource r_i when assigned to execute a task t_j, and this equation means that the total time used by each human resource cannot exceed his/her initial availability.

We schematize in Figure 3.3 the end-to-end process to achieve a priority-based and reliability-based resource allocation for our approach.

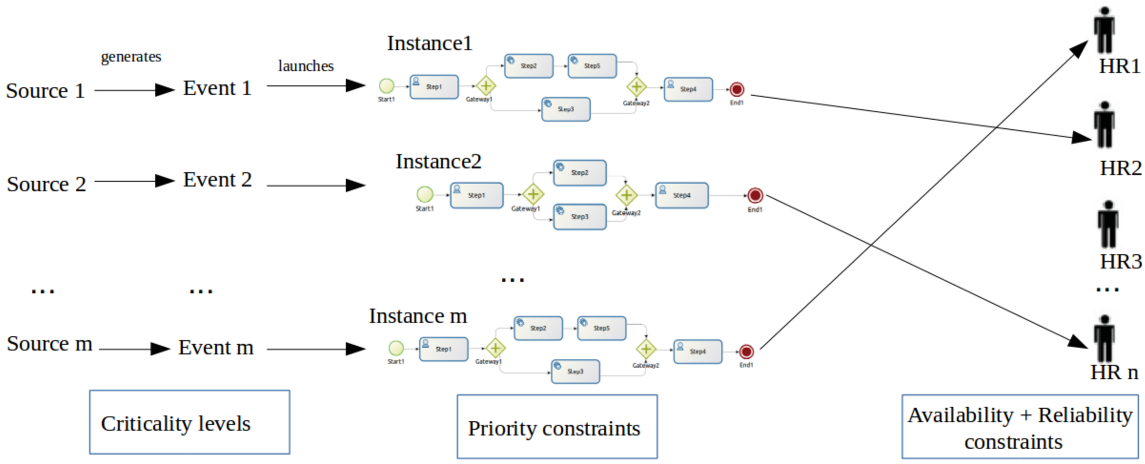

Figure 3.3 – Priority-based scheduling of process instances under human resource constraints

3.2.3 Machine Learning for process instance priority determination

- Basic Definitions

Machine learning is a branch of the artificial intelligence research domain. Using mathematical methods, machine learning enables systems to learn from data and generate knowledge from experience. With time and more experience, the system can learn and improve and sharpen a model that can be used to predict outcomes of questions using previous learning [243]. Machine learning algorithms are organized into different categories based on the learning type: Supervised Learning, Unsupervised Learning, Semi-supervised Learning, Reinforcement Learning, and Transduction learning:

— **Supervised machine learning** consists on having input variables (X) and an output variable (Y), so all data are labeled, and we use an algorithm to learn the mapping function from the input to the output.

$$Y = f(X) \tag{1}$$

The main purpose of supervised learning is to get an approximation of the mapping function so that we can predict the output variables (Y) when we have new input data (X). Learning stops when the algorithm reaches an acceptable level of performance. The process of an algorithm learning from the training dataset in this type of machine learning algorithm can be seen as a teacher supervising the learning process, which explains the name of this learning type.

— **Unsupervised machine learning** consists on having only input data (X) but no corresponding output variables, so there is no "supervising teacher", and all data are unlabeled. That is why the main goal of this learning type is to model the implicit structure or distribution in our data to learn more about these data.

— **Semi-Supervised machine learning**, this particular case of learning, which is between supervised learning and unsupervised learning, is used when we have a considerable amount of input data (X) but only some of the data are labeled (Y). In this case a mixture of both learning techniques can be used [244].

— **Reinforcement machine learning** problems concern how an agent must learn behavior and take action in a dynamic environment in order to maximize a cumulative reward. In this type of learning problem, the agent emulates the leaning behavior of humans by following a trial-and-error process, instead of having, in advance, a set of desired actions [245].

— **Transduction learning** is used in the field of statistical learning theory to refer to the prediction of new outputs based on training inputs, training outputs, and new inputs [246].

Supervised learning and unsupervised learning algorithms are the most used in real-world problems. They can be further grouped into different categories that we can find in real-world machine learning problems (see Table 3.2).

- Clustering problem

As we have mentioned in Table 3.2, we resort to clustering techniques when we want to discover the inherent grouping or clusters in an unlabeled data set. Clustering is the most important technique for unsupervised learning. It means dividing a large data set into distinct and manageable groups called "clusters", based on certain similar patterns.

Clustering techniques help us to discover hidden patterns or relationships between the data points in our data set that contain unlabeled data. We use these patterns similarities to determine whether a data point belongs to the same cluster as another data point. This similarity is measured based on distance. In fact, there are several methods to measure this distance from one data point to another. Among these methods we have : Euclidean distance, Hamming distance, Manhattan distance, Minkowski distance, Mahalanobis distance ([248], [249]).

In the literature, we find several clustering algorithms. According to Xu et al., clustering algorithms can be classified into two categories, the traditional clustering algorithms, and the modern clustering algorithms. The first category contains nine subcategories which contain 26 commonly used algorithms (see Table A.1 of Appendix A). The second one contains 10 subcategories that include 45 algorithms (See Table A.2 of Appendix A). [249].

Machine Learning Algorithm	Learning Problem	Definition	Example
Supervised Machine Learning	Classification	A classification problem is when the output variable is a category (spam and no spam)	Naive Bayes Classifier, Support vector machines (SVM), Random forest, K Nearest Neighbors (K-NN) ....
	Regression	A regression problem is when the output variable is a real value (weight, ...)	Linear regression, Random forest, ...
	Recommendation	Recommender Systems (RSs) are software tools and techniques used to provide suggestions for items to be of use to a user [247]	Collaborative Filtering, Content-based Recommendations, Context-aware Recommendations, ...
	Time series prediction	It describes a model that predicts the future value based on current and past data samples, in order to see how a given variable changes over time	Autoregressive Moving Average (ARMA), Moving Average (MA), Autoregressive (AR)
Unsupervised Machine Learning	Clustering	A clustering problem is where we want to discover the inherent groupings or clusters in the data (clusters of customers by purchasing behavior)	k-means, Density-Based Spatial Clustering of Applications with Noise (DBSCAN), Gaussian Mixture Model ...
	Association	An association rule learning problem is where we want to discover rules that describe large portions of our data (customers that by $Product1$ also tend to buy $Product2$)	Apriori algorithm for association rule learning problems

Table 3.2 – Machine learning algorithms taxonomy

In our approach, we have chosen K-means algorithms. However, this choice was based on a comparative study that we conducted. In fact, we have chosen to test and compare five clustering algorithms among the most widely used clustering algorithms in the literature. K-means, DBSCAN, Fuzzy C-mean, mean shift, and Spectral clustering. In order to analyze each clustering model of the five selected, and compare their differences and adaptability to our case study, we decided to apply these five algorithms on our case study to analyze the alerts (falls) of each patient. This comparative study is detailed in the Appendix A. This comparative study concludes that, for our case study and based on the dataset we have, K-means is the most suitable clustering algorithm for our approach.

- Instance priority determination Step

As mentioned previously, in order to schedule our business process instances according to their priority, we estimate this priority based on the criticality of the events that launch these instances. We proceed to a dynamic clustering in order to score and estimate the priority of the incoming event based on the cluster of its source.

We opted for clustering algorithms to discover groups in our dataset, we choose K-means clustering algorithms, and we tested several criteria such as the frequency of falls or total number of falls, in order to have the most representative clustering for our data. We apply the K-means algorithm on a set of events sources in order to classify those sources on different clusters using a score that we calculate for each event's source (a patient in our case) based on the frequency of previously generated events and their criticality value given previously by the agents (human resources) in the qualification step (see Figure 3.5).

This first step of our proposed method uses basic iterations of the K-means algorithm. The basic steps of the K-means algorithm are shown in the following pseudocode (see Algorithm 1):

Algorithm 1 K-Means clustering algorithm

Input:
S = s_1, s_2..... s_m // list of data points (list of sources which generate the different events)
K //Number of clusters

```
choose K Random data points from S as initial clusters centroids
repeat
    Assign each data point s_i to the cluster which has the closest centroids.
    Calculate the new centroids of each cluster.
until Convergence //no more changes for centroids
Output: Set of K clusters
```

The event criticality ranged from low level (0) to very serious (3), and there is a bijection between the event criticality levels and the instance priority. Two scenarios are encountered when applying this approach:

1. The sources of the incoming events belong to different clusters: in this case, the score of each cluster helps us to determine the criticality level of each event, which

helps us to estimate the priority level of the business process instance launched by this event. So, the instance launched by an event that was generated from a source that belongs to the critical cluster has higher priority than the other instance.

2. Both sources, which generate the events, belong to the same cluster: in this case, the criticality level of each event is determined by the comparison of the score (used to cluster the sources) of each event source.

3.2.4 Genetic Algorithm for human resource allocation

- Basic Definitions

Swarm intelligence is a discipline that is inspired from the collective and simple behavior of a swarm in nature and also the self-organazing interactions among agents. Swarm intelligence was first used in 1988 by G.Beni and J.Wang for cellular robotic systems [250], since then several Swarm intelligence Algorithms (SI) have been proposed to address complex optimization problems. Those algorithms can be classified into different categories: Insect-based algorithms such as (Ant Colony Optimization Algorithms, Bee-Inspired Algorithms, Firefly-Based Algorithms, Glow-Worm-Based Algorithms), Animal-based algorithms such as (Bat-Based algorithms, Monkey-Based algorithms, Lion-Based algorithms, and Wolf-Based algorithms) and Bird-Based swarm algorithms (Levy flights, Cuckoo search). SI algorithms have been applied in different domains and have proven their efficiency in solving NP-hard problems in several real-world applications. The growing popularity of these SI-based algorithms is due to multiple reasons inspired by the swarm behavior [251]: flexibility, self-learning capability, versatility, self-organization, inside out and outside in Interaction(interaction between the member of the swarm, and the interaction with the environment), collective vigilance (the swarm members are always open and alert to new information), Independence, Dynamic system (a system that reacts to changes in the environment). At first, those bioinspired algorithms deal with stationary optimization problems, but nowadays most real-world optimization problems occur in a dynamic environment [252].

In addition to Swarm Intelligence algorithms that have proven their efficiency, we find Genetic Algorithm (GA). It is a metaheuristic that was proposed in 1975 by John Holland and belongs to the evolutionary algorithms group. GA is a search optimization algorithm based on the mechanics of the natural selection process. The underlying concept of this algorithm is to imitate the concept of "survival of the fittest". In fact, this algorithm simulates the processes observed in a natural system where the strongest tend to adapt and survive, while the weak tend to perish [253]. In this population composed of many individuals, each individual is uniquely determined by its gene, and couples of individuals give birth to offspring and thus generate a new population. The genes of the offspring are the product of their parents mixed genes, which ensures that the genetic traits that contribute to the survival of the individual are fixed in the population.

The basic steps of the genetic algorithm are shown in the following pseudocode (see Algorithm 2:

Algorithm 2 Genetic algorithm pseudo-code

Begin

Randomly generate an initial population of different individuals
Evaluate the fitness of each individual of the population
repeat
    Select two parents from the population
    Generate offspring by the selected parents
    randomly Mutate the offspring
    Evaluate the fitness of the offspring
    Replace the less important individuals in the initial population by the best ones from the offspring
until convergence criterion is met { time limit or specific number of iteration}

Resource allocation and task assignment are among the most important optimization problems in Business Process execution that has been extensively studied in the literature; many researchers have proposed several algorithms for scheduling problems, resource management, and task assignment [254] [255] [256] [242]. Genetic Algorithm and Swarm Intelligence algorithms are one of the methods used in this optimization problem that have proven their efficiency [238] [257] [252], [258], as we have said previously.

- Instance Tasks and Resource matching Step

Meta-heuristics present a potential solution for scheduling problems when exact methods are unable to find an optimal solution within a reasonable computational time.

The use of a meta-heuristic in our approach is intuitive as we are facing an optimization problem, and meta-heuristics have proven their efficiency and their capability to obtain near-optimal results, through several works previously done by researchers. But we opted for the Genetic algorithm instead of other metaheuristics such as the Artificial Bee Colonies algorithm (ABC) or the Ant Colony Optimization algorithm (ACO), as it was more adaptable to our case. In addition, the phases of GA offer more flexibility in order to propose modified or adapted algorithm versions by researchers. In fact, we also proposed in this approach an adapted version of the genetic algorithm previously described. Our optimization approach consists of the following phases:

— **Input parameters and Population initialization** Like other population-based search and optimization algorithms, the initial phase of the genetic algorithm begins with generating the initial population and setting the initial parameters. A population in the genetic algorithm represents all possible solutions to the problem, and an adequate representation of a population of candidate solutions increases the efficiency of the GA results. For our approach, each individual from the initial population is encoded as a vector where the first element of this vector represents the human resource index and the second one represents the task index. An individual in our case is represented as a possible one-to-one matching between a human resource and process instance tasks. Therefore, our population will have the following representation (see Figure 3.4).

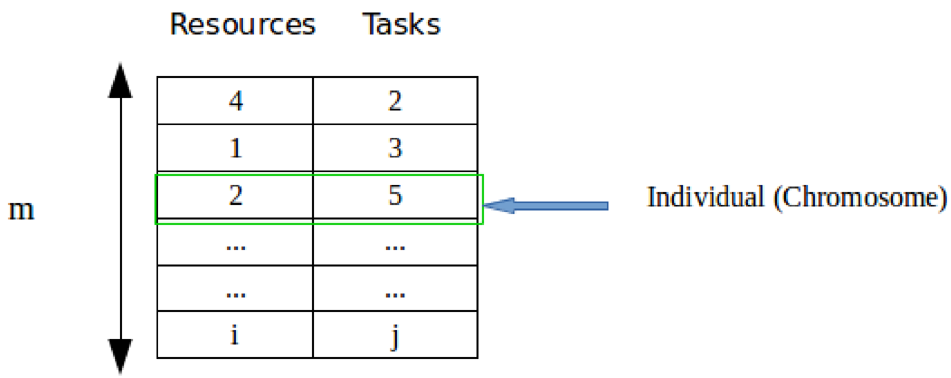

Figure 3.4 – Representation of population of candidate solutions

— **Population fitness-based evaluation:** As we mentioned before, our purpose is to ensure one-to-one matching between a human resource and process instance tasks. We evaluate the fitness value of each individual in the population. This fitness represents the total cost-reliability ratio of the available human resources that will be allocated to current tasks (see Equation 3.2).

— **Parent Selection and Population reproduction:** In this phase, the individuals of the initial population (parent) are sorted according to their fitness values. Among the different selection techniques in literature (Tournament Selection, Roulette Wheel Selection, Rank Selection, ...), we apply rank selection. This technique consists of sorting individuals according to their fitness score, and then we randomly choose the parents from the individuals with higher ranks.

— **Crossover phase:** is a step in the genetic algorithm that consists of selecting two random individuals (chromosomes) and switching between their elements (genes) to generate a new population. In our approach, we can only use the one-point crossover strategy for the individuals of our population given their representation (see Figure 3.4).

— **Mutation phase and New generation:** is an operation in the genetic algorithm that consists on randomly modifying an individual. In our case, we opted for selecting the first element of the individual (chromosome) that represents the human resource index and modify it with an index of another available human resource. To obtain the future population, we use the 'elitism' with a fitness-based selection approach, which consists of keeping the fittest individuals of the current population, and those individuals replace the least fit offspring in the new generation.

— **Termination Condition:** The termination condition is basically used to prevent the genetic algorithm from endlessly iterating. It can be stopped automatically after defining specific iterations. We can use either the fitness value or the time [258] as a termination condition in a genetic algorithm. In the literature, we also find some research work that proposes other termination condition or stopping criterion, such as variance [259]. But time is a crucial factor in our case study, since we are dealing with critical events (falls of elderly people). So, we use a temporal termination condition. After a fixed time (e.g. 10 minutes), the algorithm is stopped, and the cost-reliability ratio (fitness value) corresponding to the individuals is retained as the final result. The individual with the lowest ratio is kept as the result of the genetic algorithm. This approach allows us to limit the duration of the algorithm, which can run for a very long time.

3.3 Case study : Scenario of the incident management process

In this section, we present a real-life scenario from our case study at Angel Assistance [2], to illustrate our problem and highlight the challenges we are trying to solve with the proposed approach.

The case study of our research work belongs to the silver economy domain, which is a new industrial sector officially launched in 2013 in France [260]. The aim is to create personalized services and new technologies to improve disability-free life expectancy and to help dependent elderly people as well as their caregivers on a daily basis. This domain represents a solid proof for why priority-based process instance management is critical.

Most countries around the world live in the demographic transition of an aging population. According to the United Nations, the number of people with 80+ years of age will triple between 2015 (126.4 million) and 2050 (446.6 million). If we take France as an example, in 2015 the number of people over 60 years old is 12 thousands, which represents 18% of the French population and will represent more than 1/3 of the population by 2060 [3]. Since demographic change is becoming a global phenomenon, several companies are focusing on developing products and services to create age-friendly societies.

The risk of disease, loss of capacity, and falls increases with age. Losing physical capacities due to age or some kind of accidents can lead to serious falls of elderly people and those falls can have adverse repercussions. In fact, The physical consequences of a fall differ from one individual to another. They can represent a decrease in mobility and an increase in daily life activities dependency. Falls also have some psychological consequences such as a loss of self-confidence, which can accelerate functional capacities decline. Falls among seniors result in a significant number of hospitalizations, with hip fractures being the main cause. Furthermore, falls are the leading cause of injury-related death.

Every year, more than 12 million people over 65 years of age fall and one in two people over 80 years of age is a victim. As the leading cause of accidental death in the over 65s, falls often have an impact on physical condition but also on psychological condition [4]. Loss of confidence, fear of falling again, withdrawal, the consequences of a fall are multiple, often serious, and linked to the importance of the injury and the health status of the person.

Several studies have been conducted in the field of silver economy, in order to determine a standard definition of a fall and the number of falls over a specific period to consider

2. http://c2ime.eu/portfolio-item/angel-assistance/
3. https://www.insee.fr/fr/accueil
4. https://www.pourbienvieillir.fr/sites/default/files/830.pdf

an old person as a repetitive case. In [261] [262] [263] [264][265] a fall represents "an unintentional change in position resulting in the return to rest at a lower level or on the ground". To characterize the repetitive aspect of a fall, we must determine the number of falls and the time interval between falls. As we can see in Table 3.3, most published studies consider at least two falls to retain repetitive character, with an interval between two falls ranging from 6 to 12 months on average.

Table 3.3 – Repetitiveness aspect of a fall.

References	Data Collected	Participants	Study plan	Falls repetition
[262]	Questionnaire A follow-up period of one year	N = 730 Age >= 55 years	Transversal	Yes >= 2/12 months
[263]	Postal questionnaire A follow-up period of 1 year	N = 1660 Age >= 70 years	Transversal	Yes >= 2/12 months
[264]	Telephone questionnaire Interval : 6 weeks A follow-up period of 36 weeks	N = 311 Age >= 70 years	Observational cohort	Yes >= 2/9 months
[265]	A follow-up period of 3 years Participants report their falls weekly on a fall calendar Phone contact in case the person is incapable of filling in his calendar	N = 1365 Age > 65 years	Observational cohort	Yes >= 2/6 months
[261]	Participants interviews Retrospective (12 months)	N = 377 Age =78 +− 3 years	Transversal	Yes >= 2/12 months

Quick intervention after a fall, for example, using a fall detector, could avoid 26 % hospitalizations, that is, 160 M€ and 9,400 deaths per year. There are several solutions for fall detection, such as:

- Passive Solutions: where the senior must press a beeper to notify in case of an incident.
- Active solutions: These solutions require the use of sensors (accelerometer, biological signals) or environmental detectors (presence, ground, doors, etc.). In case of a particular variation of the signals, the device triggers an alert.
- Video-surveillance Solutions: the camera sensor analyzes the senior's behavior and triggers the alert accordingly.

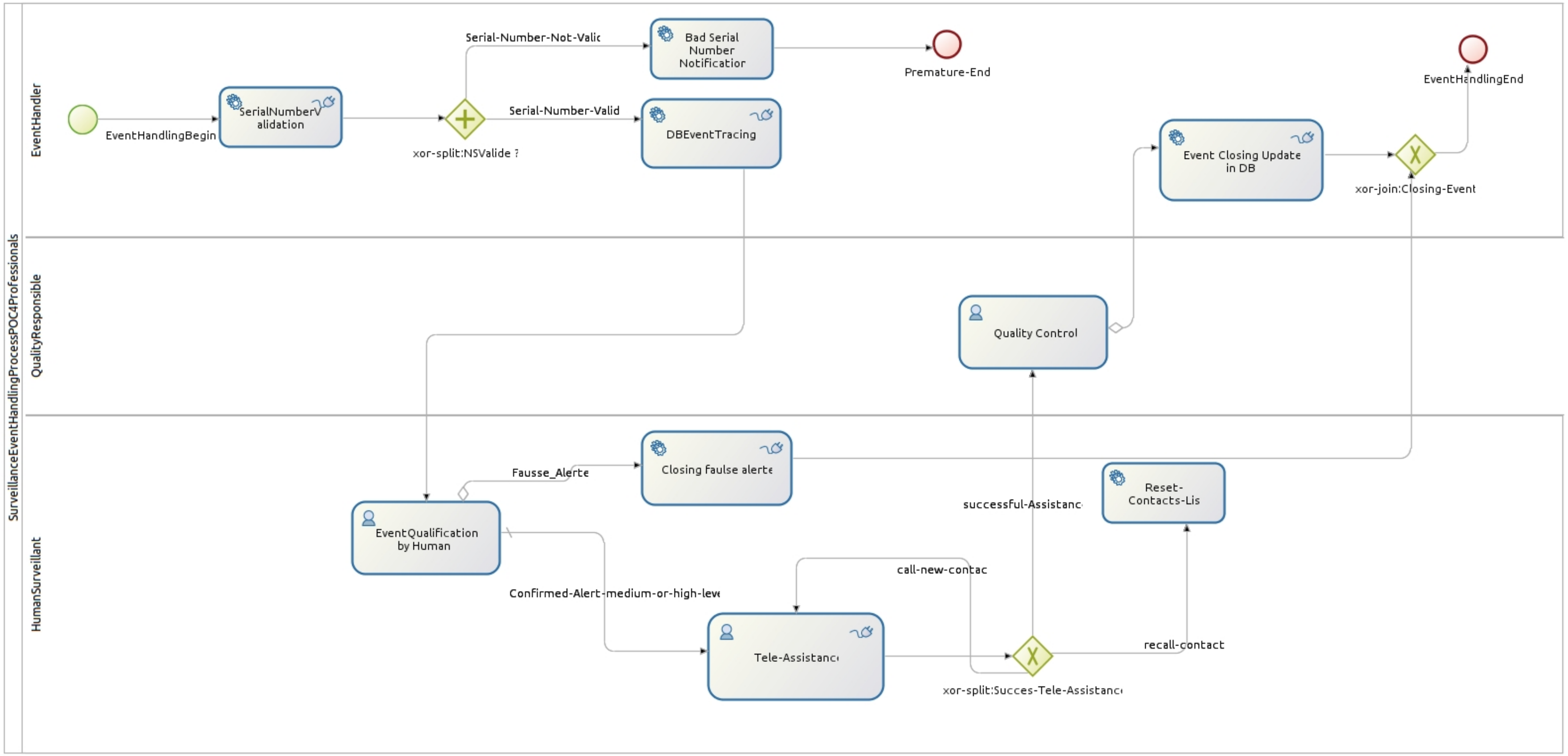

Figure 3.5 – Motivating Scenario : Qualification and Assessment of the risk level of incidents process

Some of these solutions (bracelets, presence detectors, active floors, etc.) are "blind". This means that they do not allow to know if a fall is serious or not, based on the received alert. Only the image delivered by the video fall detectors allows us to remove doubt about the incident, and therefore avoid unnecessary interventions and therefore minimize the overall cost of the service.

Predicting and preventing falls among elderly people is the main objective of our case study; in fact, to apply our approach, we will use a data set and a business process model from a video surveillance company. This company edits an automatic fall detection system for older people and offers a 24/7 automatic alert solution and rapid rescue without the intervention of the person in danger. This process is compliant with the ISO 9001 corrective / preventive action process. ISO 9001 is a standard that introduces new approach to corrective/preventive actions and continual improvement [266]. Corrective action means dealing with the problem after it occurs to find the root cause of this problem. However, preventive action means dealing with the problem before it happens in order to prevent it. The global business process of our illustrative example is simple, but it represents several hard functional constraints such as: Business scaling, real-time data analysis, and the obligation to maintain limited resources for the viability of the business.

Figure 3.5 depicts our incident management process, using BPMN ((Business Process Model and Notation), Object Management Group (OMG)), which aims to manage falls alerts from detection to assistance and resolution. This incident management process is based on an analysis in real time of alerts received from 24/7 streaming cameras (IoT devices) for detecting elderly people's falls. Waiting too long (sometimes even for a few minutes) can be so risky as it can complicate the situation and also can be very painful for the person. Therefore, a quick rescue is mandatory to help the person after a fall or an incident. To achieve this prompt intervention, smart video surveillance cameras are installed at the client's home or in patients' rooms in geriatric services. These devices detect suspicious scenes that may be a fall or an incident, take a picture of the scene, and then automatically send an alert to the video surveillance center. That received alert is handled by a human agent, who qualifies the alerts into four categories, as described below, and after that he/she determines whether an assistance action is necessary or not according to the criticality level of the alert. That is why, each received alert requires quite vigilant treatment, in order to be sure of its category, because the margin of error in this type of system must be very small, as those falls, in case of a delayed intervention or an incorrect qualification, may have an adverse impact on the person concerned :

— False alerts (level 0): Empty place.
— False alerts (level 1): Active person.
— Alerts with average level (level 2): Seated person.
— High-level alerts (level 3): Person lying down.

The growing needs of these types of companies (24/7 streaming HD camera, increasing number of clients, unpredictable falls of elderly people) also increase the need to have more dynamic, adaptable, and proactive business processes that ensure an appropriate response to emerging customer events while maintaining an effective management of resources and without compromising one business process value (time, cost, quality, efficiency, flexibility, etc.) over the other. It turns out that time and resources are the most critical values in these cases, and a non efficient management of resources preclude the organization from achieving an effective scheduling, and this consequently hinder the continuous improvement of these business processes.

Video surveillance systems have proven their efficiency, as they can detect hundreds of falls and risky situations and then assist the person in danger in less than five minutes. However, these ambient cameras videos generate a lot of false alerts, triggered by an active person or a moving curtains for example, that are sent to the back-office alert workflow management system. Given that the human agent handles the received alerts in a first-in-first-out (FIFO) order, sometimes true and critical alerts may stay on a waiting list for a few minutes until the agent handles all the false alerts or less critical ones that were received before the true one, as the incoming events are intercepted, queued, and launch the process instances. So, if we do not integrate a mechanism to help human intelligence by prioritizing the event generated by a (very) serious case, high latency will induce delays that can be disruptive, depending on the severity of the case.

3.4 Validation

In this section, we present the results of our experiments to demonstrate the effectiveness of the combination of the two proposed steps. All our experiments were conducted on an Intel(R) Core(TM) i5- 540M 2.53GHz. For the first step in our proposed approach, which aims to estimate the priority level of each business process instance based on the criticality level of each incoming event that launches these instances, we used the K-means algorithm that we coded in the R language. For this, we took a data set of patient falls over the period from 01-02-2016 to 12-06-2017; this data set consisted of 238228 observations generated by 81 patients: 89312 alerts are of level 0 (low), 148466 of level 1(average), 275 of level 2 (serious) and 175 of level 3 (very serious). This data set represents historical data gathered from our previous business process past instances (see Figure 3.5). More details about our data set are represented in the Appendix C. As shown in Figure 3.6, we obtain four clusters with the K-means algorithm based on the score of each patient calculated using the total number of his/her falls, taking into consideration level 2 and 3 only.

Analyzing historical data for each patient helped us to cluster the patients into categories and find similarities between different patients. Each time a new event has been processed, the clustering is dynamically restarted in batch to ensure that the clusters are continuously updated and conclusive regarding the evolution of the patient's health level. In fact, this helps us keep the score and the cluster of each patient up to date in our database. Because we are using these two criteria to estimate the priority of incoming events, in order to execute the instances linked to those events in priority order instead of first in first out order, as shown in Figures 3.7 and 3.8.

The first figure (see Figure 3.7) represents the contents of the JSON file that we send to our REST API (we have used Postman[5] which is an API platform for building and using APIs) to sort the incoming event by priority using the score and the cluster ID of each patient. The second Figure (See Figure 3.8) shows the results received. As we can see, the score obtained for each event corresponds to the result of our clustering, and

5. https://www.postman.com/product/rest-client/

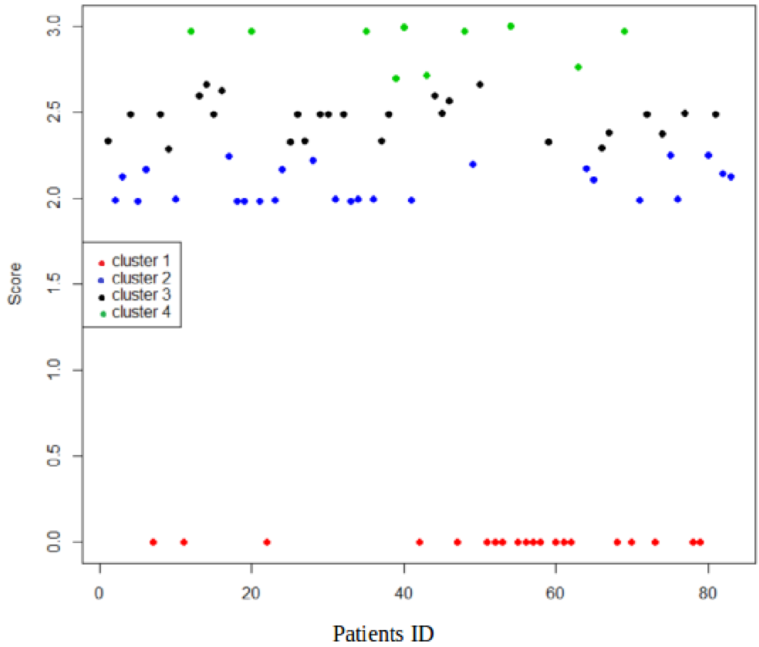

Figure 3.6 – Clustering of patients according to their score

these events are sorted according to the score and the cluster ID of their sources.

In the second step, which represents the human resources allocation step in our approach, the Genetic Algorithm and all other algorithms were coded in Java programming language. To experiment our Genetic Algorithm-based approach for human resources allocation we used 8 human resources (see Table 3.4) with the same sorted events from our first experiment (see Figure 3.8). The results obtained from this matching operation (Resource, Task) respect the two constraints that we propose in our approach, which are the reliability score and the initial availability, which is linked to the time slot of availability for each human resource.

Table 3.4 – List of Human Resources

Human resource ID	Reliability score	Initial availability	Time slot of availability
1	0.13	4 hours	8AM - 12 (Noon)
2	0.20	4 hours	8AM - 12 (Noon)
3	0.25	8 hours	2PM - 8PM
4	0.19	3 hours	2PM - 5PM
5	0.57	4 hours	8AM - 12 (Noon)
6	0.31	4 hours	4AM - 8AM
7	0.12	2 hours	10AM - 12 (Noon)
8	0.43	6 hours	6PM - 12 (Midnight)

We obtain the following result (7, 88876), (1, 88875), (2, 88874), (5, 88873). Among

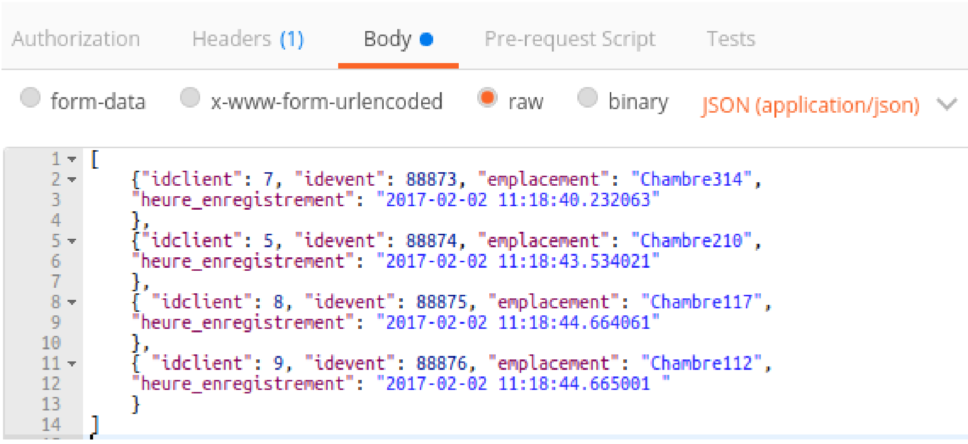

Figure 3.7 – Event priority determination: Received Events


```
[
    {"idclient": 9, "idevent": 88876, "emplacement": "Chambre112",
    "heure_enregistrement": "2017-02-02 11:18:44.665001 ",
    "cluster": 4,
    "score": 3
    },
    { "idclient": 8, "idevent": 88875, "emplacement": "Chambre117",
    "heure_enregistrement": "2017-02-02 11:18:44.664061",
    "cluster": 4,
    "score": 2.7
    },
    {"idclient": 5,  "idevent": 88874, "emplacement": "Chambre210",
    "heure_enregistrement": "2017-02-02 11:18:43.534021",
    "cluster": 2,
    "score": 2.3
    },
    { "idclient": 7, "idevent": 88873, "emplacement": "Chambre314",
    "heure_enregistrement": "2017-02-02 11:18:40.232063",
    "cluster": 2,
    "score": 2.1
    }
]
```


Figure 3.8 – Event priority determination : Sorted Events

the available human resources, only those with high reliability score were selected.

In addition to the constraints related to human resources (availability and reliability) and to business process instances (priority), response time is also an important criterion that we should take into consideration in our approach since we are dealing with a critical tasks that should be allocated to human resources in near real-time. For this, we conduct another series of experiments in which we keep a fixed number of human resources, but we have alternately modified the number of tasks and the number of generation that we used within our genetic algorithm. Figure 3.9 represents the results obtained.

We observe that our priority-based scheduling approach allows us to schedule up to 20 events in just a few seconds. Increasing the number of generations causes a slight increase in processing time, but the final result of resource allocation is the same. Thus, we have limited the number of generations in our genetic algorithm to 50.

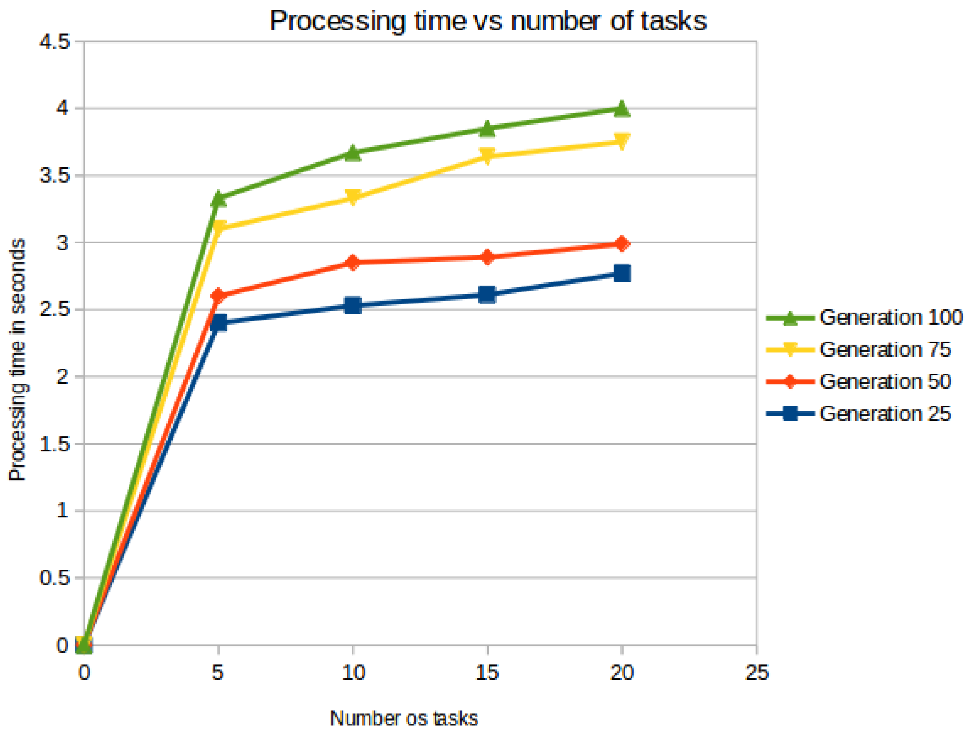

Figure 3.9 – Variability of the processing time according to the number of tasks and the number of generation

3.5 Conclusion

Scheduling and resource management are important steps in the path of business process improvement, as they directly influence business process performance. After reviewing the literature and highlighting its shortcomings, we realized that the existing approaches do not take into consideration information gathered from past process executions, given the stateless aspect of business processes in most cases, and more precisely in incident management business processes. Extracting useful knowledge from these existing event logs can help achieve effective instance scheduling decisions. Besides, the concept of priority has not been sufficiently addressed in the literature, as a scheduling optimization criteria.

In this chapter, we introduced a two-phase approach to ensure effective scheduling in the case of critical tasks that must be executed by human resources. The first phase represents a solution to determine event priority to ensure effective scheduling of instances in the business process. This solution is based on the analysis of historical data from past business process execution using unsupervised machine learning algorithms for clustering, in order to manage the priority of several events that launch business process instances. The second phase is about resource allocation. In fact, the problem of scheduling in business processes has several constraints at the same time, such as resource availability, resource reliability, and time. As this problem is considered an optimization problem, we propose a genetic algorithm to solve it to achieve an effective match between the most critical process instance and the most available human resource.

So far, our contribution ensures that the events that launch business process instances are processed according to their order of priority, by exploiting the result of our clustering step to estimate the criticality of the incoming events. However, there is always an uncertainty level regarding the criticality/priority level of events generated from sources that belong to the same cluster. In the next chapter, we explore how to manage this level of uncertainty.

Chapter 4

Business Process uncertainty mining through Fuzzy Logic

> Knowledge is an unending adventure
> at the edge of uncertainty.
>
> Jacob Bronowski

There is a great excitement that new technologies based on artificial intelligence and machine learning are increasingly used in real world applications : health-care, banking, facilities, weather forecasting, etc. Especially in this era of digital transformation where enterprises are looking, more and more, for advanced and effective technologies that facilitate and improve decision-making process. However, as machine learning algorithms become further integrated into real world systems, these systems have some difficulties caused by uncertainty. Uncertainty can be seen as "a situation which involves unknown or imperfect information", which can lead to a plethora of possibilities. All these possibilities are what make human beings free to choose and make decisions. Besides, humans beings have no problems dealing with uncertainty and understanding words that are related to approximate reasoning in human decision making, such as "Low", "high", "slow", "fast", "about", etc. Humans are able to make decisions in imprecise and uncertain environment, but what about machines and algorithms?

4.1 Why Fuzzy Logic?

To understand the reason behind using fuzzy logic in this contribution, we need first to present a list of terms and definitions that are going to be used throughout this chapter.

Conceptual Background

Fuzzy Logic is a mathematical discipline used to handle nonlinear uncertainties that exist in physical systems to model human experience and human decision making behavior. The emergence of Fuzzy logic is linked to the publication of a paper by Lotfi A.Zadeh in

1965. In this paper Zadeh introduced the concept of a fuzzy set, and latter on the fuzzy logic in the broad sense and fuzzy system modeling (FSM).

The difference between Fuzzy sets and classical sets is the concept of boundaries. In fact, a classical set represents distinct and definite collection of objects, and satisfy two requirements [267] :

— All members of the set must be distinguishable from one another.
— The set must have a sharp boundary, which means that any given object is either a member of this set or not.

Whereas for Fuzzy sets the boundaries are not mandatory sharp, which leads to ambiguity and uncertainty. For example, *"the class of animals clearly includes dogs, horses, birds, etc, as its members, and clearly excludes such objects as rocks, fluids, plants, etc. However, such objects as starfish, bacteria, etc. have an ambiguous status with respect to the class of animals. The same kind of ambiguity arises in the case of a number such as 10 in relation to the "class" of all real numbers which are much greater than 1."* [268].

Fuzzy logic is characterized by fuzzy sets, membership functions, linguistic variable, fuzzy operations, fuzzy If-Then rules, and Fuzzy Inference System. We will see all these concepts in more details in next subsection, but first let us dig deeper into uncertainty since it represents the reason behind using fuzzy logic in our approach.

Uncertainty

Nowadays, all objects are becoming increasingly connected in different fields, which has led to a huge amount of data and events, that are conveyed by all these sensors. The need for technologies and methods, to bridge the gap between collected data and their exploitation, is steadily increasing. That is why, we are witnessing an emergence of "big" data technologies, Machine learning techniques, Artificial Intelligence methods and data science techniques, that are increasingly reaching to every other domain. However, when applying one of these methods, we can be confronted with some uncertainty challenges.

Uncertainty as concept is linked to all the branches of science: in Social Science (Psychology, sociology, Cognitive, etc.), Life Science (Biology, etc), Formal Science (Mathematics, logic, etc.), and Computer Science and data science which draws on Formal Science such as mathematics and logic. Uncertainty concerns process management fields also, due to uncertain event data and event logs. We can distinguish two different types of uncertainty [279]:

— **Strong uncertainty**: the possible values of the attributes are known, but the probability that the attribute takes a certain instantiation is unknown or cannot be observed.
— **Weak uncertainty**: both the possible values of an attribute and their respective probabilities are known.

Several researchers have tried to define uncertainty in more precise way and identify its different types and dimensions. We present in Table 4.1 an non exhaustive list of uncertainty definitions from different domains (mathematics, Information systems, process mining, and psychology...).

With the emergence of Big data, Internet of Things, the need to handle uncertainty is becoming more insistent. In fact, uncertainty exists in every phase of "big" data

Uncertainty definition	**References**
A part of Incompleteness which is product of Errors, it is a specific type of Vagueness	[269] [270]
The basic part of ignorance. Uncertainty is a property that results from a lack of information about the world for deciding if the statement is true or false	[271]
Uncertainty relates to the content of value of information. It is a lack of the truth or the confidence of information	[272]
The lack of available information about the state of the world for determining if a classical statement (which can only be true or false) is actually true or false	[273]
Uncertainty can be identified with three basic types : non specificity, strife and fuzziness	[274]
Uncertainty is to understand the circumstances under which each of the inconsistent propositions can appear to be warranted and yet turn out to be false	[275]
Lack of knowledge, bias and psychological perception	[276] [277] [278]
An explicit absence of trustworthiness	[37] [279]
The psychological perception that creates fear and the motivation of communication	[280] [281]

Table 4.1 – Definitions and types of uncertainty

learning, "big" data analysis, and in machine learning application. This uncertainty can result either from the data collection phase (due to environmental conditions or sampling issues,...), or the learning phase when using [10]:

— **Low value data** : when using data sets unrelated to the current problem for example.
— **Low veracity data** : when using data sets that contains uncertain and/or incomplete data.

As we can see in figure 4.2 and figure 4.1 , to tackle uncertainty and facilitate the process of decision making under uncertainty, fuzzy logic is considered as an approach that can efficiently handle vagueness and imprecision that are difficult to manage and/or to model.

4.2 Proposed Approach

We apply this approach on the same case study we have presented in details in chapter 3. The global business process of this case study is simple but it represents several hard

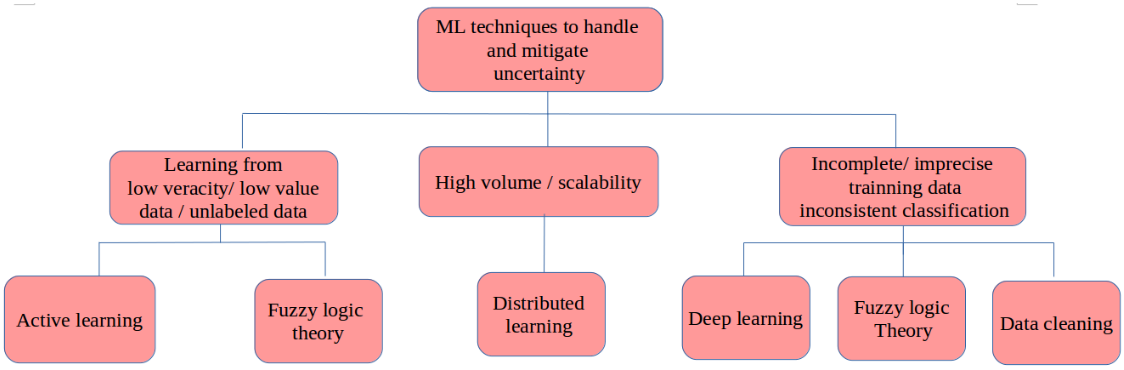

Figure 4.1 – Handling uncertainty with machine learning [10]

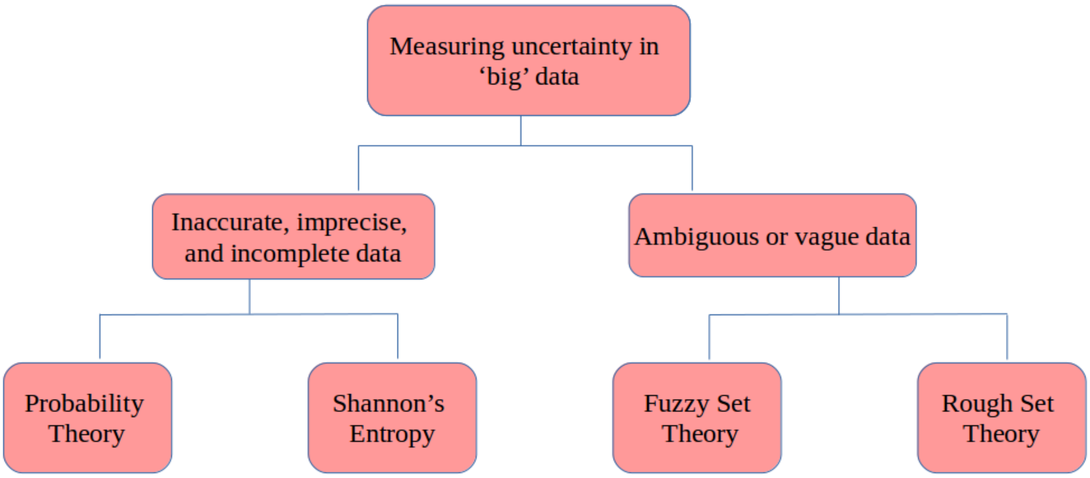

Figure 4.2 – Uncertainty measuring approaches in 'big' data [10]

functional constraints such as: real-time data analysis and the obligation to maintain limited resources for the viability of the business. An instance of this process is handled by one agent (from end-to-end). That is why, we focus on estimating the priority of the whole instance instead of the different tasks of this business process.

This Section provides details about our proposed approach. Our goal is to efficiently mitigate uncertainty and manage the communication between events received from IoT devices and business process instances using Fuzzy Logic, which provides a process for formulating the mapping from inputs to output through a Fuzzy Logic system.

Our augmented BPM through Fuzzy logic in an IoT-BPM architecture is composed of three main layers: (i) The Edge/Fog based IoT Layer which ensures incident data acquisition and filtering. We use a message broker to facilitate the communication and data exchange between the IoT layer (source) and the Fuzzy Inference System (target). Message broker is responsible for routing, storing, retrieving and transforming the information. (ii) The control layer is decomposed into a series of two FIS, where the output

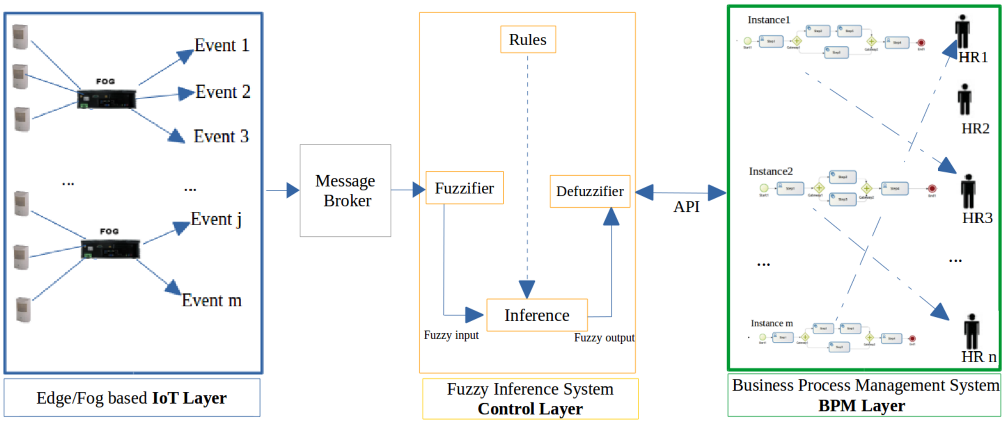

Figure 4.3 – IoT-BPM Architecture for priority-based event management through FIS

of the first FIS is one of the inputs of the second FIS. These two FIS are used here as a decision-making technique to ensure the processing of the IoT generated events, based on several criteria, in order to determine the criticality level of each event source (FIS 1) and then the priority level of those generated events (FIS 2) and their triggered business process instances. (iii) The BPM layer used to coordinate the execution of business process instances and the resource allocation, taking into consideration the priority levels previously determined by our FIS layer. The communication between the control layer and BPM layer is ensured through an application programming interface (API) (see Figure. 4.3). In this approach, we focus only on the implementation of the control layer, as it represents the core phase of our architecture.

4.2.1 Definitions related to uncertainty mining/Fuzzy logics

As we saw in the previous sections, Zadeh proposed Fuzzy logic as a methodology to model human experience and human behavior in decision-making process, in order to represent uncertainty, imprecision and provide a formal approach to deal with the intrinsic inaccuracy of several problems. This concept is characterized by [282]:

Fuzzy Sets: represent objects collection, which is characterized by a membership function. The membership function is a type variable, which uses words instead of numbers to represent its knowledge in fuzzy systems that express the statement of conditional rules.

Linguistic variables: are parameters that allow the representation of numerical values with fuzzy sets. They are used to control the system and to show its performance. They are divided into levels.

Fuzzy Operations: are the same operations in classical sets, which are union, intersection, containment, and complement.

Fuzzy If-Then rules: Fuzzy rules consist of three parts antecedent, fuzzy proposition, and consequence. The fuzzy if-then rule expressed in the following statement:

if x_1 is A and/or x_2 is B then x_3=C. where, A, B, C are linguistic values, and x_1,x_2,x_3 are linguistic variables.

Fuzzy Inference System (FIS): is the regrouping of fuzzy rules, the linguistic variables of membership functions and fuzzy reasoning. The FIS is designed for constructing complex and nonlinear relationships between the input and output. There are many FIS used in applications such as Mamdani, Tsukamoto, and Sugeno [283]. The difference between these types of FIS is in the aggregation and defuzzification processes.

According to [284] , the fuzzy processing procedure is realized in four important steps, including Fuzzification, Fuzzy If-Then rule, Aggregation, and Defuzzification. This process is represented as follows:

— Define the input (Problem identification) and the output (Interpretation and verification) of the system.
— Select the type of input and output membership functions (MFs).
— Convert the input numerical values into linguistic variables (levels) using MFs, this step called Fuzzification.
— Define the IF-THEN rules and apply them to the fuzzy input parameters.
— Use Defuzzification to convert fuzzy values to numerical values as the output.

Fuzzification: is the first step for fuzzy modeling. It represents the process of transforming a crisp or real value into fuzzy sets by using membership functions. In this step, the crisp input values are transformed into linguistic variables.

Fuzzy If-Then rule: Fuzzy rules consist of three parts, antecedent, fuzzy operation, and consequence. For example: if x1 is A and/or x2 is B then x3=C. where, A, B, C are linguistic values, and x1, x2, x3 are linguistic variables.

Aggregation: is the combination of the output of every rule into one fuzzy set before the defuzzification step.

Defuzzification: is the process of converting a fuzzy output of a FIS into a crisp output.

4.2.2 Formulation of BP uncertainty problem

Our fuzzy system is divided into two Fuzzy Logic Systems, the first one is for the Patient Criticality Level determination and the second one is for the Event Priority Level determination. Each system consists of four important steps including Fuzzification, If-Then rules, Aggregation, and Defuzzification. Figure 4.4 describes different blocks diagram of FIS for Patient Criticality Level and Event Priority Level determination respectively. More details about these two Fuzzy Inference systems are presented in the following sub-sections.

4.2.3 Business process instance criticality fuzzy selection

Inputs and output of FIS 1

For modeling the control system, the first FIS (FIS 1) has two fuzzy parameters as inputs: Patient Status and Device Reliability Level.

+ The **Patient Status** parameter represents all the characteristics of a specific patient. It contains:

— **Room location** of the patient.

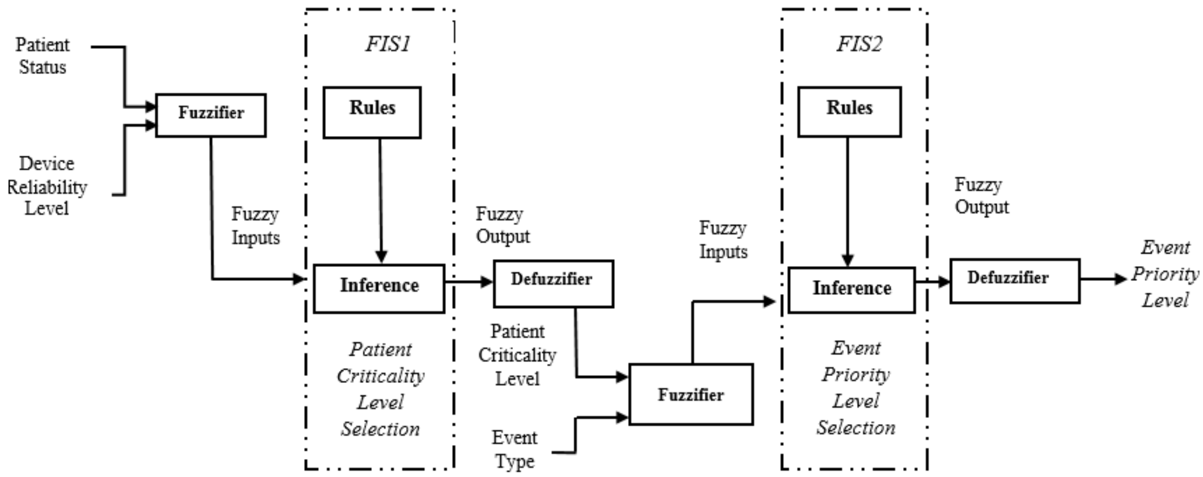

Figure 4.4 – Fuzzy Inference System (FIS) for Criticality Patient Level and Event Priority Level determination

- **Event Id:** represents the identification of each event.
- **Emergency level of the previous fall:** presents the level of emergency of the previous fall.
- **Duration of the previous fall:** represents the number of days between the current and the previous critical fall.
- **Repetitive faller:** represents if the patient is considered as repetitive faller or not. Several studies have been conducted in the field of silver economy, in order to determine a standard definition of a fall and the number of falls over a specific period to consider a person as a repetitive case. In [262] [263] [264] [265] [261] a fall represents "an unintentional change in position resulting in coming to rest at a lower level or on the ground". To characterize the repetitive aspect of a fall, we must determine the number of falls and the time interval between falls. As we have seen in Table 3.3, in chapter 3, the majority of published studies consider at least two falls to retain repetitive character, with an interval between two falls ranging from 6 to 12 months on average.
- **Patient Score:** represents a weighted mean calculated for each patient based on his/her several previous falls of levels 2 and 3 only.

+ **The Device Reliability Level** parameter, presents all features of the device: its identification, location, date of the first use, a total of false alerts ranged from 2015 to 2018 and whether the device has been changed or not since its first use. This reliability level is determined based on the state of each device. In fact, The state of the device deteriorates with time, so the device can either generate a lot of false alerts or not detect the real ones.

The FIS 1 is characterized also by one output parameter, which is the chance value of Patient Criticality Level. The Patient Criticality Level output refers to the criticality level of each patient that generates an event at a time slot t.

The numerical values of these inputs are converted into linguistic variables using membership functions during Fuzzification step.

Membership Functions

The fuzzy parameter that describes the Patient Status input is illustrated in figure 4.5. The linguistic values for this fuzzy set are Low, Average, Serious and Very Serious.

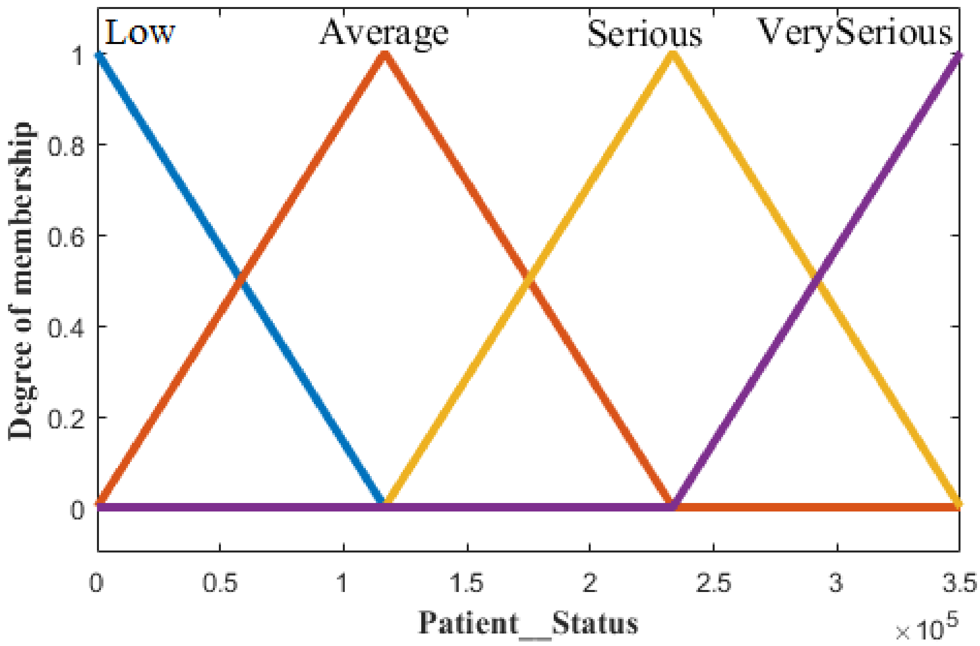

Figure 4.5 – Membership functions for patient status input

The second fuzzy input parameter is the Device Reliability level. The fuzzy parameter that describes this input is illustrated in Figure 4.6. The linguistic values for this linguistic variable are Low, Medium, High and Very High.

The output of the first FIS is Patient Criticality Level. The fuzzy parameter that illustrates this output is described in Figure 4.7 This output is divided into four membership functions as follows: Low, Average, Serious and Very Serious.

The memberships functions of all inputs and output are represented by triangular membership functions because this type is most frequently used and give better results [285]. For the defuzzification process, we have used here the method of the Center of Area (CoA) [286], called also the Center of Gravity (CoG) method. The concept of this method is that the fuzzy controller calculates the range of the output variable and within the area under the scaled membership functions as described in equation (4.1).

$$\text{CoA} = \frac{\int_a^b \mu_A(x) x \, dx}{\int_a^b \mu_A(x) \, dx} \tag{4.1}$$

Where, x is the value of the linguistic variable, and a and b represent the range of the linguistic variable.

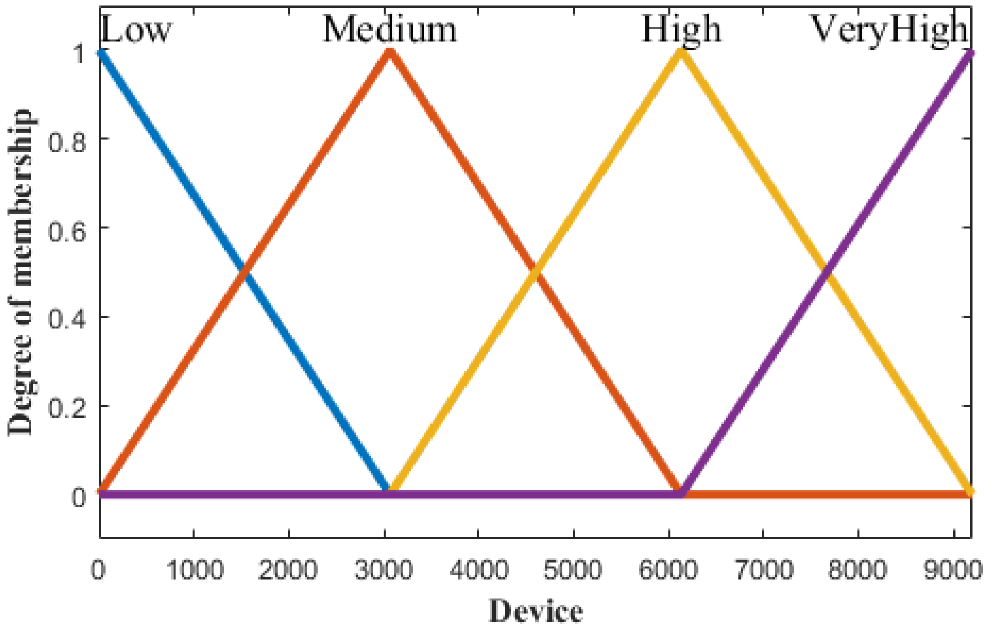

Figure 4.6 – Membership functions for device reliability level input

Fuzzy If-Then rules

The chance values (Patient Criticality Level) is accomplished by using predefined fuzzy if-then rules to handle the uncertainty. We have two inputs, each divided into four linguistic variables, thus we obtain 2^4=16 possible chance values as shown in table 4.2. The defuzzification step uses these values to obtain crisp output values.

4.2.4 Business process instance priority fuzzy inference

The second FIS modeling follows the same steps of FIS 1modeling as discussed above: Fuzzification, Fuzzy rules, and Defuzzification.

Inputs and output of FIS 2

The second FIS has two fuzzy input parameters: Patient Criticality Level, which is the output of the first FIS and Event Type, and one output which is the Event Priority Level:

— Event Type input represents the type of the last event generated by the device in question. These events could either be false alerts or true alerts.

— Event Priority Level output is used to choose between two or more events that will trigger two or more business process instances. Those instances may need the same resources at the same time (in case of limited (human) resources). The event with the lowest priority must wait for the resource occupied by executing the event with the highest priority.

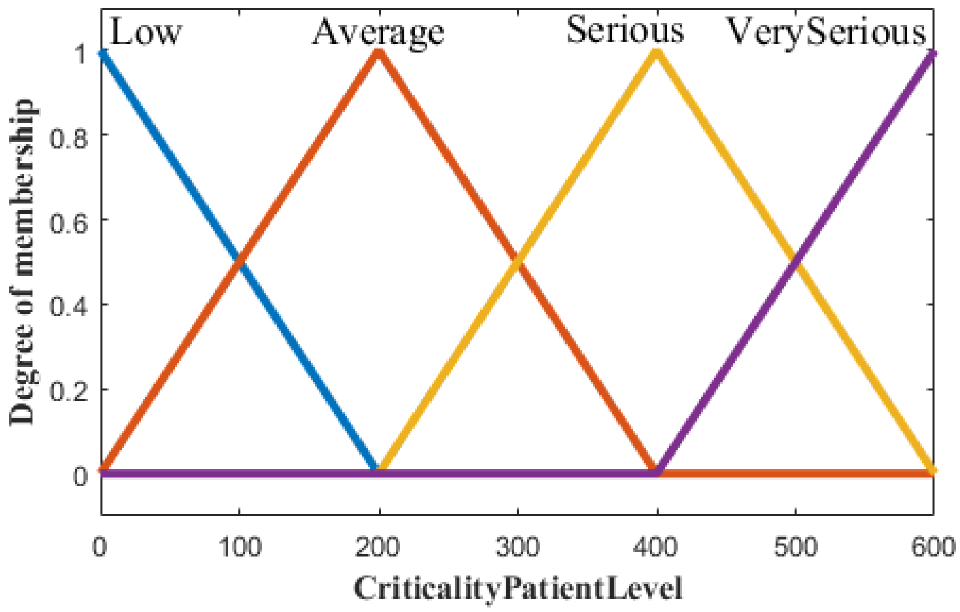

Figure 4.7 – Membership functions for Patient Criticality Level output

Table 4.2 – FIS1 If-Then rules for Criticality Patient Level Selection

Rules Number	Patient Status	Device Reliability level	Criticality Patient Level
1	Low	Low	Low
2	Average	Low	Low
3	Serious	Low	Average
4	VerySerious	Low	Serious
5	Low	Medium	Low
6	Average	Medium	Average
7	Serious	Medium	Serious
8	VerySerious	Medium	VerySerious
9	Low	High	Low
10	Average	High	Average
11	Serious	High	Serious
12	VerySerious	High	VerySerious
13	Low	VeryHigh	Low
14	Average	VeryHigh	Average
15	Serious	VeryHigh	Serious
16	VerySerious	VeryHigh	VerySerious

Membership Functions

The fuzzy parameter that describes the Patient Criticality Level input is illustrated in Figure 4.8. The linguistic values for this fuzzy set are Low, Average, serious and Very serious.

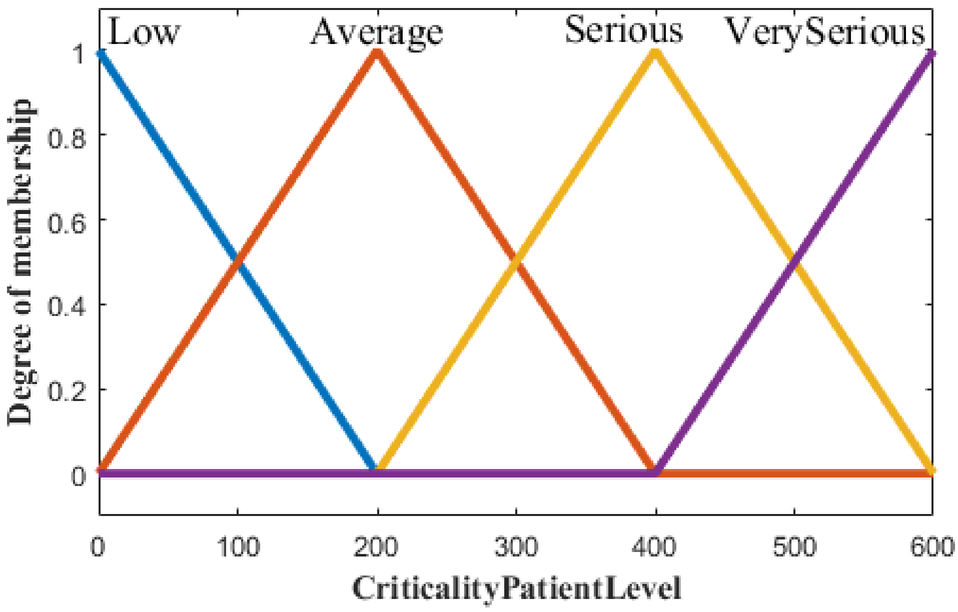

Figure 4.8 – Membership functions for Patient Criticality Level input

The second fuzzy input parameter is the Event Type. The fuzzy parameter that describes this input is described in Figure 4.9. The linguistic values for this linguistic variable are True and False.

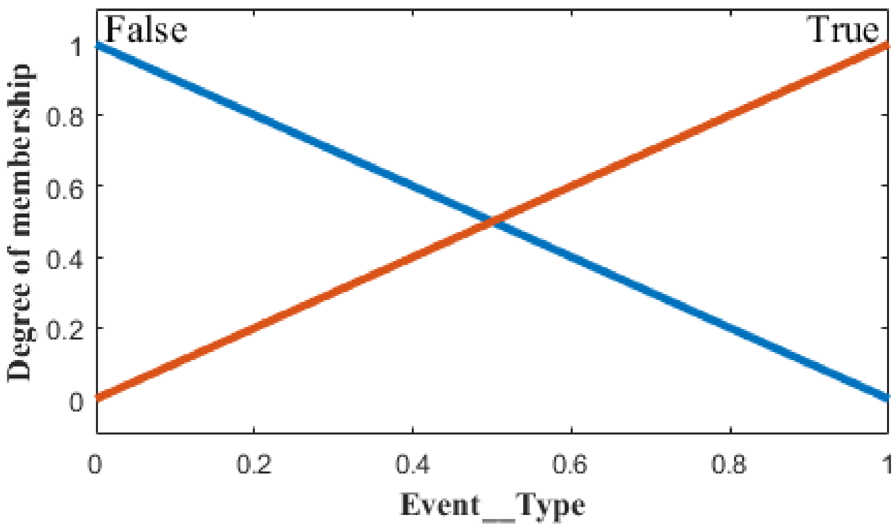

Figure 4.9 – Membership functions for Event Type input

The output of the second FIS is Event Priority Level. The fuzzy parameter that illustrates this output is represented in Figure 4.10.

This output is divided into eight membership functions as follows: Very Low (VL), Low (L), Little Average (LA), High Average (HA), Serious (S), High Serious (HS) and

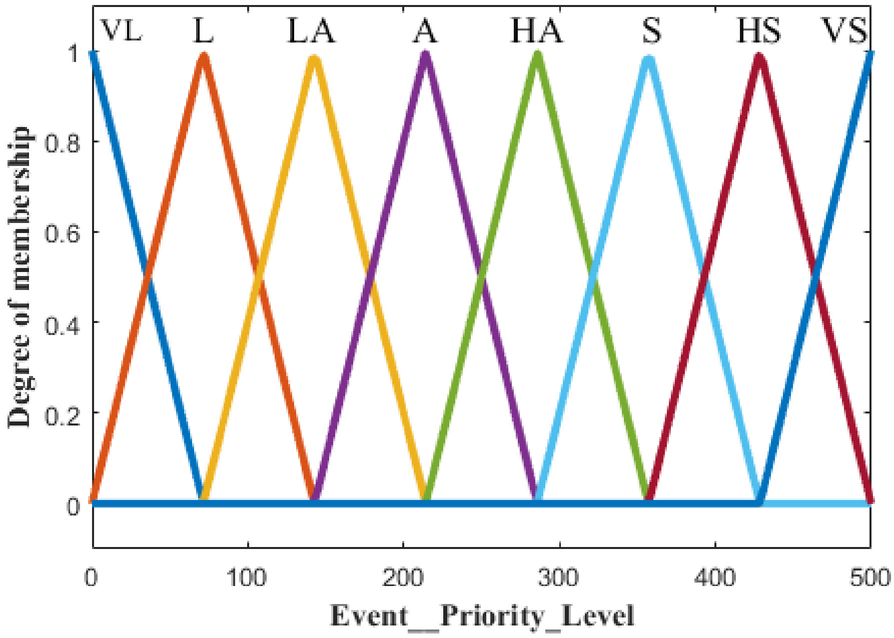

Figure 4.10 – Membership functions for Event Priority Level output

Very Serious (VS).

Fuzzy If-Then rules

According to the number of inputs and the number of linguistic variables for each input (2x4) 8 rules are determined. The chance values (Event Priority Level) are accomplished by using predefined fuzzy If-Then rules to handle the uncertainty as shown in table 4.3.

Table 4.3 – FIS2 If-Then rules for Event Priority Level Selection

Rules Number	Criticality Patient Level	Event Type	Event Priority Level
1	Low	False	VeryLow
2	Low	True	Low
3	Average	False	LittleAverage
4	Average	True	HighAverage
5	Serious	False	Average
6	Serious	True	HighSerious
7	VerySerious	False	Serious
8	VerySerious	True	VerySerious

4.3 Validation

In this section, we present the results from our experiments, in order to demonstrate the effectiveness of our fuzzy logic based approach. In order to be able to compare the two approaches we proposed, we have used the same case study with the same process and the same data set.

4.3.1 Simulation settings

The performance of our proposed approach is evaluated using MATLAB. The fuzzy logic toolbox allows users to create Fuzzy inference for estimating conclusions problems. The simulation of our system is built through the Graphical User Interface (GUI) Tools using five GUI tools that are being executed to have simulations for the input and output: building, editing FIS, Membership function, Rules, and reviewing Rule and surface.

In order to achieve our goal which is the estimation of business process instances priority through the estimation of the criticality level of the events that trigger those instances, we used two Fuzzy Inference Systems. The first FIS is dedicated to determining the Patient Criticality Level, and the second FIS is dedicated to determining the event priority level based on the result of the first FIS. We worked with two datasets as input for the Fuzzy Inference System: the first dataset is an event log from our business process, the second dataset contains IoT-device characteristics.

4.3.2 Simulation Scenario Description

To evaluate our proposed approach, we analyzed our historical data of each event source (IoT device in each patient room) and discuss several cases to estimate the priority level of each received event. To this end, we choose eleven different patients with different IoT device in eleven different rooms. We report four cases of simulation to observe the Event Priority Level in different situations (Low, Average, Serious and Very serious):

Case 1: Low patient status (i.e. a patient considered as a non-critical case) and Low device reliability level.

Case 2: Very Serious patient status (i.e. a patient considered as a very critical case) and Very High device reliability level.

Case 3: Low Patient Criticality Level and False event type (i.e. if the event was generated by a device with Low reliability level).

Case 4: Very Serious Patient Criticality Level and true event type (i.e. if the event was generated by a device with High reliability level).

4.3.3 Analysis and discussion

Case 1: If (Patient status is Low) and (device reliability level is Low) then the Patient Criticality Level is Low. For example, if the value of the patient status is low (138000), and the device reliability level is low (297), then the Patient Criticality Level is 102 (See Figure. 4.11). In this case, the Patient Criticality Level is also Low, it can be explained that the patient status is non-critical.

Figure 4.11 – Rule viewer for Case 1

Case 2: If (Patient status is Very Serious) and (device reliability level is Very High) then Patient Criticality Level is Very Serious. For example, if the value of the patient status is Very Serious (319000), and the device reliability level is Very High (8810), then the Patient Criticality Level is 458. In this case, the Patient Criticality Level is Very Serious, which signifies that the patient status is more critical (See Figure. 4.12).

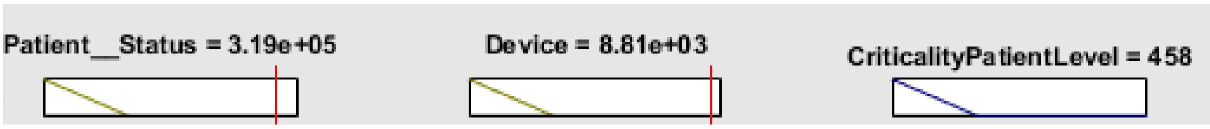

Figure 4.12 – Rule viewer for Case 2

Case 3: If (Patient Criticality Level is Low) and (event type is False) then the Event Priority Level is low. For example, if the value of the Patient Criticality Level is Low (29,8), and the event type is True (0.123), then the priority level is 116. In this case, the priority level is Low, which means that the patient status is non-critical and has the lower priority level (See Figure. 4.13).

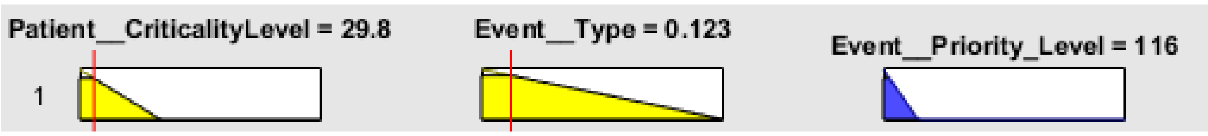

Figure 4.13 – Rule viewer for Case 3

Case 4: If (Patient Criticality Level is Very Serious) and (event type is True) then Event Priority Level is Very Serious. For example, if the value of the Patient Criticality Level is Very Serious (456), and the Event Type is True (0.959), then the priority level is 423. In this case, the priority level is Very Serious, which means that this patient status is more critical and has the highest priority level (See Figure. 4.14).

As we can see in Figure 4.15, the patients that we have randomly chosen for this simulation (represented in the figure by the ID of the events generated by their devices) have almost the same score value (2.5), except the patient with event Id (313671) which is considered as a less critical patient (belongs to cluster 2 which is given here for comparison purposes). However, those ten patients do not have the same criticality level. So, it is important to efficiently manage the criticality levels of each patient even when the device reliability level is low (until this device is changed). As we can see the criticality level values fall into the range of [65, 500]. We divide the range into 3 parts. The first part is when the criticality is greater than 400. It corresponds to the highest criticality level. In

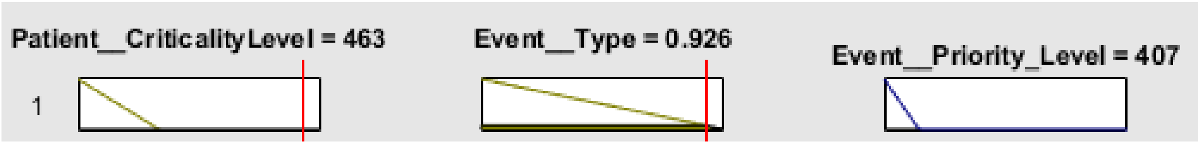

Figure 4.14 – Rule viewer for Case 4

this situation, the events are marked with a higher priority level. Then, if the criticality is less than 100. we say that it is at the lower criticality level. Finally, medium criticality level corresponds to criticality values, which is between [100, 400].

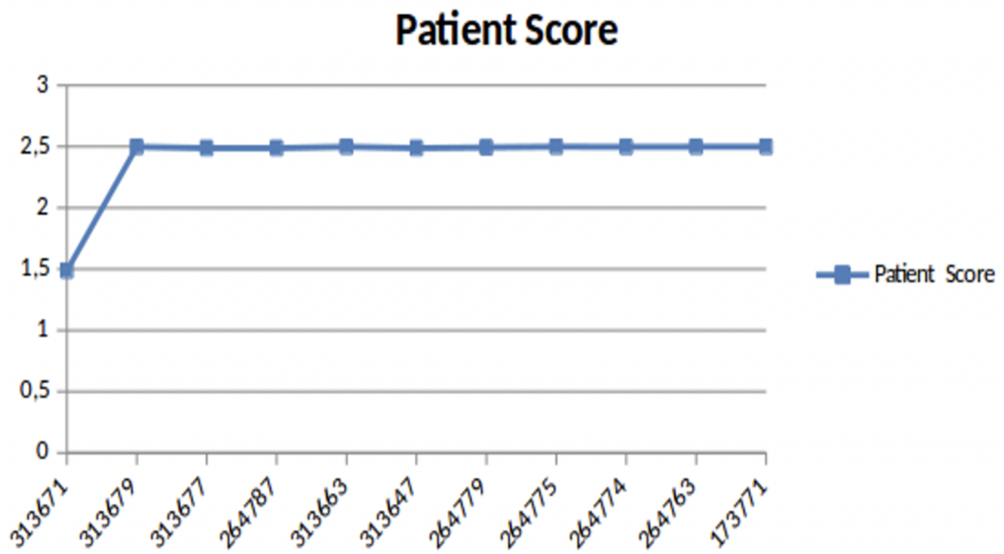

Figure 4.15 – Patients Score.

There are 3 important points to consider in Figures (4.15, 4.16). The first point is that the score all alone is not enough to determine the criticality/priority level of an event, due to the classification uncertainty as we have discussed in the previous chapter (See Figure (4.15)). For example, the event Id (264787) has the same patient score compared to other events (See Table 4.4), but it has the lower level of criticality and lower event priority. The second point is that the events generated by patients with the highest criticality level, have the highest priority level, as we can see for the event Id (313679). The third point is that the events generated by patients with close values of criticality level have finally different priority level. If we take, for example, the events with Id (264779) and (173771) they are both generated from two different patients with approximately the same criticality level (64,81) and (64,69) respectively. However, they have different priority levels (162,45) and (93,81) respectively.

In our previous work [29] [57], that we have explained in more details in chapter 3, the criticality level was estimated using the score value to create clusters of priority (with K-Means algorithm for clustering). As explained before, we have obtained four clusters with the K-means algorithm based on the score of each patient calculated using the total

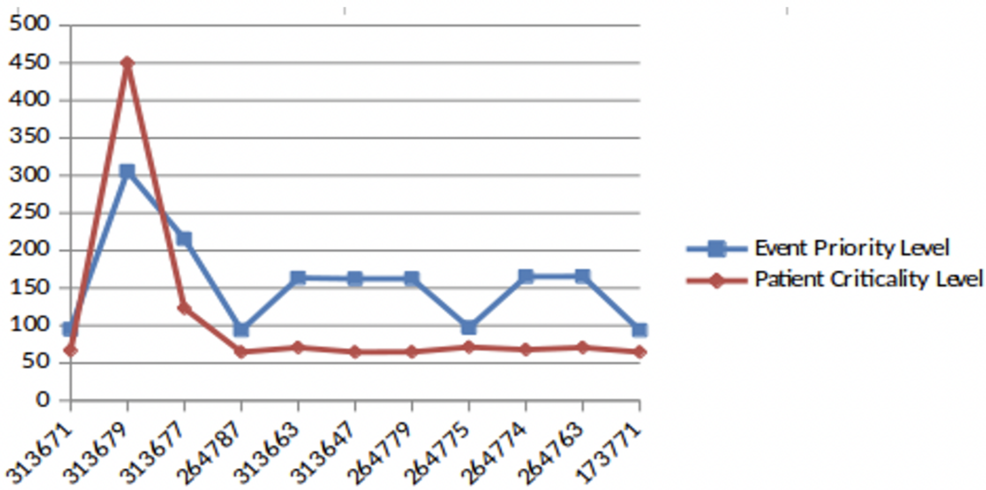

Figure 4.16 – Event Priority Level vs Patient Criticality Level.

number of his/her falls, taking into consideration level 2 and 3 only. So the patients in the same cluster have the same profile and then the same criticality/priority level.

The Event Priority Level is estimated based on the criticality level of the patient (source) that generates this event. With the clustering-based approach, there is a bijection between criticality levels and priority levels. As described in chapter 3, two scenarios were encountered when applying this approach: (i) the patients belong to different clusters. (ii) both patients belong to the same cluster. The subtle difference between the criticality level in the second scenario generates uncertainty regarding their priority level determination (See table 4.4). As we can see in this table, the patients belong to the same cluster (Cluster 3) in spite of their different scores, and in this case their generated events will have eventually the same priority level. However, with our fuzzy logic based approach, we have managed to mitigate this level of uncertainty (See Figure 4.16).

When using fuzzy logic, the decision-making process for determining criticality/priority level is based on multicriteria (Status patient, events type, etc), which are represented by linguistic variables and is based also on uncertainty management problem in the case of the events that have the same characteristics (score). The whole decision-making process in Fuzzy Logic is inspired by human reasoning. While the k-means method is based on the partition of n observations into k cluster for making decision. This algorithm works by choosing k initial cluster centers and then assigns every data point to the nearest cluster (based on distance) based on the provided features that define the resulting clusters (based on feature similarity), see in table 4.5.

The simulation results prove that the integration of Fuzzy Logic in the IoT-BPM architecture has been advantageous for the priority-based event management. From the results above, we can consider that the performance of the proposed architecture, at this step, has been achieved. The highest priority level is assigned to the highest criticality level (i.e a critical patient). Thus, it has a positive impact on managing these events and

Table 4.4 – Event Id and Clusters of patient with the same profile based on their score value

Event Id	Patient Id	Patient Score	Patient Criticality level	Event Priority level	Patient Cluster
313671	70	1.4875	66.66	94.86	Cluster 2
313679	49	2.4979	449.90	305.71	Cluster 3
313677	31	2.4875	123.02	215.41	Cluster 3
264787	21	2.4875	64.69	93.81	Cluster 3
313663	27	2.4987	70.70	163.36	Cluster 3
313647	29	2.4875	64.69	162.37	Cluster 3
264779	30	2.4937	64.81	162.45	Cluster 3
264775	19	2.4995	71.28	97.16	Cluster 3
264774	16	2.4975	67.88	165.19	Cluster 3
264763	12	2.4986	70.70	165.52	Cluster 3
173771	22	2.4989	64.69	93.81	Cluster 3

on reducing the waiting time of some critical cases, especially when the organization has a limited number of human resources.

Both methods have advantages and disadvantages. In fact, K-means is considered as an easy to implement algorithm that gives easy to interpret results. However, it can be difficult to predict the K value; besides, the initial inputs (k, features, order of the data, data set quality, tidiness, etc.) have a huge impact on the final clustering results. The same thing applies to fuzzy logic. In fact, it is characterized by its flexibility. However, in a high-complex system, the use of fuzzy logic becomes an obstacle to verify the reliability of the system.

4.3.4 Complexity analysis

The computational complexity of the fuzzy logic method is presented as follows: O(N_{rule} N_{dim}), where N_{rule} is the number of fuzzy rules and N_{dim} is the number of dimensions of the input [287]. In our case, FIS1 uses 2 inputs and 16 rules; the complexity is on the order of 32. In addition, FIS2 uses 2 inputs and 8 rules, the complexity is of the order of 16. Thus, the computational complexity of the Fuzzy Inference System is the sum of the complexity of FIS1 and FIS2 and is on the order of 48, the running time of the algorithm for n iterations is O(n). For the K-means method, computational complexity is expressed as O (n^2), where n is the input data size. Therefore, its performance is directly proportional to the square of the size of the data set used as input. Generally, K-means performance deteriorates when using very large data sets. With this computational complexity, K-means can be seen as a greedy algorithm [288]. Improving the computational complexity of the K-means algorithms is basically linked to the initial inputs (selecting of better initial centroids, choosing an effective number of clusters, etc.),

Table 4.5 – Comparison: Fuzzy Logic Vs K-Means

Fuzzy Logic	K-Means
Uncertainty-based decisions	Clustering-based decisions
Human reasoning: inspired by the processes of human perception and cognition	Based on distance (Euclidean distance or Manhattan distance): to identify the set of objects with similar characteristics
Linguistic Variables	Features
Building decisions using multicriteria	Building decisions through partitioning n observations into k clusters

and to the execution conditions (using parallel and distributed execution environment, reducing the number of scans over the data set, especially for very large data sets, etc.). As we can see, the fuzzy logic method offers a lower computational complexity in our case compared to the K-means algorithms.

4.4 Conclusion

The Internet of Things (IoT) world is growing at a breathtaking pace. This new paradigm shift affects all layers of enterprise architecture, from infrastructure to business. Organizations today face new challenges to maintain their quality of service and competitive advantage over other rival organizations. Business Process Management (BPM) is a field that, among others, will be affected by this new technology. Both the IoT and the BPM communicate through events, and effective and efficient management of those events ensures a better communication channel between the IoT physical layer and the Business layer. However, the huge amount of those IoT generated events and sometimes the subtle difference between their criticality level generate uncertainty regarding their priority level determination.

In this chapter, we presented an IoT-BPM architecture to ensure effective communication between the IoT layer and the BPM layer through event management by handling uncertainty using a Fuzzy Inference System. This approach is based on the integration of a fuzzy inference system to determine the criticality level of the events generated by IoT on the one hand (FIS 1), and the priority level of the instances of the business process triggered by those events on the other hand (FIS 2). The main advantage of this approach is that the Fuzzy logic allows us to handle the degree of uncertainty regarding the criticality/priority level of the events generated by some sources that may have the same characteristics. This proposed approach can be applied for (IoT) generated events triggering any other entity not only business processes, for example, telecommunication application for management services, or any system that deals with concurrent access to shared resources (human and/or machine) where priority/criticality level determination is important for the viability of the business.

Chapter 5

IoT - BPM Integration Architecture Towards an Augmented BPM

> Always look back and learn, look forward and walk and keep learning.
>
> Samuel Butler

As we have seen throughout the chapters of this thesis, BPM is a well-established discipline in both academia and industry. BPM already provides different methods and solutions to manage and analyze data and event data. Among these methods, we have Business Process Intelligence (BPI). In fact, BPI systems provide solutions to enhance decision making throughout a wide range of business activities, by analyzing, predicting, monitoring, controlling, and optimizing business processes [289]. Although BPI has proven to be efficient for post-execution prediction of future process behavior, it is unable to manage and process large amounts of real-time data and events generated from different sources [290]. This becomes more difficult when IoT devices are integrated into a BPM architecture. Another solution provided by the BPM field when dealing with real-time event data is Business Activity Monitoring (BAM). In fact, BAM is used to analyze data related to activities that have been executed. It complements ex-post analysis of process execution by continuously identifying specific situations at run-time and responding to them by triggering specific actions [291]. However, this technology remains less effective in use cases that include IoT-generated events. The limitations of traditional BAM in IoT case studies can be seen from two aspects: prediction and proactivity. The first aspect is manifested in the identification of complex event correlation [292] . In fact, by sensing their environment, IoT devices generate a massive volume of event data that need to be processed and analyzed to extract useful information and detect (complex) event patterns in real time. However, traditional BAM does not provide rule-based engines. This limitation becomes more apparent when events are generated from diverse data sources, because BAM lacks the flexibility to integrate multiple heterogeneous data sources [290] [293] [294]. The second aspect is related to the absence of proactivity in BAM solutions. In fact, using BAM in reactive way is no longer sufficient, especially when we have this huge amount of real life data and events.

In a business environment where every single event is important and need to be processed, Event Driven Architecture (EDA) needs to be adopted. Now with the emergence of IoT, events are becoming increasingly important for current information systems (SI), especially for organizations that integrate IoT devices and sensors in their business operations (video surveillance, health care, etc.). EDA is the successor of Service-Oriented Architecture (SOA). The idea behind this paradigm is that everything is an event, and all the different components of this architecture interact with each other through events. In this architecture, each component is either an event consumer or an event provider. Event consumers subscribe to an intermediate event handler, and providers publish to that handler. When the event handler receives an event from a provider, the handler forwards it to the consumer [295]. The difference between EDA and SOA is that SOA is based on the *"request/response"* concept where the consumer of the service sends a request to the producer and the producer sends a response that contains either the result or feedback. While EDA is based on the *"publish/subscribe"* concept where the communication pattern between the consumer and the provider is reversed. Therefore, in this architecture, consumers do not start the communication channel, but receive events published by event providers, which means that communication is done in a unidirectional way [296].

The main interest of the EDA is to manage events and data of a real-time process in an efficient way. This message-driven architecture enables the introduction of a higher level of event processing using the Complex Event Processing (CEP) engine. CEP is used to exploit and correlate large event streams generated by heterogeneous data sources in order to produce useful information.

5.1 Augmented Event-Driven Business Process Management

Although the Internet of Things is becoming the hot spot area for technological innovations and economic development promises for many industries and services, it is still in its infant stage. During the last years, both academic and industrial worlds have been interested in this field and its integration with other domains. However, there is still a lot of research work to do to propose methodologies, design models, and architectures in order to ensure an efficient and smooth integration and bidirectional communication between the IoT field and other fields. BPM is a field, among others, that started to be affected by this new technology. However, dealing with events in the BPM field is not recent. The concept of Event-Driven Business Process Management (EDBPM) represents an enhancement of Business Process Management, by including other concepts such as SOA (Service Oriented Architecture), EDA (Event-Driven Architecture), SaaS (Software as a Service), BAM (Business Activity Monitoring), and CEP (Complex Event Processing) [297]. This concept was first used in 2003 in a white paper by Bruce Silver Associates in connection with the FileNet P8-BPM platform [298]. The idea behind EDBPM was limited to a single event processing because the concept of CEP was not well known back then. So, early applications of EDBPM focused mainly on business process monitoring and Key Performance Indicators (KPIs), or metrics measurement [299]. Despite the fact that the concept of CEP was used for the first time in 2002 by David Luckham in his book "the power of events" [300]. It was until 2007 that an integration of CEP into the BPM field was considered for the first time [301]. The integration of CEP technology into BPM is then known as EDBPM [302] [292].

Recently, the EDBPM research area has been growing significantly again due to the omnipresence of IoT devices. We find applications that integrate IoT, CEP, and (ED)BPM in several research areas and business sectors such as healthcare, logistics, manufacturing, banking, smart cities/homes, cultural heritage, agriculture, etc. ([303], [304], [305], [306]).

5.2 Why Complex Event Processing - CEP?

To understand the reason behind using Complex Event Processing in this contribution, we need first to present a list of terms and definitions that are going to be used throughout this chapter.

5.2.1 Conceptual Background

Event

An event, also called *atomic event* is an instantaneous record of an activity in a system [307] at a point in time, and represents any change that occurs or will occur in this system. Whereas a complex event is a set of events that are related to each other by event operators such as aggregation, causality, semantic or time [308]. The field of CEP aims to address several issues related to events such as event filtering, routing, transformation, and also detecting complex events while processing atomic events and using predefined event patterns [309].

Complex Event Processing

Complex event processing is a widely used technology; it becomes an increasingly active research field [310] especially with the popularity of *"publish/subscribe"* systems in 1997, as we have seen in the beginning of this chapter [311]. Complex Event Processing represents a set of methods, techniques, and tools to process events as they occur. The objective of this concept is to process and analyze events generated from different sources in order to extract useful information [312] [313]. CEP is widely used to detect and deal with different business anomalies, threats, and opportunities [312] by analyzing event streams instead of traditional static data stored in databases. CEP engines provide the scalability and (near) real-time processing to filter, combine, and extract actionable knowledge, known as situation of interest, from a stream of events (see Figure 5.1).

CEP Engine and Rules Engine

In an event processing context, CEP engines are used to filter events, aggregate them, and drive valuable knowledge from them.

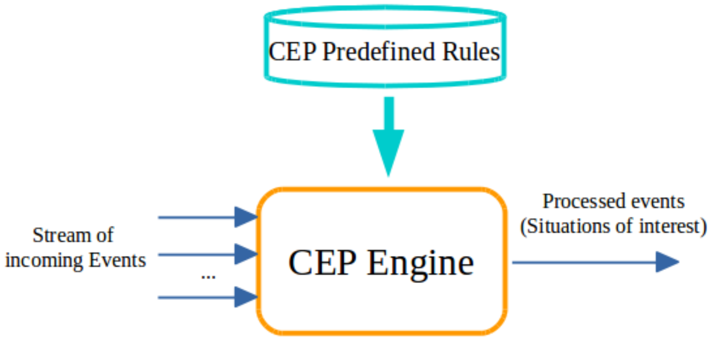

Figure 5.1 – CEP basic architecture

As we have seen before, CEP is based on the *publish/subscribe* concept. *Publish/subscribe* systems processes inbound events without taking into account the history of events already received. However, CEP systems are used to overcome this limitation. In fact, CEP systems are considered as a potent extension of traditional *publish/subscribe* systems [26]. A complex situation of interest can be easily expressed using CEP engines and rules. For example, in our case study, queries such as: *prioritize a case if the patient has some particular needs (Wheelchair, walker, etc)* or *trigger an alert if the patient is a recidivist faller*.

The strength of CEP is basically based on the concept of rules and operators. CEP engines are mainly based on a set of rules provided by a rule engine. A rule engine represents a part of a CEP engine that generates rule models. These rules are used to create and/or modify business logic in a Business Process Management System (BPMS) [314] for example. CEP rules are based on CEP operators. Among these operators we have: *aggregation operators, sequencing operators, logic operators, single-item operators, windowing operators, and flow management operators*. A combination of two types of languages allows to express CEP rules [26]:

— **Declarative Languages:** this type of language is a derivative of languages such as SQL and relational algebra [315]. This extension is created by adding some ad-hoc operators in order to support stream-specific operations. For example, sequencing and time window constraints. Declarative languages are very useful for users who want to express exactly their specifications (what they want).

— **Pattern-based Languages:** this type of language is used to express the conditions of rules using complex patterns. These patterns could integrate operators such as logical, timing, and sequencing operators.

Most of the time, these rules are manually predefined by domain experts, and then in in CEP systems such as Esper [6], Siddhi [7], FlinkCEP [8], or Oracle. Since defining these rules manually can be error prone and time consuming, there are many recent approaches that propose an automatic CEP rule learning and generation [312] [316] [317] [318] [26]. However, automatic CEP rules learning and generation is beyond the scope of our work in this thesis.

6. https://www.espertech.com/esper/
7. http://siddhi.sourceforge.net/
8. https://nightlies.apache.org/flink/flink-docs-release-1.14/docs/libs/cep/

A CEP engine provides several functionalities:
— Event-pattern detection
— Event filtering
— Event-relationships detection
— Event aggregation and transformation
— Modeling event hierarchies
We have based our approach on event pattern detection. From our stream of received events, we must trigger an alert if an event has the characteristics of a serious incident or anomaly.

5.3 Proposed Approach

Event-driven business process management is adapted mainly in organizations that have a real-time based activities that involve some sensors or IoT devices that collect data and generate new events by sensing their environment [57]. However, a real-time system must have three main characteristics to ensure better functioning within any organization [319]: 1) *High availability*, 2) *Low latency* and 3) *Horizontal scalability.* These three characteristics are mandatory for efficient and real-time scheduling and event management in BPM. So in order to ensure a (near) real-time priority-based business process instances management, we resort to an integration of four concepts: IoT (to sense the environment), CEP (to detect situations of interest, since it is considered as the standard course for real-time analysis and situation detection [320]), Machine Learning (to analyze our data, find patterns in it and then make predictions, to facilitate decision making) and BPM (to manage our business processes). The idea behind this integrated quaternity of technologies (see Figure 5.2) is: 1- to accompany data an event explosion resulting from IoT, 2- to examine large data sets in order to uncover hidden patterns, unknown correlations between collected events, either at a very technical level (incident/anomaly detection, predictive maintenance) or at business level (customer preferences, market trends, revenue opportunities), 3- to capitalize business value from the data generated by IoT sensors, 4- to provide improved operational efficiency, better customer service, competitive advantages over rival organizations.

A critical use case for IoT is to alert organizations when a product or service is at risk. Early detection is essential to remedy the problem before it becomes a real issue or to quickly perform a clean-up when a failure occurs [321]. Therefore, to turn this conceptual integrated quaternity of technologies into a concrete reality, we propose an end-to-end IoT-BPM architecture, that we have named IoDEP (IoT-Data-Event-Process). This integration architecture follows Haze Architecture and Cascading Analytics incarnated by a DIKW (Data Information Knowledge Wisdom) discovery pattern crossing the architecture from the device and then Fog/Edge to the cloud, and a learning feedback loop that feeds forward insights to adjust either Fog/Edge or device algorithms as described in Figure 1.2 in chapter 1. As we have seen throughout the chapters of this thesis, IoT-BPM integration involves bidirectional communication. It is possible to acquire data/event from sensors (e.g., monitor and control IoT devices) to manage business process instances and to send instructions to those devices (e.g., reset, adjustment, or shut them down).

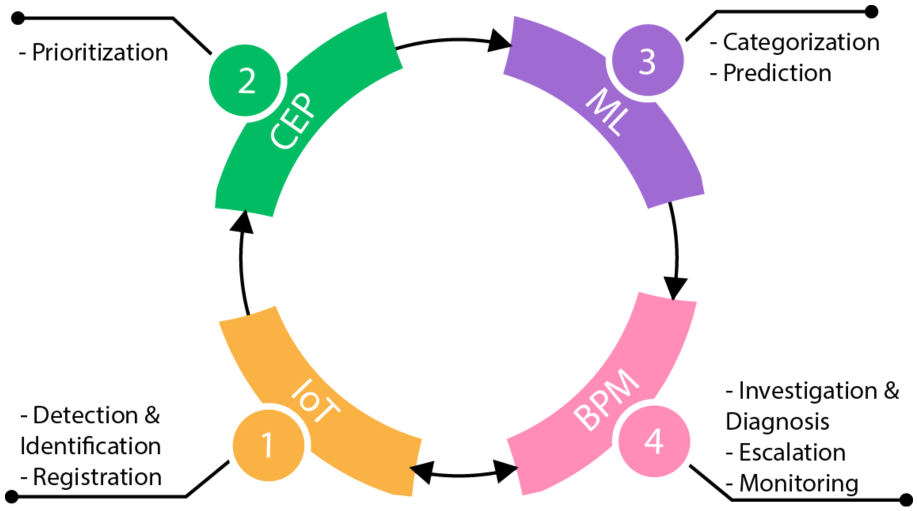

Figure 5.2 – Quaternity view of BPM, IoT, Machine learning and CEP

One of the requirements of our proposed architecture is scalability without imposing an architecture redesign. That is the reason behind using Haze Architecture and Cascading Architecture, as it ensures fluidity and dynamism.

5.3.1 Research Design/Modeling Methodology

As we said previously, to handle the challenges faced by BPM when dealing with IoT objects (see Chapter 1), IoT and BPM need to meet in the middle. The incorporation of two heavy paradigms, such as IoT and BPM, generates, without a doubt, a modeling methodology issue. This issue becomes more complicated when this integration involves other technologies. There are three types of modeling methodologies: 1- The Top-down approach, 2- The Bottom-up approach, 3- and The Meet-in the-middle approach. We propose in this approach a meet-in-the-middle methodology, to facilitate the integration of IoT and BPM. In fact, the meet-in-the-middle approach is considered as a method of refinement going alternately from top to bottom to bottom to top. The combination of deductive and inductive iterative sprints in this approach allows reuse/mutualization and disruptive thinking. In this contribution, we propose the integration of IoT and BPM via an end-to-end architecture, aiming to provide a meet-in-the-middle environment. This environment is capable of capturing data and events from IoT sensors when they are sensing their environment, creating actionable and useful knowledge, and allowing this knowledge to be used in the business layer through business processes.

5.3.2 Functional Requirements

Business processes are supposed to run smoothly in different business situations and contexts. This constantly changing environment requires business processes that can be easily adapted to the appropriate action to be taken. However, without being coupled to other technologies, business processes are still deficient regarding the critical ability to provide assistance to their users, due to the lack of two important aspects (as we have seen in Chapter 1) :

— **Context-awareness** : The emergence or even omnipresence of IoT solutions in different businesses forces organizations to adapt their processes to a high level of connectivity. Context-awareness is a fundamental characteristic of ubiquitous computing [64], and it is the key to benefit from the raw data collected by sensors, as it allows storing contextual information related to these raw data and deciding which data should be processed, to facilitate interpretation [65] especially at the level of business process.

— **Knowledge feedback loop**: traditional BPM systems have different limits, as they do not facilitate the use of knowledge extracted/generated from data by business processes after their execution. As a result, a tremendous amount of data and event data that are constantly collected within the organization are not exploited to improve business processes. In fact, these data represent for enterprises a real engine of growth. However, a large amount of raw data is not valuable; data must go through a whole process to extract value from it. The analysis of large data helps organizations extract information and then knowledge, because the real value is how organizations will use these data and turn their organization into an information-centric company that relies on the insights derived from data analysis for their decision-making [68].

As we have seen in Chapter 1, incident and supervision management processes are considered as an event-driven business process. The instances of these processes are, in some cases, launched by IoT generated events ('Big' data push paradigm: i.e., data are triggering processes). To handle incident management in IoT-BPM architecture, the Information Technology and Infrastructure Library (ITIL) repository has been considered. In fact, according to ITIL V3 [322], incident management processes are composed of the following functions:

— **Detection / Identification:** Detect the incident and identify the first elements of the classification.

— **Registration:** Record basic details of the incident and propagate the incident alert as necessary.

— **Categorization:** Categorize incidents, assign impact and urgency, and thereby define priority and match against known errors and problems.

— **Prioritization:** the incident is prioritized for better utilization of resources and the Support Staff time

— **Investigation and Diagnosis:** Assess incident details, collect and analyze all related information, and resolve (including any workaround) or route to online support.

— **Escalation:** Escalate (functionally or hierarchically) where/when necessary.

— **Resolution and recovery:** Resolve the incident and take recovery actions.

— **Closure:** When the incident has been resolved, the system should ensure that the details of the action taken to resolve the incident are concise and readable, the classification is complete and accurate according to the root cause, and the resolution/action is achieved.

In addition, the IoT-BPM Architecture of incident management processes integrates additional features to achieve the highest levels of IoT maturity. Among those features, we have :

— **Monitoring and Communication :** All information, metrics and key performance indicators applicable to the incident are assessed, recorded and reported (time spent on the incident, support actors, date and time of closure, number and type of reoccurring incidents, average time to achieve incident resolution, percentage of incidents resolved at first-line support that meet the Service Level Agreement, etc.).

— **Prediction:** Predicting an incident before it happens will enable anticipatory incident management. This may help avoiding the incident by actioning problem management, or at least, this moves the predicted incident resolution closer to the incident detection insuring proactive incident management. Thus, by collecting and combining connected devices/sensors data with historical context data, IoT-BPM Architecture could provide a wide variety of ad hoc, proactive and anticipatory incident, anomaly, and problem management.

Table 5.1 represents the functional requirements according to each layer of our proposed architecture.

		IoDEP Architecture				
		IoDEP Architecture Front-End		IoDEP Architecture Back-End		IoDEP Architecture Continuum
		Connected Device/Sensor Processing & Analytics	Fog/Edge Processing & Analytics	Cloud Data Processing & Analytics	Human Processes	Learning Feed back loop
Incident Management Process Functions	Detection/Identification	**X**	**X**			**X**
	Registration		**X**	**X**		
	Categorization		**X**	**X**		**X**
	Prioritization			**X**		**X**
	Investigation & Diagnosis				**X**	**X**
	Escalation			**X**	**X**	
	Resolution & Recovery				**X**	
	Closure				**X**	

Table 5.1 – ITIL Incident Management process Functions vs IoDEP Architecture Analysis Matrix

5.4 Overview of the Layered Architecture

This section provides a high-level overview of the different layers of our proposed IoDEP architecture. Figure 5.3 presents the layers of our architecture. In fact, the purpose of this layered architecture is to meet the requirements detailed in the previous section. Our proposed architecture requires five layers (see figure 5.3). IoT sensing layer or Edge Layer, IoT sensor data acquisition Layer, Detection, identification and registration Layer, Categorization and Prioritization Layer, and Cloud layer. In the following, we will present a detailed explanation of each layer of this architecture.

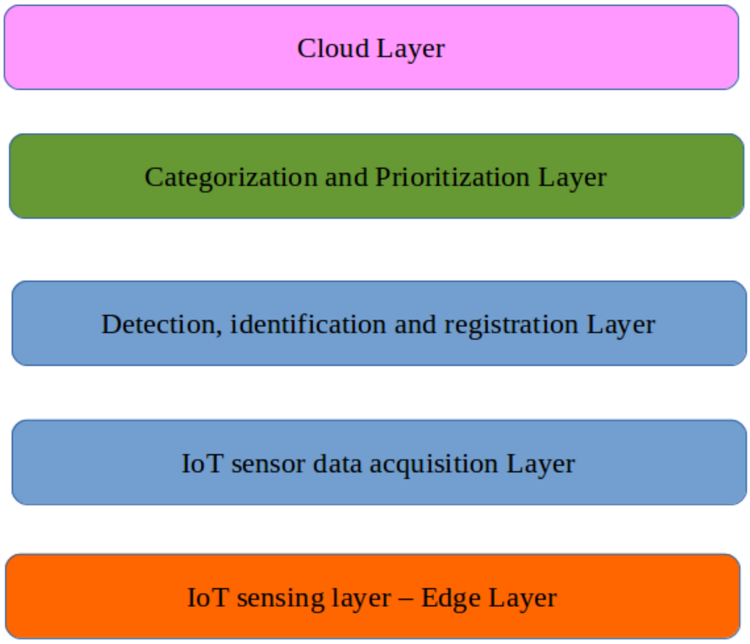

Figure 5.3 – The architecture layers

5.4.1 The Edge Layer

This layer concerns all IoT sensors and devices, each one of these has a precise role depending on their environment, location, and the purpose behind their use, such as hospital, home, geriatric services, retail store, and facilities. IoT devices collect data and generate new events by sensing their environment, after that, those events could launch specific business process instances according to the particularities of the detected situation.

As mentioned above, when dealing with critical use cases, early detection is essential to remedy the issue before it becomes a real problem or quickly do cleanup when the failure hits [323]. The time lag between sensing the environment and sending a notification, a message, or a signal is very critical. In fact, latency is one of the most challenging requirements for connected IoT devices. One of the most effective solutions to reduce this lapse of time is to put some computing on the device or at least bring computing near the device (edge computing).

Pushing data to the edge eliminates long-distance data transmissions to the cloud, which reduces network congestion and latency. Edge nodes are endowed with fall detection capabilities. In fact, these nodes implement several advanced algorithms for image processing and interpretation ([324], [325], [326]). However, this is outside the scope of our research work in this thesis. These algorithms allow the automatic detection and qualification of risky events, following these steps:

— **Image Qualification**: This step allows us to qualify the captured image of the camera. Several types of errors are identified at this stage (Error of recovery of the image, Image of incorrect size, etc.).

— **Movement detection**: This step allows to identify the movements captured by the image, and then render the behaviors of these movements. Mathematical features are identified and extracted from images and videos. Their exploitation allows us to model the behaviors and the events.

— **Movement tracking**: This step consists of tracking objects in their attention zones. Tracking their evolution allows to recognize an abnormal event or behavior and to make a decision.

— **Decision**: at this step a decision is made on whether or not to launch an alert, based on the results of the previous steps.

— **Pre-qualification**: this step is reinforced with the learning feedback loop, by learning from past decisions, mainly false alerts. This final step will consider and exploit the overall context to correct future risk assessments.

At the end of these steps, the detected alerts/events are sent to the cloud. Intercepted events are queued in order to be qualified by a human resource.

5.4.2 The Fog Layer

When dealing with IoT devices and sensors, the cloud by itself cannot connect, process, and analyze data from thousands and millions of objects and devices of different type and nature spread over vast areas. To overcome this issue, Cisco introduced the Fog layer in 2012, in order to offload the cloud through the injection of smart devices into the network layer to provide limited computational resources at the edge of the device layer [327]. Fog computing and edge computing seem similar since both consist of bringing intelligence and processing closer to data creation. However, the location of the intelligence, processing, and computing power is the key difference between these two layers. Generally, intelligence and computing power are placed in devices such as Smart Cameras with embedded vision software (used in our case study). While in the fog layer intelligence and computing facilities are placed in the local area network (LAN) [328].

In our proposed architecture, the fog layer is dedicated to real-time stream processing with CEP (see Section 5.2.1).

5.4.3 The Cloud Layer

Both fog and edge computing are extensions of cloud networks. Most enterprises are already familiar with the cloud, since it is the de facto standard in most industries. The concept of "cloud" was used in several contexts in the 1990s, but only in 2006 when it became associated with the offering of services over the Internet [329]. Cloud computing is "a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction." [330]. The Cloud layer facilitates storing and accessing data and programs over the Internet (as a service) rather than on servers of the enterprise, as we can see in Figure 5.4. In fact, it offers the ability to drastically outgrow an organization's normally available storage, without having to host any additional servers. In our IoDEP architecture, data and intelligence are pushed through layers from edge to cloud in order to be analyzed and processed. The knowledge produced goes through a learning feedback loop that feeds forward insight to adjust either Fog/Edge or device algorithms. To produce this knowledge, we use a machine learning algorithm. We have already presented this technology in Section 3.2.3 Chapter 3.

5.5 Overview of the front-end and back-end of the Architecture

From a front- and back-end perspective, our architecture can be seen as follows:
The IoDEP front-end architecture is composed of two parts:

— **Connected Device/Sensor Processing and Analytics:** insure acquisition of incident data, and incident data filtering and simple classifier.

— **Fog/Edge Processing and Analytics:** insure the following functionalities:

1. **incident data processing** : Detection of an anomaly incident.
2. **incident data analytics** : Pattern recognition/correlation/scoring (advanced supervised time-based analysis algorithms here need smaller training set, but may need more performance resources like GPU).
3. **incident data routing** : Transmission of the anomaly information through an Edge Spooler.

The IoDEP Architecture Back-End is represented by several components insuring cloud side processing and analytics of incident management:

— **Cloud Data Processing and Analytics:** insure the following functionalities:

1. **incident data routing** : Transmission of the anomaly information to the relevant back-end processing and analytics system – ESB/CEP.
2. **incident data analytics and intelligence** : Extraction, cleaning and annotation, Integration, aggregation and representation, Modeling and analysis Pattern recognition/correlation/scoring (more sophisticated supervised machine learning algorithms (e.g. deep learning) may here need big training sets) (Big Data).
3. **incident data processing** : Anomaly Human Processes (Human qualification of the anomaly information) and Enterprise Business Processes (BPMS based on Big Data analytics), and Interpretation : through on Reporting incident KPI Scoreboards based on (Data Warehousing and Data visualization).

5.6 Overview of the application Architecture

In this section, we present a concrete application of our architecture (see Figure 5.4).

The content of the Edge layer can differ from one case study to another. For our case study, IoT devices were represented by smart cameras that detect falls of patients. In other contexts, these devices and sensors could be either Smart locks, Fire and smoke alarms, or smart wristband, etc.

Our Fog layer represents the real-time stream processing layer using CEP technologies. We propose the use of Apache Flink framework as it facilitates complex event processing for real-time analysis. Our choice is based on the benchmark that we have conducted (see Appendix B).

The Enterprise Service Bus layer represents a message broker. It is an integration solution implementing a totally distributed architecture where the applications or services to be integrated are distributed on different machines or information systems, and its role is to ensure communication and interoperability between these different applications whatever their communication protocols. ESB is mostly adapted to asynchronous communications, publish/subscribe messaging, and message queues. Solutions such as Mule ESB [9], PEtALS [10], JBoss ESB [11], Glassfish ESB [12], or Apache Camel [13] could be used in this architecture.

Historical data in our architecture are stored in a PostgreSQL database. However, other Data Base Management System (DBMS) can be used.

For the Business Process Layer, several Business Process management systems can be used such as Bonitasoft [14] (that we have used in our case), Signavio [15], or Camunda [16].

As an application server, solutions such as Apache Tomcat [17], or WildFly [18] can be used in this architecture. As a web server to distribute web content, examples such as Apache Web Server [19] can be integrated into this architecture.

And finally, to communicate the results, a BI Dashboard layer and interactive Dashboards can be build using PHP [20], or Shiny [21] R package given that we have used the R language in our experiments.

This approach has been partially implemented. We did not conduct a proof-of-concept (POC) validation. That is why we did not present results in this chapter. However, we have implemented the CEP part; for more details, see Appendix B.

9. https://developer.mulesoft.com/
10. https://petals.linagora.com/
11. https://jbossesb.jboss.org/
12. https://docs.oracle.com/cd/E19182-01/821-0917/6nluh6gq2/index.html
13. https://camel.apache.org/
14. https://fr.bonitasoft.com/
15. https://www.signavio.com/
16. https://camunda.com/
17. https://tomcat.apache.org/
18. https://www.wildfly.org/
19. https://httpd.apache.org/
20. https://www.php.net/
21. https://shiny.rstudio.com/

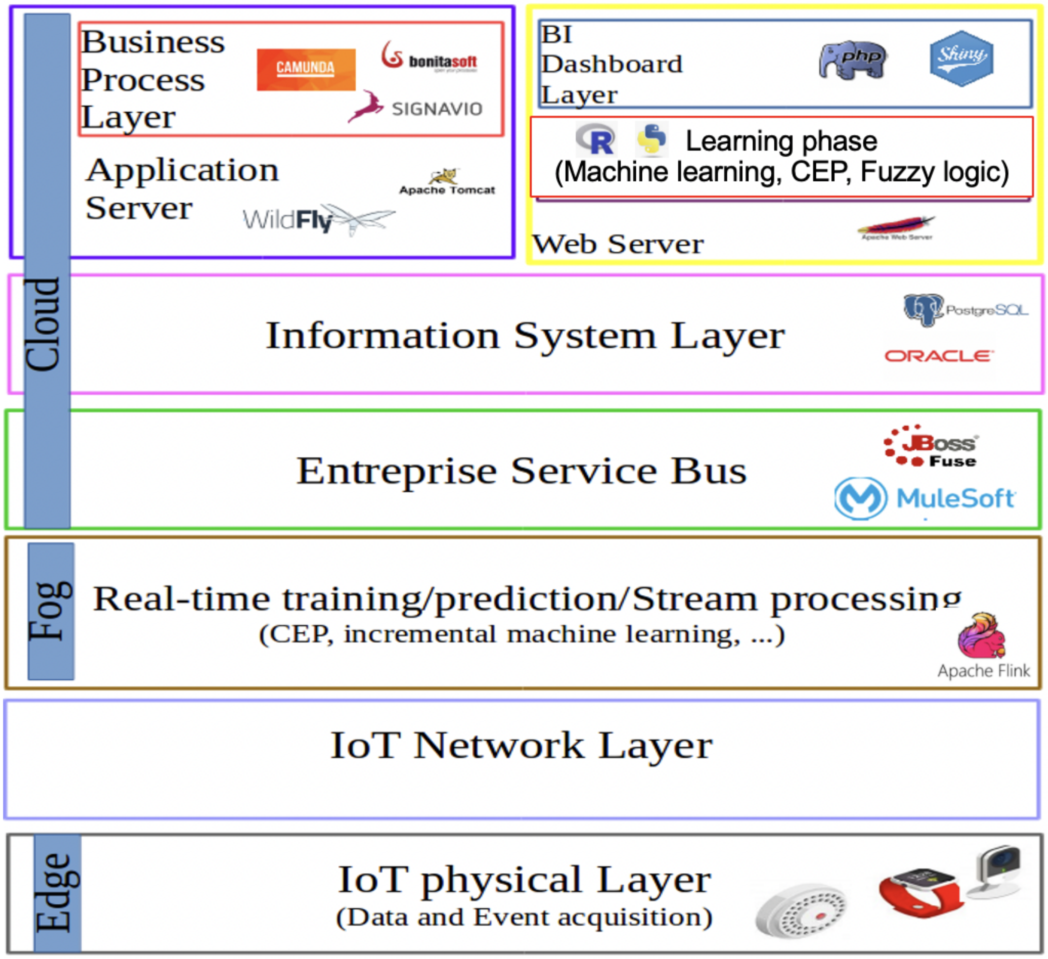

Figure 5.4 – IoDEP Applicative Architecture

5.7 Conclusion

As shown in this chapter, integrating IoT and BPM as a step toward augmented process management that benefits from data and event power is possible via an integrated architecture (IoDEP). The idea behind this architecture is to manage data and events at the same time via an integration approach that includes four concepts: IoT (to sense the environment), CEP (to detect situations of interest since it is considered as the standard course for real-time analysis and situation detection), Machine Learning (to analyze our data, find patterns in it, and then make predictions, to facilitate decision making) and BPM (to manage our business process instances).

We introduced at the beginning of this chapter the reason behind using CEP in our IoT-BPM communication approach. In fact, this bi-directional communication is established through event in one direction and data in another direction. That is why conducting a data analysis approach with event management can facilitates this communication/integration. After that we have presented the different functionalities proposed by this architecture and also the different requirements that should be addressed.

Throughout the different overviews of our proposed architecture, we argue that our approach is generic and can be used in multivariate settings, and most importantly in normal and strict environments where time and priority matter.

Chapter 6

Conclusion and Perspectives

> Puzzles* are sort of like life because you can mess up and rebuild later, and you're likely smarter the next time around. *(Thesis in this case)
>
> Adam Silvera

Several contributions have emerged from this thesis. We can summarize the research problem of this thesis through this question: ***How to exploit data and events in order to augment Business Process Management?***.

6.1 Research Summary

The amount of data produced is growing every year due to the expansion of technological advances and storage capacities. IoT devices and sensors have already generated 13.6 zettabytes of data in 2019 alone, and it is estimated that a total of 59 zettabytes of data will be consumed by the end of 2020 and that this will increase to 150 zettabytes by 2025 [22]. With the SARS-CoV-2 pandemic, this growth has been accelerated due to the use of digitalized means of communication all over the world (meetings, classes, etc.).

All processes deployed in every organization continue to generate a large amount of data (log files, event logs), and these data should be exploited and processed in order to extract the maximum of knowledge, which can subsequently be used to improve business processes. Data and event data are becoming an ubiquitous source of information. It's not surprising that Data science is branching out into all other fields.

Data represent the key ingredient of improvements and paradigm shifts in every domain. However, with the omnipresence of the Internet of Things in almost every research field, we should add events to these ingredients. In fact, data and events are becoming the core driver of new technologies advancement. As we have explained in Chapter 2, data have a huge impact on processes and decisions and also on how things are done in business and applications. As the American computer scientist and Google's Research

22. https://firstsiteguide.com/big-data-stats/

Director Peter Norvig said "*more data beats better algorithms*". That is why data science, as a multidisciplinary approach, is present in all industries and data-intensive domains. With the abundance of data, both academic and industrial research environments are more eager than ever to exploit data in a more predictive and even proactive way rather than only descriptive way. The idea behind this is to assist human intelligence by detecting situations of interest at every step of the process and taking the right decisions and actions. The same reasoning can be applied to events given that organizations are facing an event abundance also due to the ubiquity of IoT devices and sensors. This is where our work comes in handy, as we proposed an integrated architecture that links data to events and the IoT to BPM using machine learning (for data analysis) and CEP (for event management).

The answers to the research questions we raised in the Introduction have been outlined throughout the dissertation. Some answers were made explicitly clear; others were concealed within the sentences and paragraphs. In the following sections, we will revisit our research objectives and envisioned contributions we laid out in the previous chapters of this thesis, and highlight itineraries for future work.

6.2 Contributions Revisited

Business Process Improvement represents a crucial step in the life-cycle of every business process. The main objective of this improvement approach is to help organizations enhance at least one aspect of their business processes (process models, performance metrics, scheduling, etc.). In the literature, we find several methodologies for business process improvement. We started this dissertation with a literature review of the most important BPI existing methodologies. We analyzed each one of them and presented their key advantages and their critical limitations. After that, we focused on the different metrics and indicators used to measure the performance of business processes.

Since the core interest of our research work is the role of data in business process improvement, we highlighted the use of data science as a tool to improve and enhance the different aspects of a business process. Based on the study and the review conducted in the chapter 2, we have determined the aspect of a business process that we want to improve using our proposed approaches.

The research objective of the first contribution, with a focus on improving the scheduling aspect of a business process instances, was to analyze historical event logs from past business process instances execution. Using unsupervised machine learning algorithms for clustering, in order to estimate the criticality level of the incoming IoT events and then manage the priority of these events to achieve a priority-based scheduling of business process instances launched by these events. In this approach, we focus mainly on the scheduling of business process instances. Scheduling tasks and activities within a process instance have not been addressed in this thesis. The reasoning behind this choice was that, for incident process management that involves human resources (such as our case study), critical cases could be handled by one human resource to avoid wasting time. Our contribution ensures that events, which trigger ongoing business processes, are processed according to their level of priority, which is deduced and estimated based on the criticality level of the event that launch this instance. The implementation of this approach in our case study has proven its effectiveness. However, since our approach is mainly based on clustering, we have encountered difficulties in determining the priority of events that belong to the same cluster.

Handling uncertainty regarding the priority level of events from the same cluster progresses the instance scheduling approach throughout Chapter 4. This chapter started by defining uncertainty and providing the reasoning behind choosing Fuzzy Logic in our second approach to handle uncertainty. We propose a solution based on the integration of a Fuzzy Inference System in order to determine the criticality level of the IoT generated events on the one hand (FIS1), and the priority level of the business process instances triggered by those events on the other hand (FIS 2). The results show promising improvements regarding the criticality/priority level of the events generated by some sources that may have the same characteristics.

We propose throughout Chapter 5 an end-to-end IoT-BPM architecture (IoDEP). The idea behind this architecture is to manage data and events at the same time via an integration approach that includes four concepts: IoT (to sense the environment), CEP (to detect situations of interest since it is considered as the standard course for real-time analysis and situation detection), Machine Learning (to analyze our data, find patterns in it, and then make predictions, to facilitate decision making), and BPM (to manage our business process instances). This approach has been partially implemented.

6.3 Limitations and Recommendation for Future research

Although our proposed approaches have provided some innovative solutions to some intricate problems in the field of business process management, there are, nevertheless, some limitations that need to be mentioned.

The first limitation is about the learning algorithm. In fact, using the K-means algorithm was mostly driven by the size and type of our data set. Applying the same algorithm to a different and big data set may change the obtained results and may generate further complexities, such as time complexity. To overcome this limit, three solutions could be explored, depending on the requirements and capacities of each enterprise:

- The first one is using more advanced data analytic techniques and distributed machine learning, as it provides the possibility to scale to larger input data size, as well as enhance performance and increase accuracy, via a multi-node machine learning system.

- The second one is Federated machine learning, which represents a recent approach of distributed machine learning. The concept behind Federated learning is that many participants can collaboratively train a model while the training data is kept decentralized and without being moved to a centralized server or data center or even the cloud [331]. Those participants could be IoT devices, mobile phones, tablets, or even organizations.

The advantage of this type of learning system is that it allows to update the models, to offer a large-scale deployment, and to preserve data privacy and security given that the training data are decentralized. In fact, preserving data privacy is a very important issue especially in critical uses cases (Health, banking, etc.).

- The third solution is Predictive process mining. Which is a field of process mining [332] that focuses on predicting in advance certain aspects of the execution of a running or ongoing (unfinished) process, such as: Execution time of a task or instance, Next activity, or Result (outcome), etc. The purpose of this perspective is to combine the use of predictive process mining on running instances and machine learning techniques on historical data (event logs, data sets, etc.), (see Figure 6.1) to predict the priority of the ongoing activities or instances for more flexibility (in case of execution failure, for example), and also to manage the allocation of resources (priority-based matching of the available resources, when to intervene in case of execution failure or emergency, etc.).

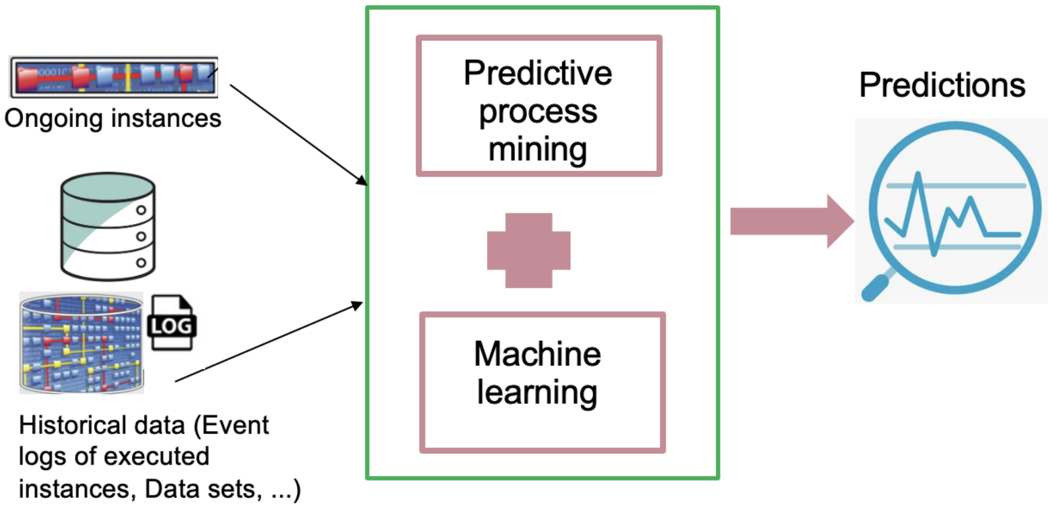

Figure 6.1 – An overview of the predictive process mining perspective

The second limitation concerns our integration approach. We argue that at the level of the IoDEP architecture, our approach is innovative and it can be considered as a first real step towards an event and data based communication between IoT and BPM. And given that IoT devices and sensors are becoming increasingly pervasive in the BPM field, our proposed architecture represents an effective improvement for Business Process Management and a concrete step towards a proactive BPM. However, we have used the basic level of Complex Event Processing (operators, rules, ...), even though we have used an advanced CEP framework such as Apache Flink. That is why we intend to improve the CEP level through automatic CEP rules learning and generation, in order to avoid manual specification of rules, and have more rules that cover more possible cases. This step will facilitate the transition to proactive/predictive management of events/instances.

The third limitation concerns the event logs used in our approach. We have applied our approaches on a static event logs and data set. However, to reduce the delay between the moment when the event occurs in the real world (and the instance is launched) and the moment when the useful information is extracted and used, we propose, as a perspective, to use Streaming process mining [333]. The purpose of this technique is to process a data stream, instead of a static event log, to extract relevant information about ongoing processes.

In this thesis, through a set of experiments, we focused on improving and augmenting some aspects of business process management using data (data analysis, machine learning) and events (complex event management). However, we focused only on instances executed by one human resource from start to end. Therefore, we intend in the future work to explore and manage the priority of more cases such as: process instances executed by several human resources, process instances executed by several human and machine resources, ongoing instances, activities, tasks, etc.

Finally, we intend to improve the proposed end-to-end IoDEP architecture, taking into account all the perspectives proposed in this section. And to validate it using other case studies from the real world.

Appendix A

Clustering Algorithms

Traditional Clustering algorithms categories	Typical algorithm
Clustering algorithm based on partition	K-means, K-medoids, Partition Around Medoids - PAM, Clustering Large Applications - CLARA, Clustering Large Applications based on RANdomized Search - CLARANS
Clustering algorithm based on hierarchy	Balanced Iterative Reducing and Clustering using Hierarchies - BIRCH, Clustering Using REpresentatives - CURE, RObust Clustering using LinKs - ROCK, Hierarchical Clustering Algorithm using dynamic Modeling - Chameleon
Clustering algorithm based on fuzzy theory	Fuzzy c-means - FCM, Fuzzy Compactness and Separation - FCS, Majorization-Minimization - MM
Clustering algorithm based on distribution	Distribution Based Clustering of LArge Spatial Database - DBCLASD, Gaussian Mixture Model - GMM
Clustering algorithm based on density	Density-Based Spatial Clustering of Applications with Noise - DBSCAN, Ordering Points To Identify the Clustering Structure - OPTICS, Mean-shift
Clustering algorithm based on graph theory	CLuster Identification via Connectivity Kernels - CLICK, Minimum Spanning Tree - MST
Clustering algorithm based on grid	STatistical Information Grid - STING, CLustering in QUEst - CLIQUE
Clustering algorithm based on fractal theory	Fractal Clustering - FC
Clustering algorithm based on model	Incremental System for Hierarchical Conceptual Clustering - COBWEB, Self-Organizing Map - SOM, Adaptive Resonance Theory - ART

Table A.1 – Categories of traditional Clustering Algorithms

Modern Clustering algorithms categories	Typical algorithm
Clustering algorithm based on kernel	Kernel K-means, kernel SOM, kernel FCM, Support Vector Clustering - SVC, Maximum Margin Clustering - MMC, Multiple Kernel Clustering - MKC
Clustering algorithm based on ensemble	Cluster-based Similarity Partitioning Algorithm - CSPA HyperGraph Partitioning Algorithm - HGPA Weight Partitioned Canopy K-means - WPCK
Clustering algorithm based on Swarm intelligence	Ant Colony Optimization - ACO Particle Swarm Optimization - PSO Artificial Bee Colony - ABC Shuffled Frog Leaping Algorithm - SFLA
Clustering algorithm based on spectral graph theory	Ng, Jordan and Weiss algorithm - NJW Shi and Malik algorithm - SM
Clustering algorithm based on quantum theory	Quantum Clustering - QC Dynamic Quantum Clustering - DQC
Clustering algorithm based on affinity propagation	Affinity Propagation algorithm - AP
Clustering algorithm based on density and distance	Density and Distance algorithm
Clustering algorithm for spatial data	DBSCAN, STING, Wavecluster, CLARANS
Clustering algorithm for data stream	STREAM, CluStream, HPStream, DenStream
Clustering algorithm for large-scale data	K-means, BIRCH, CLARA, CURE, DBSCAN, DENsity CLUstering - DENCLUE, Wavecluster, FC

Table A.2 – Categories of modern Clustering Algorithms

A.1 Clustering algorithms : A comparative study

In order to choose the most suitable clustering algorithm for our approach, we have led a comparative study. We have chosen 5 clustering algorithms. 4 algorithms from the traditional clustering category (see Table A.1) : K-means (clustering algorithm based on partition), DBSCAN, Mean shift (clustering algorithm based on density), and Fuzzy c-means (clustering algorithm based on fuzzy theory). The fifth one is from the modern clustering category (see Table A.2) : Spectral clustering.

A.1.1 K-means

K-means, also known as Hard C-means (see Subsection A.1.3), is a partitioning technique used to analyze data based on the distance between different data points in the input data set. This algorithm was described by Hartigan in 1975. The idea behind the K-means algorithm is to divide a dataset composed of M data points in N dimensions into K clusters, in such a way that the within-clusters sum of squares is minimized [334].

The most complicated part of this algorithm is to determine the right value of K, which represents the number of clusters. In the literature, we can find several methods to select the most optimal number of clusters for this algorithm, such as the Elbow method [335], the Average Silhouette method [336], or the GAP statistic method [337]. When applying the K-means algorithm in our approach, we have used the Elbow method to determine the value of K.

The basic steps of the K-means algorithm are shown in the following pseudocode:

Algorithm 3 K-Means clustering algorithm

Input:

S = s_1, s_2..... s_m // list of data points (list of sources which generate the different events)

K //Number of clusters

```
1: choose K Random data points from S as initial clusters centroids
2: repeat
3:     Assign each data point s_i to the cluster which has the closest centroids.
4:     Calculate the new centroids of each cluster.
5: until Convergence //no more changes for centroids
   Output: Set of K clusters
```

A.1.2 DBSCAN

DBSCAN or Density-Based Spatial Clustering of Applications with Noise is a density-based clustering algorithm . A cluster for this algorithm is a group of high data point density, that are separated from other data points by a region of low data point density. So each "isolated" dense grouping is considered as a cluster.

This algorithm uses two important parameters [338] :

— **minPts:** the minimum number of data points clustered for a region to be considered dense, and to form a cluster.

— **eps (ε):**this parameter measures the distance that will be used to locate the points in the vicinity of any given point.

The idea behind this algorithm is to randomly choose a point and check if its ε-neighborhood contains at least $MinPts$ points. If it does, this point is considered as part of this cluster. We then go through the rest of the ε-neighborhood points in order to complete the cluster.

The basic steps of the DBSCAN algorithm are shown in the following pseudocode:

Algorithm 4 DBSCAN Pseudo-code

Input:
X = A set of points // the dataset
eps = The neighborhood distance
minpts = The minimum number of points
Output:
Discoverd clusters and potential outliers

procedure DBSCAN (X, eps, minpts)
foreach *unvisted point* $x \in X$ **do**
mark x as visited
$N \leftarrow GETNEIGHBORS(x, eps)$
if $|N| < minpts$ **then**
mark x as noise
else
$C \leftarrow \{x\}$
foreach *point* $x' \in N$ **do**
$N \leftarrow N \backslash x'$
if x' is not visited **then**
mark x' as visited
$N' \leftarrow GETNEIGHBORS(x', eps)$
if $|N'| \geq minpts$ **then**
$N \leftarrow N \cup N'$
if x' is not yet member of any cluster **then**
$C \leftarrow C \cup \{x'\}$

A.1.3 Fuzzy c-Means

Classical clustering techniques result in sharp partitions or clusters. Each instance or data point belongs to one and only one partition, since the clusters are disjoint. This is what we call *hard* or *crisp* clustering. In 1981 James C.Bezdek proposed a Fuzzy clustering approach which is considered as a *soft* clustering [339]. In this type of clustering, we use a membership function. Each cluster is a fuzzy set of all the patterns. In fact, higher membership values signify greater confidence in the assignment of the pattern to a specific cluster [340]. In the literature, we find several fuzzy clustering algorithms. However, Fuzzy C-means is the most popular.

Fuzzy C-Means (FCM) is considered as an extension of the K-Means algorithm. FCM is based on the concept of optimizing an objective function. Fuzzy C-means allows data to belong to more than one group, taking into consideration the uncertainty of cluster assignment. The resulting partition is hence a fuzzy partition. Each cluster is associated with a membership function that indicates the degree to which each data point belongs to the cluster, and each cluster is represented by a cluster center or a prototype.

The general formulation of the fuzzy C-means clustering algorithm is as follows [341] [342]:

Let v_i be the prototype of cluster A_i and let V be the set of all C cluster prototypes.

The objective function of Fuzzy c-means to minimize is :

$$J(U,V) = \sum_{k=1}^{n} \sum_{i=1}^{c} (u_{ik})^m d^2(x_k, v_i), \tag{A.1}$$

Where d^2 represents a distance function.

The condition $\sum_{i=1}^{c} u_{ij} = 1$ for all k is satisfied.

The parameter m represents the fuzzifier. As we have seen before, larger values of m yield a larger number of fuzzy partitions. The following two equations must be solved to achieve this minimization.

Prototypes of the clusters must have the following form:

$$v_i = \frac{\sum_{k=1}^{n} (u_{ik})^m x_k}{\sum_{k=1}^{n} (u_{ik})^m} \tag{A.2}$$

In addition, the required condition on the membership values is as follows:

$$u_{ik} = \frac{\left(\frac{1}{d(x_k,v_i)}\right)^{\frac{2}{m-1}}}{\sum_{j=1}^{c} \left(\frac{1}{d(x_k,v_i)}\right)^{\frac{2}{m-1}}} \tag{A.3}$$

In each iteration of this algorithm, cluster memberships and cluster prototypes are updated.

The basic steps of FCM algorithm are shown in the following pseudo-code. The inputs of this algorithm are the dataset X, the number of clusters C, the fuzzifier parameter m, and the stop criterion ε.

A.1.4 Mean Shift

Mean shift is an unsupervised technique proposed by Fukunaga and Hostetler in 1975 [343]. It is one of the best data analysis methods used for computer vision, object tracking [344], image processing, analysis, and segmentation [345], [346]. But it is also used for other types of data clustering and segmentation such as patients or consumers segmentation [347]. Unlike the K-means algorithm, Mean shift assigns data points to their clusters without having a prior knowledge of the number of clusters [348].

Algorithm 5 Fuzzy C-Means pseudocode

Fuzzy C-Means(X, C, m, ε) //Randomly
$v^0 \leftarrow \{V_1^0, ..., V_c^0\}$
$t \leftarrow 0$
REPEAT
for $k \leftarrow 1$ *to length*(X) **do**
if $d(x_k, v_i) = 0$ for some i
then $u_{ik}^t \leftarrow 1$ and $u_{jk}^t \leftarrow 0$ for $j \neq i$
else Compute u_{ik}^t applying Equation A.3
$t \leftarrow t + 1$
Compute V^t applying Equation A.2 using U^{t-1}
until $\sum_{i=1}^{c} \|v_i^t - v_i^{t-1}\| < \varepsilon$ //$\|*\|$ is any vector norm

The idea behind this algorithm is to shift each data point towards centroids to be the mean of the other data points in its neighborhood [349]. Mean shift is based on the concept of Kernel Density Estimation (KDE), which is a non-parametric method used to estimate the probability function of a random variable. A weight is assigned to each data point using a weight function called a Kernel (such as the Gaussian Kernel). The addition of all these kernels generates a density function or what we call a probability surface [23].

On a high level, Mean shift algorithm proceeds following the steps below :

- **Step 1:** The first step consist on defining a window or a cluster or what we call a bandwith of the kernel, using the terminology of this algorithm, and placing the cluster on a data point.
- **Step 2:** After that, we calculate the mean for all the data points located in this window.
- **Step 3:** Shift the centroid of the window to the location of the mean.
- **Step 4:** Repeat the last two steps until no shift results in a higher density (number of points in the window) or until there is convergence.
- **Step 5:** Delete overlapping windows. When several windows overlap, we keep the window with the most points and delete the others.

A.1.5 Spectral Clustering

Spectral clustering is a class of techniques based on eigen decomposition of affinity, dissimilarity, or kernel matrices [350]. Spectral clustering methods are considered as unsupervised clustering algorithms which are based on the weighted graph partition problem. This technique relies on a similarity matrix to divide data points into disjoint clusters [351]. The data points assigned to the same clusters have a high level of similarity, and the data points assigned to different clusters have a low level of similarity. Unlike conventional clustering algorithms (i.e. K-Means) which assume that the data points that belong to a cluster are spherical about the center of this cluster. Spectral clustering makes no assumptions about the clusters and their forms. Data points compactness is another

23. https://analyticsindiamag.com/hands-on-tutorial-on-mean-shift-clustering-algorithm/

difference between spectral clustering and conventional clustering techniques. In fact, in spectral clustering data points have to be connected but may not necessarily have convex boundaries or be compact.

The core idea of the spectral clustering algorithm is to form a similarity matrix S, compute the Laplacian matrix L and after that compute eigenvectors of L, to perform dimensional reduction and then clustering of the data points.

The spectral clustering algorithm proceeds following the steps below [352] :

Given a set of points $\{X_i\}_{i=1}^{N}$ in $\mathbb{R}^d$ (where d is the dimensionality).

— **Step 1:** Form the similarity matrix $S \in \mathbb{R}^{N*N}$ defined by :
$$S_{ij} = exp(-\|X_i - X_j\|^2/2\sigma^2) \text{ if } i \neq j, \text{ and } S_{ij} = 0$$
Where The scaling parameter σ^2 is used to control how rapidly the affinity S_{ij} decreases with the distance between X_i and X_j.

— **Step 2:** Define D to be the diagonal matrix whose (i, j)−element is the sum of S's i-th row, and construct the matrix $L = D^{-1/2}AD^{-1/2}$.

— **Step 3:** Find $V_1, V_2, ..., V_k$, the top k eigenvectors of L, and form the matrix $X = [V_1V_2...V_k] \in \mathbb{R}^{N*k}$ by stacking the eigenvectors in columns.

— **Step 4:** Form the matrix Y from X by renormalizing each of X' s rows to have unit length, $Y_{ij} = X_{ij}/(\sqrt{\sum_j X_{ij}^2})$.

— **Step 5:** Treat each row of Y as a point in $\mathbb{R}^k$, cluster them into k clusters using K-means

— **Step 6:** Assign the original point X_i to cluster j if and only if the row i of the matrix Y is assigned to cluster j.

A.2 Clustering algorithms : Implementation results

In this section, we present a summary of the results obtained from our experiments. To illustrate the behavior of each clustering algorithm, among the 5 chosen one, and to analyze their differences and adaptability to our case study based on the obtained results, we have conducted a series of experiments.

Experimental settings:

We worked with a data set of patients falls (from 01-02-2016 to 12-06-2017), this dataset contains 238228 observations generated by 81 patients. The historical data in this data set are gathered from our incident management process past instances, and they are partitioned as follows: 89312 alerts are of level 0 (low), 148466 of level 1(average), 275 of level 2 (serious), and 175 of level 3 (very serious). Since the serious and very serious alerts are the most important in our study, we have applied our clustering algorithms only on these two levels.

The data set was stored in a PostGreSQL database and all analyses and algorithms implementation were conducted using R and different R packages [353], [354], such as: (tidyverse [355], RPostgreSQL, ggplot2, dplyr, caret, etc.). And all our experiments were conducted on an Intel(R) Core(TM) i5- 540M2.53GHz. All data have been anonymized.

Experimental results for K-means:

As we have seen before, to determine the value of K, which represents the number of clusters, we have used the Elbow method (see Figure A.1).

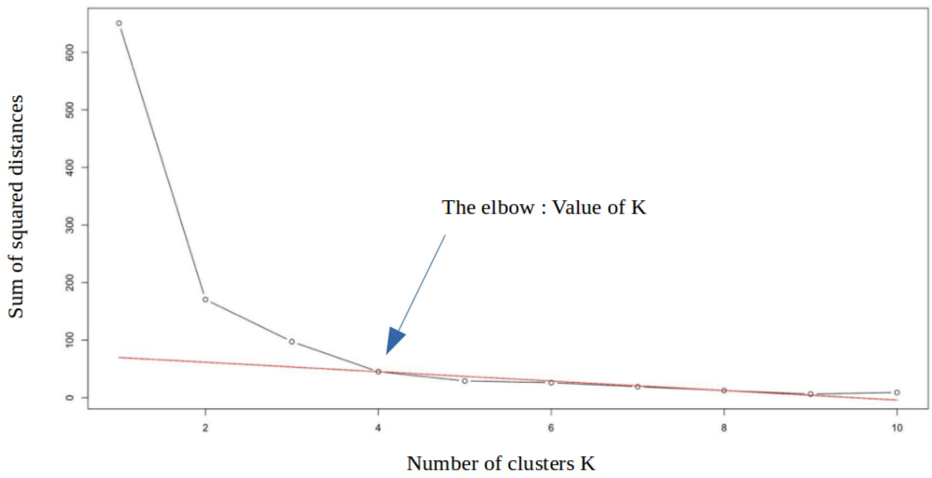

Figure A.1 – Elbow method - sum of squared distances

The plot represents the variation of the sum of squares with the number of classes. We notice that the most adequate number of clusters would be 4, since this point corresponds to an inflection point and the sum of squares seems to stabilize from this point on.

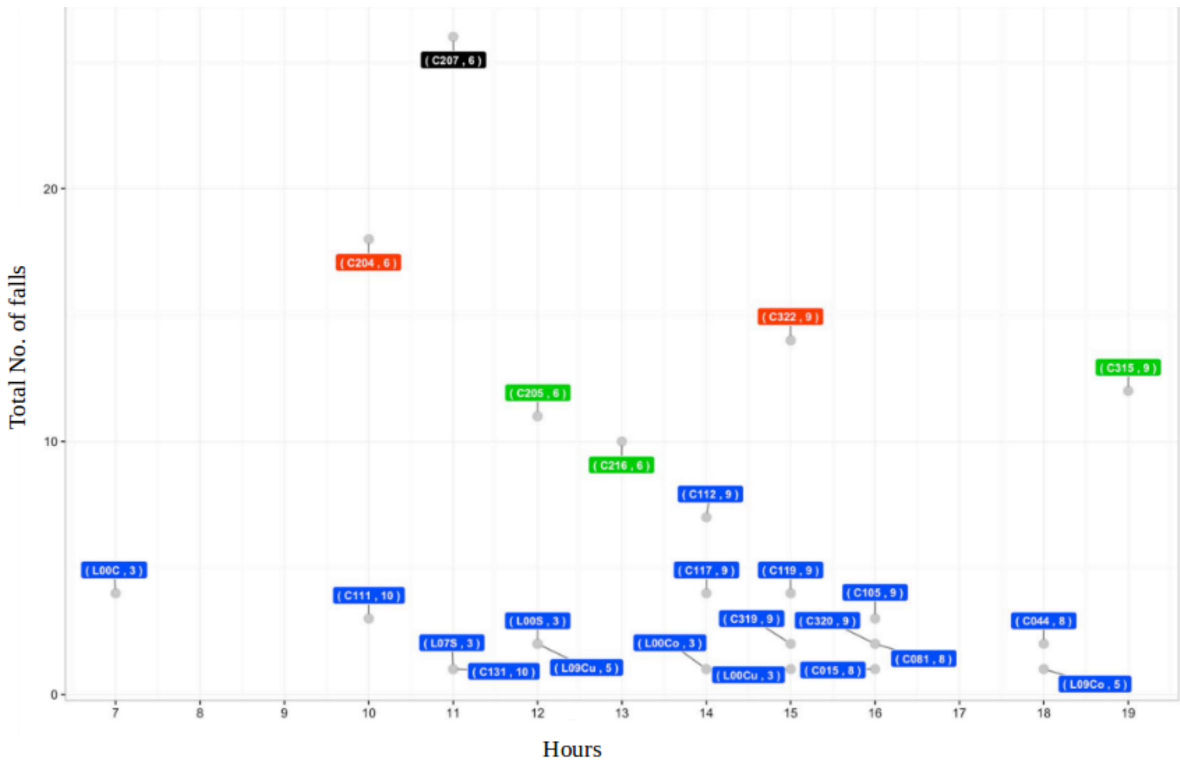

Figure A.2 – Clustering with K-means

As we can see in Figure A.2, patients used in this clustering are divided into 4 clusters: patients with fewer than 8 falls (the blue cluster), 9 to 12 (the green cluster), 13 to 18 (the red cluster) and more than 24 falls (the black cluster). The latter represents the most critical cases.

Experimental results for Fuzzy C-means:

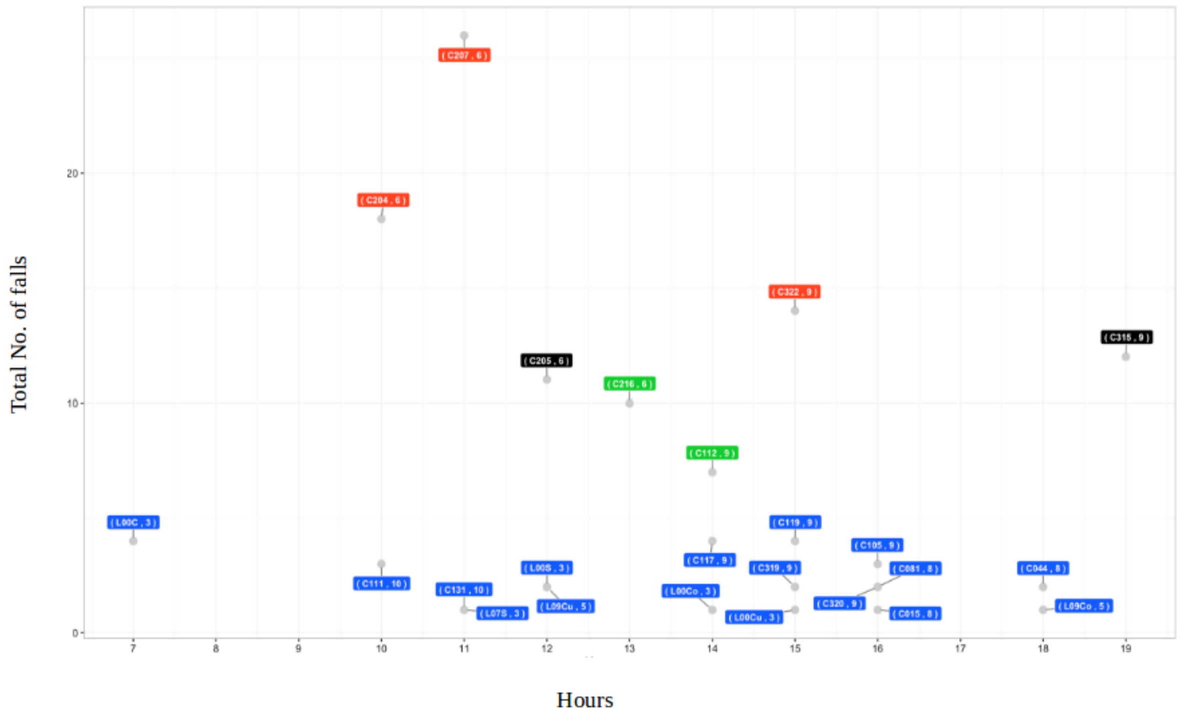

Figure A.3 – Clustering with Fuzzy C-means

At first glance, the results K-means and Fuzzy C-means look very similar. However, the subdivision obtained is not the same. In fact, patient (C112,9) for example belongs in the k-means results to the first cluster that contains patients with falls less than 8. Whereas in Fuzzy c-Means (see Figure A.3), it belongs to the second cluster which contains a number of falls between 7 and 10. The range of the first cluster was thus narrowed. The fourth cluster (black) in K-means contains only one outlier value that exceeds 20 falls. Whereas in Fuzzy c-Means, we find 3 patients with a total number of falls starting from 15.

Although Fuzzy C-means Clustering differs from the previous one because of several factors including the sensitivity to noise, the results obtained are still plausible since there is no overlap between the obtained clusters.

Experimental results for DBSCAN:

As we have explained in the previous section (see Sub-Section A.1.2), eps (ε) and $MinPts$ are the two important parameters in the DBSCAN algorithm. Unlike K-means that proposes several methods to determine the K value of clusters, we do not have an automatic way to determine the most adequate value for the $Minpts$ parameter. The value of $Minpts$ should be set using domain knowledge and data set familiarity [24]. For ε, we can find several methods in the literature to automatically calculate this value. For example, we can determine esp by calculating the average distance between each data

24. https://medium.com/@tarammullin/dbscan-parameter-estimation-ff8330e3a3bd

point and its K nearest neighbors, where k represents the value of $Minpts$ previously selected [356].

Figure A.4 represents the result of the clustering with DBSCAN using the following values: $MinPts = 4$ and $\varepsilon = 0.6$.

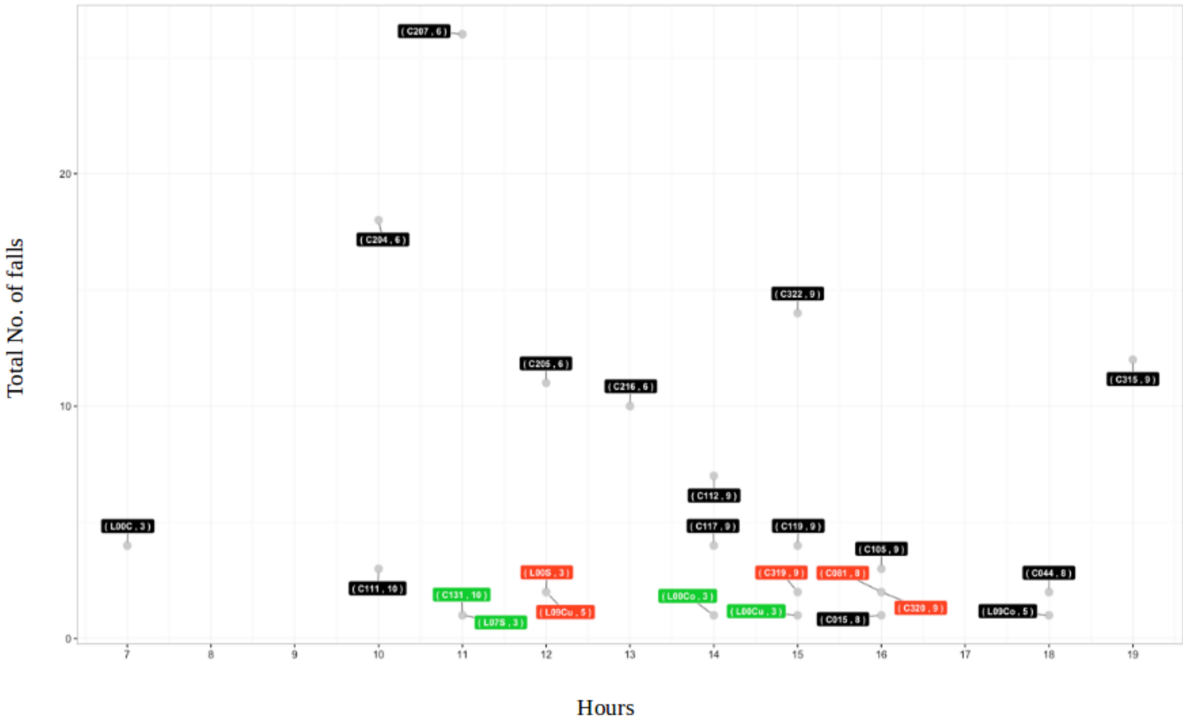

Figure A.4 – Clustering with DBSCAN : MinPts = 4 and eps = 0.6

With these two values, we have obtained 3 clusters as we can see in Figure A.4. Nevertheless, we cannot distinguish the properties of each cluster separately, since the clusters overlap. These results are meaningless and inconclusive; that is why we tried to modify the value of eps parameter.

Figure A.5 represents the result of the clustering with DBSCAN using the following values: $MinPts = 4$ and $\varepsilon = 1$.

With the two new values of $MinPts$ and ε, we have obtained 2 clusters as we can see in Figure A.5. We notice that the two clusters are more or less distinct (except for patients (C117.9) and (C119.9) who have the same number of falls, but belong to two different clusters). These results can be considered as correct but not rigorous and accurate enough. Because, as we can see in Figure A.5, there are even more sub-clusters included in these two obtained clusters. Example: patient (C207,6) with 26 falls should not belong to the same cluster as patient (C117,9) with 4 falls). Therefore, this result is not compelling enough. So we tried this time to modify the value of $MinPts$ parameter, by choosing a bigger value.

Figure A.6 represents the result of the clustering with DBSCAN using the following values: $MinPts = 6$ and $\varepsilon = 0.6$.

As we can see in Figure A.6, by setting the $MinPts$ parameter to 6 instead of 4, the model returns a single cluster grouping all the points in our sample. The result is not

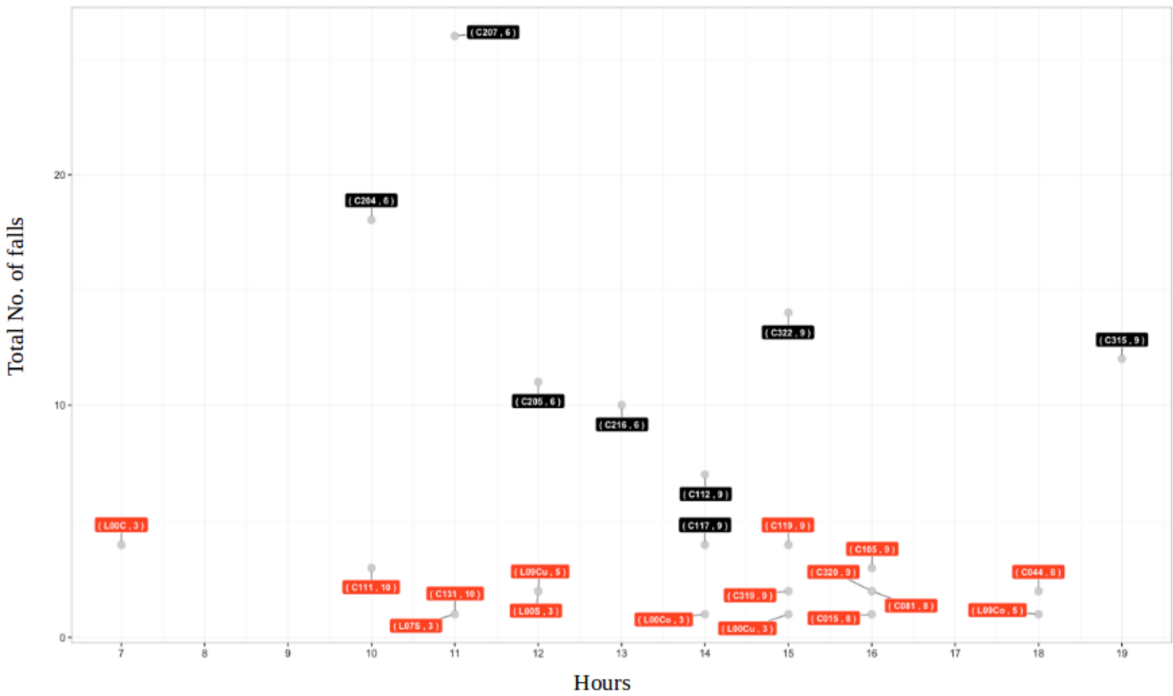

Figure A.5 – Clustering with DBSCAN : MinPts = 4 and eps =1

conclusive at all because no information can be extracted from it. We have tried other combinations of these two parameters, but the results were not conclusive each time.

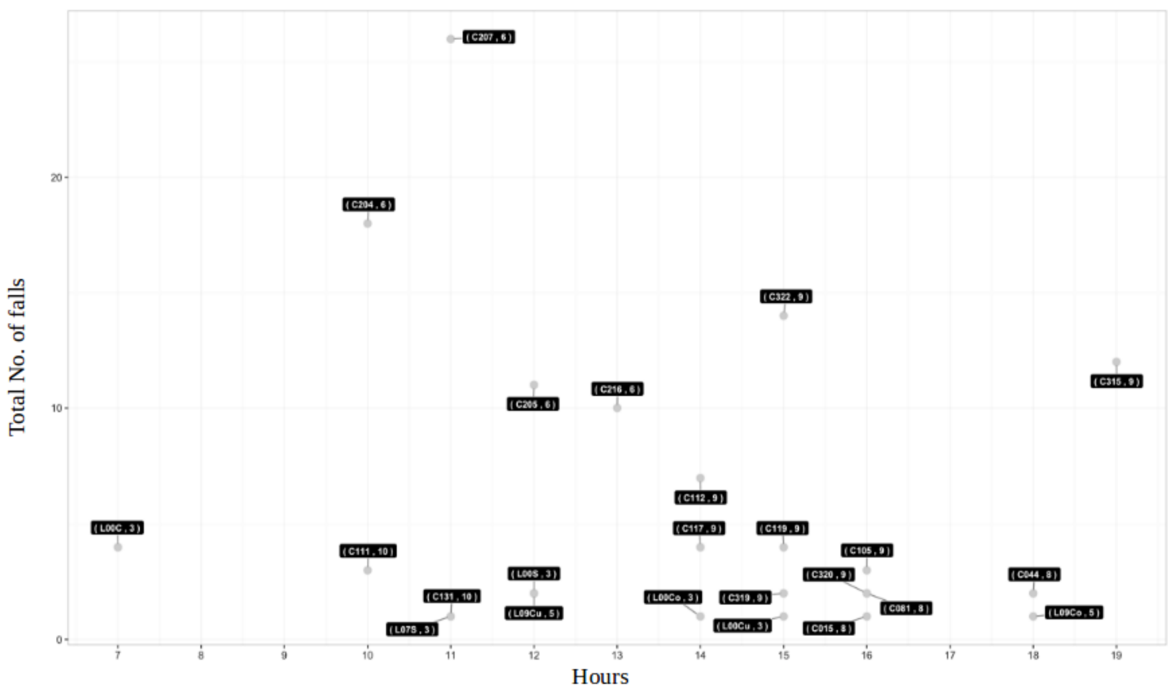

Figure A.6 – Clustering with DBSCAN : MinPts = 6 and eps = 0.6

Basically, the DBSCAN algorithm in our case turned out to be inadequate for the representation of our data. This is largely due to the small size of the data set being used.

Experimental results for Spectral Clustering:

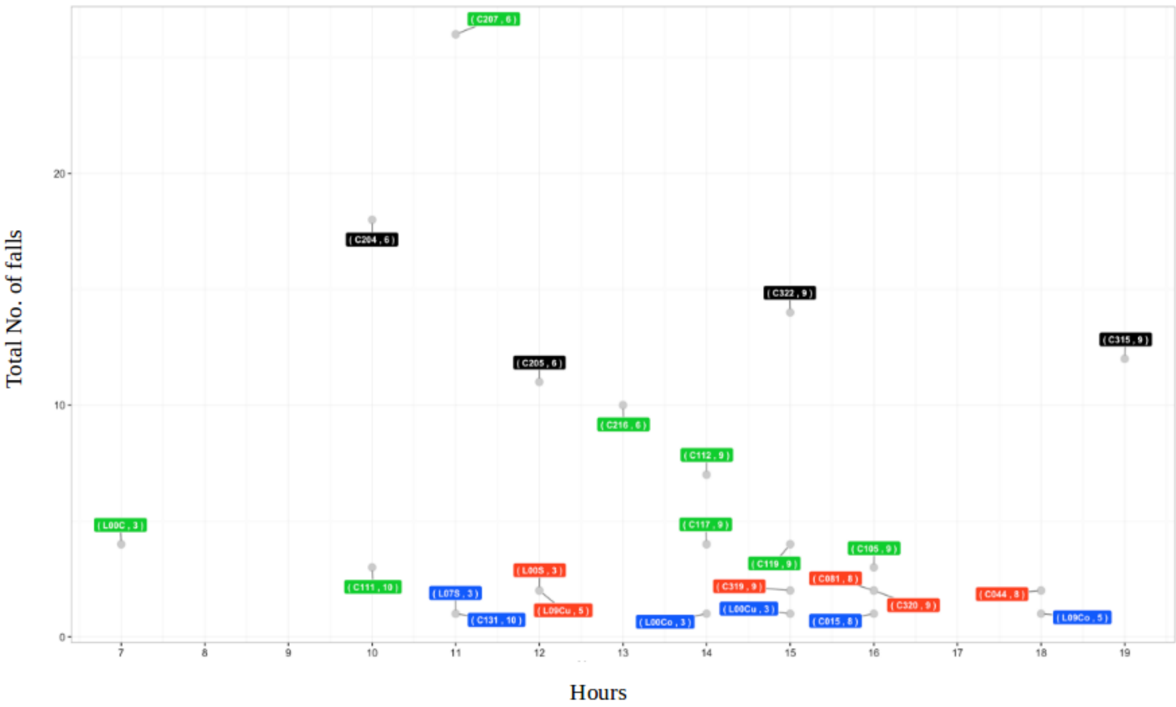

Figure A.7 – Clustering with Spectral Clustering

With Spectral Clustering we obtained 4 clusters. As we can see in Figure A.7, we notice that the four clusters are relatively disjoint and distinguishable, except for the outlier value of patient (C207,4) with 26 falls, but which belongs to a cluster grouping patients with a number of falls between 4 and 10 (the green cluster).

Like Fuzzy C-means Clustering, Spectral Clustering is also sensitive to noise. However, the results can still be considered as plausible since there is no overlap between the obtained clusters.

The conclusion to draw from all these experimental results is that the accuracy of Spectral Clustering in our case is questionable. DBSCAN is not suitable for the representation of our data. Despite the fact that we have found few applications of the Mean shift algorithm in fields other than image processing and computer vision, applying this algorithm to our case study was not obvious and the results were not conclusive at all.

For K-means and, to some extent, Fuzzy C-means, the obtained results were more accurate and concordant with our data set. That is why we have chosen K-means as the clustering algorithm to apply in our case study.

Appendix B

CEP implementation - Chapter 5

This appendix is dedicated to present and discuss the experiments that we have conducted in order to test and analyze the performance of integrating CEP in our architecture and compare our proposed approach to our first approach, which aims to estimate the priority level of each business process instances based on the criticality level of each incoming event that launches these instances (without CEP) that we have detailed in Chapter 3.

To compare the two approaches, we tried to reproduce the same experimental environment. In fact, all our experiments were conducted on an Intel(R) Core(TM) i5-540 M 2.53 GHz. And both approaches have been tested with a data set from our health care case study. More precisely we have a data set of patient falls over the period from 01-02-2016 to 12-06-2017; this dataset is consisted of 238228 observations generated by 81 patients: 89312 alerts are of level 0 (low), 148466 of level 1(average), 275 of level 2 (serious), and 175 of level 3 (very serious).

To implement this approach, we used FlinkCEP, which is a Complex Event Processing library implemented on top of Flink, used to detect event patterns in a stream of events in order to get hold of only what is important in the flow of data. The choice of FlinkCEP was based on several criteria such as performance and response time, the programming language used in this library, and finally the programming community (see Table B.1). In our benchmark we compared FlinkCEP to other CEP libraries or frameworks such as Spark CEP, Siddhi CEP and Esper CEP. (see Tables B.2 and B.3).

Criteria	**Description**
Performance	The intrinsic performance of the tool to perform the required treatments and processing needed.
Viability	Ensuring backward compatibility avoids coding an entire application, and allows services to evolve together without breaking the system.
Resources	Does the tool provides enough resources (documentations, tutorials, conferences, etc.) to facilitate our work ?
Community	It is a very important aspect of any technology as it indicates sustainability, strength and openness

Table B.1 – Criteria of CEP solution benchmark

Criteria	Spark-CEP	Esper-CEP
Definition	- It is a scalable, fault-tolerant streaming processing system. - It supports batch and streaming workloads.	A platform dedicated to complex event processing and also event stream processing (ESP).
Performance and Characteristics	- It provides a better performance when using big data. - High flow rate, ideal for many applications where sub-latency is not required. - It is able not only to process large-scale data streams, but also to query this large-scale data using SQL-like syntax. - Required an external data stores	- Provides an event processing language (EPL) like SQL which is an expressive and extensible language. - Low latency and high throughput. - Exceeds 500,000 events per sec on a dual 2GHz CPU. - Light-weight in terms of memory. - It has data storage and database functionality integrated.
Community	Very large community and massive improvements (More than 1,000 contributors)	A relatively small community (10 Contributors)
Language	Scala - Java - Python - R - SQL	Java - C# - .NET - EPL

Table B.2 – CEP solutions benchmark - 1

Criteria	Siddhi-CEP	Flink-CEP
Definition	A Complex Event Processing engine that used Streaming SQL queries in order to capture events from diverse data sources	A Complex Event Processing library implemented on top of Flink.
Performance and Characteristics	- Low latency and high throughput. - 100k to millions of events per second. - 300,000 events per second (For Uber).	- Flink claims to be 2.5 times faster than Spark. - Flink handles Streams better than Spark. - Required an external data stores
Community	94 Contributors	63 Contributors
Language	Java	Scala - Java - Python

Table B.3 – CEP solutions benchmark - 2

Event-pattern detection with CEP :

One of the functionalities offered by CEP engines is event-pattern detection, figure B.1 illustrates this concept. We have used this functionality to detect critical incidents (falls of patients) as explained in chapter 5.

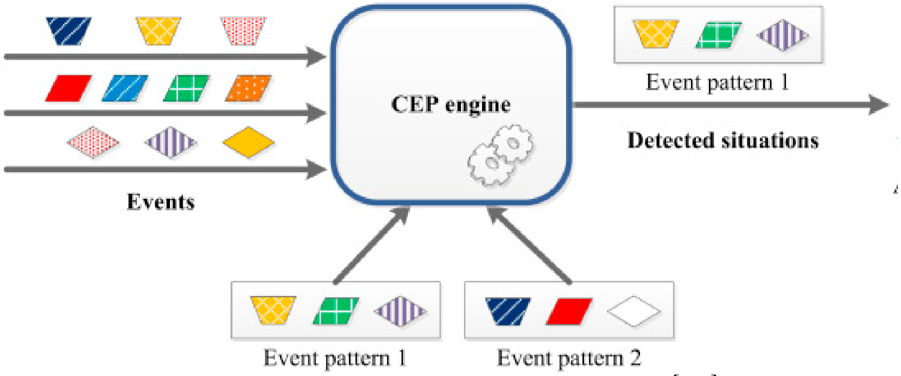

Figure B.1 – Event pattern detection [11]

CEP Rules :
Based on the context of our case study, we have defined some rules that help us estimate the priority level of the incoming event. To define those rules, we take into consideration the available information about the patient, his/her past incidents (falls), and the cluster to which this patient (event source) belongs. Taking into consideration the results of our clustering (detailed in chapter 3 and Appendix A), we tried to manually define some rules based on the domain knowledge. So, basically, we come up with the following rules:

- **IF** the event source belongs to cluster 4 (the critical cluster) **THEN** the new event generated by this source could be serious.
- **IF** The patient has some particular needs (Wheelchair, walker, etc.)**THEN** the new event generated by this source could be serious.
- **IF** The last event generated by this patient within one month was a serious or very serious alert **THEN** the new event generated by this source could be serious.

Figure B.2 represents an example of a CEP script for critical events (falls) detection. This script shows an example of rule definition and a part of critical event detection.

CEP Message broker :

As we have seen before, to manage the different incoming events within a CEP solution, we need a message broker. In this experiment, we have chosen RabbitMQ [25]. It is an open-source message broker, lightweight and easy to deploy.

25. https://www.rabbitmq.com/

```java
        DataStream<MonitoringEvent> inputEventStreamClean = inputEventstream.flatMap(new Tokenizer());
        //Définition des pattern
        Pattern<MonitoringEvent, ?> warningPattern = Pattern.<MonitoringEvent>begin("start")
                .subtype(MonitoringEvent.class)
                .where(new SimpleCondition<MonitoringEvent>() {
                    @Override
                    public boolean filter(MonitoringEvent value) {
                            return Integer.parseInt(value.getAncienneChute())>=CHUTE_GRAVE;
                    }
                }).or(new SimpleCondition<MonitoringEvent>() {
                    @Override
                    public boolean filter(MonitoringEvent value) {
                            return value.isChaiseRoulante();
                    }
                }).or(new SimpleCondition<MonitoringEvent>() {
                            @Override
                            public boolean filter(MonitoringEvent value) {
                            return value.isDeambulateur();
                    }
                            })
                .or(new SimpleCondition<MonitoringEvent>() {
                        @Override
                        public boolean filter(MonitoringEvent value) {
                            return EntityManager.getInstance().hasCurrentYearFallTwice(value.getIdClient());
                }
                            });
```

Figure B.2 – Example of our CEP script for critical events (falls) detection

Global Schema of Event-pattern detection with CEP solution :

Figure B.3 illustrates our first attempt to integrate the CEP engine into our IoT-BPM architecture.

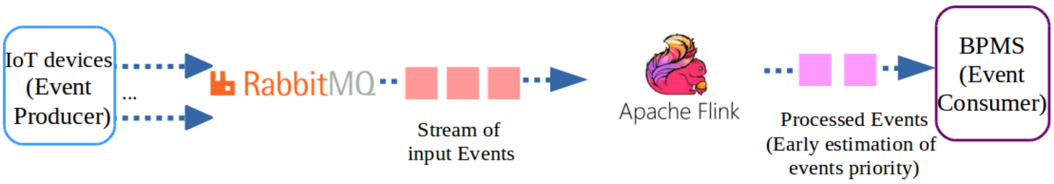

Figure B.3 – Priority-based event management with CEP

The different modules of this architecture operate as follows:

— **Event Producer:** events in this approach are mostly generated by some sensors or IoT devices by sensing their environment (Smart cameras).

— **Lightweight Message Broker:** To manage the amounts of events received and that need to be processed by the CEP engine, we use a message broker that ensures the communication between the source and the target based on a publish/subscribe mechanism. This asynchronous mechanism implemented by message brokers allows the source and target messages to be completely decoupled. Besides, the message brokers can also store the messages locally until they can be processed by the target element. That is why we have chosen RabbitMQ.

— **CEP engine:** Flink CEP is used in this case to filter and process incoming events based on the predefined rules, to detect the events with the highest priority among the incoming stream of events.

— **Event consumer:** represents in this approach a business process management system (BPMS) where processes are managed, executed, and monitored.

Experimental results:

The purpose of this series of experiments is to demonstrate the interest of integrating CEP in the IoT-BPM architecture. So, we will compare two solutions. *Solution 1* represents our first contribution detailed in Chapter 3 (without CEP), and *Solution 2* represents the integration of CEP as detailed in this Appendix and in Chapter 5.

In our experiment for both solutions, we have simulated several event streams with different total number of events (100, 200, 300, 400, 500) (generated from the historical events in our data set). Our objective is to compare the evolution of the computation time of both approaches as a response to increasing the number of input events (concurrent access and non-concurrent access).

Non-Concurrent Access (NCA) :

Total event number	**Computation time (sec) - Solution 1**
100	19.0
200	19.45
300	25.48
400	31.3
500	37.4

Table B.4 – Computation time (sec) for solution 1 - NCA

Total event number	**Computation time (sec) - Solution 2**
100	22.07
200	32.14
300	44.25
400	55.6
500	70.62

Table B.5 – Computation time (sec) for solution 2 - NCA

Concurrent Access (CA) :

Total event nomber	**Computation time (sec) - Solution 1**
200	48
400	72.1
600	108

Table B.6 – Computation time (sec) for solution 1 - CA

Total event nomber	Computation time (sec) - Solution 2
200	27
400	34
600	60.3

Table B.7 – Computation time (sec) for solution 2 - CA

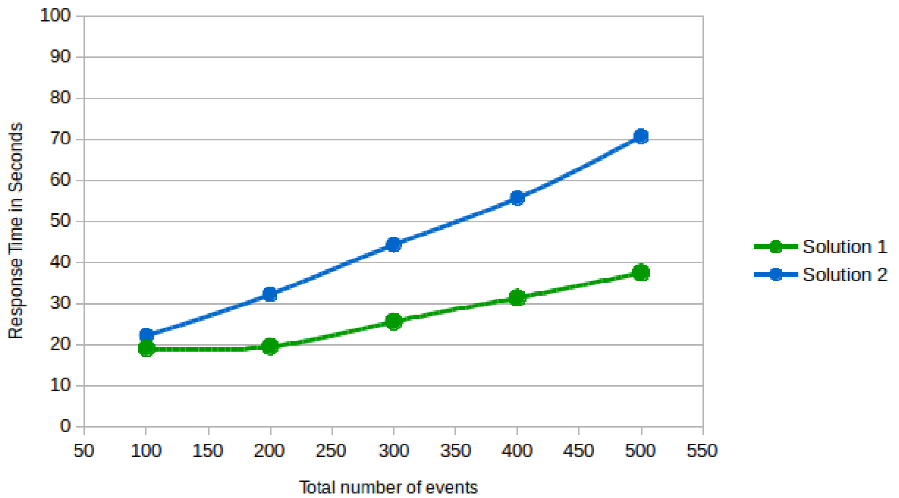

Figure B.4 – Input events with non-concurrent access

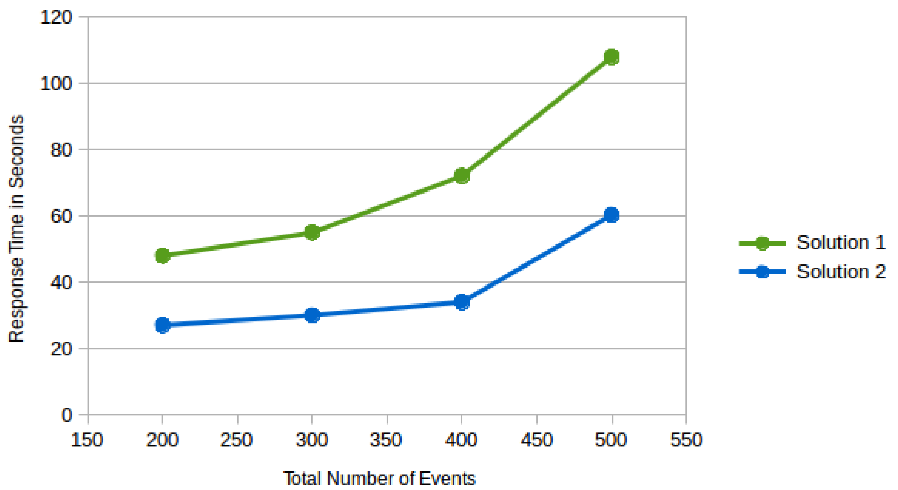

Figure B.5 – Input events with concurrent access

As we can see in Figure B.4, solution 1 presents better results compared to solution 2 when we have a non-concurrent access of the incoming events. However, when we have concurrent access, the CEP-based approach (solution 2) presents better results, especially when we increase the number of incoming events.

Although solution 1 seems to be more efficient at low input event volumes, the CEP solution can perform better, especially if implemented in a Big Data architecture.

For incident management systems, providing a balance between (near) real-time event processing and scalability is very important to achieve an efficient and optimized business process instances scheduling and event management in BPM. Moreover, in real cases, we deal most of the time with concurrent access of incoming events. So, this confirms the efficiency of our assumptions that CEP can provide better results when integrated to an IoT-BPM architecture, and it can also provide better results compared to traditional approaches for business process instances scheduling.

Interested readers can check the complete solution that we have implemented from GitHub [26].

26. `https://github.com/Abir-IA/CEPFlink_EventManagement`

Appendix C

Angel Assistance - Data Set

As stated throughout the chapters of this thesis, one main real-world case study has been used.

The data collection system in this case can be simplified as follows: In each patient's room, there are intelligent video surveillance cameras that are active 24 hours a day and detect abnormal movements. These cameras send their data, which are analyzed, and can trigger alerts. In case of an abnormal signal, an alert will be sent confidentially to designated caregivers, allowing them to intervene quickly through a specialized call center.

Their database (see Figure C.1 gathers all the different information related to the entire video surveillance process. This database thus describes two major processes:

— The first process is about the detection and assessment of the risk level of the patients' falls and the assistance that is provided to them.
— The second process involves the maintenance of the monitoring devices that are installed in the different rooms of patients.

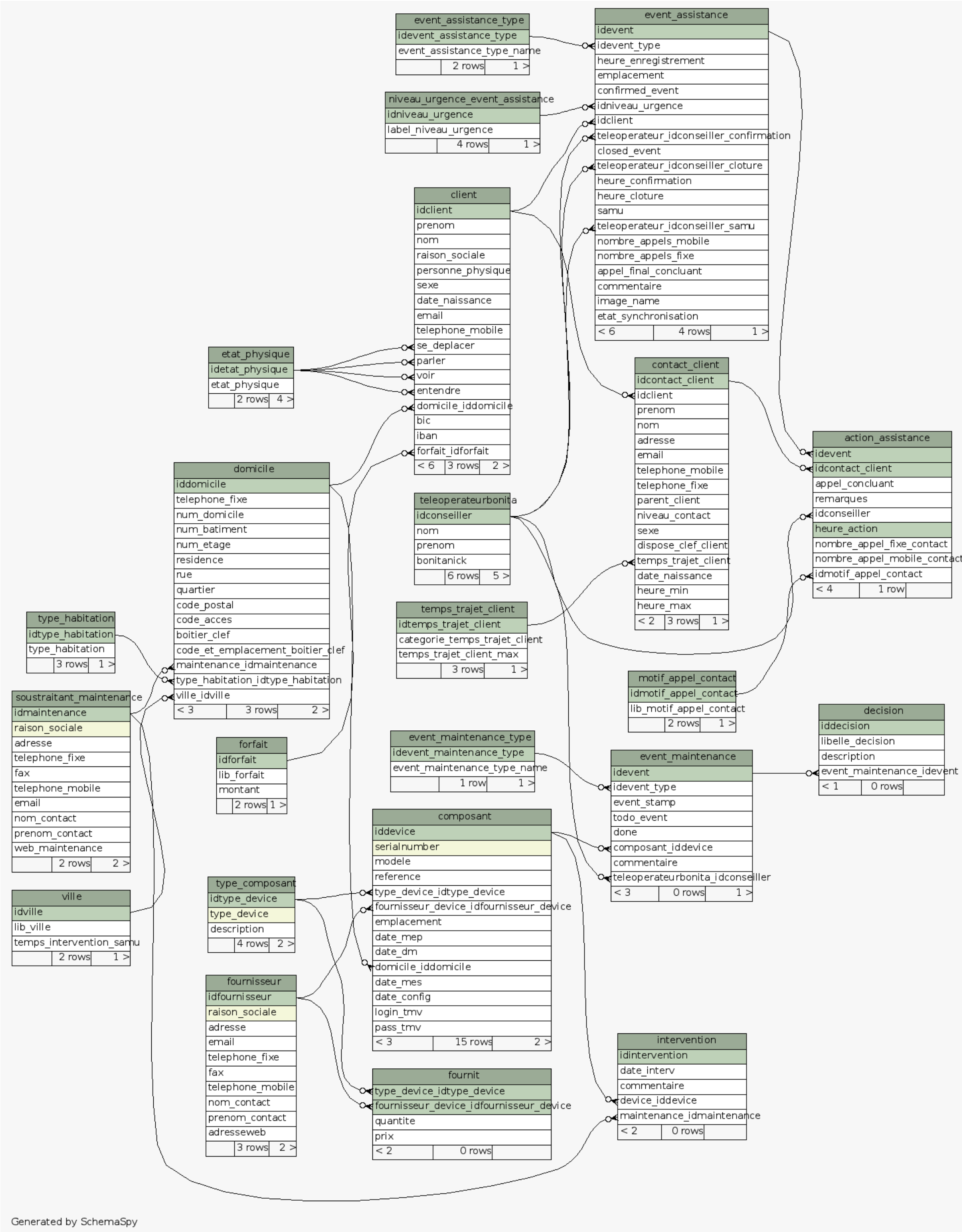

Figure C.1 – Data Base relationships schema

Appendix D

List of Publications

This Appendix lists our accepted publication throughout this thesis.

Conference Proceedings

- Ismaili-Alaoui, A., Baina, K., Benali, K. (2022, Avril). Traitement des événements complexes pour une gestion proactive des instances d'un processus métier. Congrès INFORSID (INFormatique des ORganisations et Systèmes d'Information et de Décision) (INFORSID'22)

- Ismaili-Alaoui, A., Kasmi, O., Baina, A., Baina, K., Benali, K., and Bellafkih, M. (2019, November). Priority-based Event Management using Fuzzy Logic for an IoT-BPM Architecture. In 2019 IEEE 12th Conference on Service-Oriented Computing and Applications (SOCA) (pp. 111-118). IEEE.

- Ismaili-Alaoui, A., Baina, K., and Benali, K. (2019, April). Harnessing the power of data and event data for Business Process Improvement. In CAISAM 2019-Complexity Analysis of Industrial Systems and Advanced Modeling.

- Ismaili-Alaoui, A., Baina, K., Benali, K., and Baina, J. (2018, June). Towards smart incident management under human resource constraints for an iot-bpm hybrid architecture. In International Conference on Web Services (pp. 457-471). Springer, Cham.

- Ismaili-Alaoui, A., Benali, K., Baina, K., and Baina, J. (2018, April). Business process instances scheduling with human resources based on event priority determination. In International Conference on Big Data, Cloud and Applications (pp. 118-130). Springer, Cham.

- Ismaili-Alaoui, A., Lakhrouit, J., and Baina, K. (2015, June). Toward an improved BPM using Big Data technology. 5th . International Symposium ISKO - Maghreb (pp. 156-161).

Journal publications

— Ismaili-Alaoui, A., Baina, K., and Benali, K. "IoDEP : Towards an IoT-Data Analysis and Event Processing Architecture for Business Process Incident Management". In the International Journal of Advanced Computer Science and Applications (IJACSA) - Volume 13 No 4 April 2022.

Appendix E

List of Acronyms

This appendix contains the list of all acronyms used throughout the thesis.

— **BPM:** Business Process Management
— **BPI:** Business Process Improvement
— **BP:** Business Process
— **BPMN:** business process model and notation
— **BPMS:** Business Process Management Systems
— **BAM:** Business activity monitoring
— **EDBPM:** Event-Driven Business Process Management
— **CEP:** Complex Event Processing
— **GA:** Genetic Algorithm
— **PSO:** Particle Swarm Optimization
— **BRS:** Best Resource Selection
— **ACO:** Ant Colony System
— **ABC:** Artificial Bee Colony
— **HRAP:** Human Resource Allocation Problem
— **ML:** Machine Learning
— **IoT:** Internet of Things
— **EDA:** Event Driven Architecture
— **SOA:** Service Oriented Architecture
— **RPA:** Robotic Process Automation
— **ABC:** Activity Based Costing
— **ERP:** Enterprise Resource Planning
— **CRM:** Customer Relationship Management
— **ABC:** Activity Based Costing
— **AI:** Artificial Intelligence
— **RFID:** Radio Frequency Identification
— **FSM** Fuzzy System Modeling
— **FIS:** Fuzzy Inference System
— **ITIL:** Information Technology and Infrastructure Library
— **DIKW:** Data Information Knowledge Wisdom
— **KDD** Knowledge Discovery from Databases

— **YAWL:** Yet Another Workflow Language
— **XML:** Extensible Markup Language
— **WfMC:** Workflow Management Coalition
— **BPRI:** Business Process Run time Interface
— **BPQL:** Business Process Query Language
— **XPDL:** XML Process Definition Language
— **EPC:** Event-driven Process Chain
— **OMG:** Object Management Group
— **BPD:** Business Process Diagram
— **BPR:** Business Process Re-engineering
— **TQM:** Total Quality Management
— **TOC:** Theory of Constraints
— **KPI:** Key Performance Indicators

Appendix F

Research Methodology

This appendix details the research methodology used throughout this thesis and shows how the story of this work has culminated.

BPM is a broad research domain. That is why pinpointing the main issues on which we will work, defining the research questions, and finding these problems at the first place, was the fruit of a thorough and profound research process.

1. **Problem Definition**: As with every research project, we began our work with many attempts to identify and properly formulate the problems we are tackling. This step comes obviously after a first phase of reading and exploration of the existing work in the literature linked to our research domain. During this first step, we had to approach our subject from a practical point of view as well in order to take into consideration the needs of our partner [27]. Our main objective was to address the different challenges faced by our partner regarding the enhancement of their business processes and the management of the execution of these processes by taking into account their resource requirements.
2. **Research Questions**: To get into more detail and seek solutions to the problems we have outlined in the first step, we have to ask some research questions at a more concrete level and attempt to answer these questions throughout our thesis.
3. **State-of-the-art**: Defining the problems in the first step of our research process systematically leads us to this step. Since the state of the art provides a comprehensive description of the current knowledge on the studied subject through the analysis of existing similar or related published works. Whenever we wanted to use, understand, and inspect a technology, we thoroughly reviewed the literature in order to better understand the existing research works and also to know how to position ourselves regarding the existing approaches. For example, when we wanted to check the existing Business Process Improvement approaches, and to explore the link between IoT and BPM, and the integration of CEP in an IoT-BPM architecture, all these issues have led us to a thorough review and inspection of the state-of-the-art approaches. To maintain an up-to-date view of all the existing research work, the state-of-the-art step has been revisited throughout the thesis.

27. Angel Assistance

These three fist steps are linked; in fact, the first step allowed us to thoroughly define the research questions, and the third step enabled us to tackle most of them.

4. **Solutions**: The review of the literature we have done throughout our research work has helped us answer many of our research questions but has pushed us to dig deeper to answer other questions. Exploring the different approaches in the state of the art, with their strengths and weaknesses, has given us some insight about the possible contributions we can propose to enrich our research area, which is BPM. In fact, through the approaches that we have proposed, we have managed to deal with the process instances execution based on their priority, to promote the integration of BPM and CEP in an IoT environment and the exploration of machine learning algorithms within the BPM world.
5. **Validation and Evaluation**: The validation of our work was performed through different scenarios from our Angel Assistance case study.

These five steps outline the research methodology that we followed throughout this thesis. It is noteworthy that they are not sequential, but that we went back to each step several times as needed.

Appendix G

French Summary

Cette annexe représente un résumé de la thèse intitulée "Methodology for an augmented business process management in IoT environment", rédigée en anglais. Les contributions de cette thèse se concentrent principalement sur trois aspects au niveau d'un processus métier à savoir *la gestion de la priorité des instances*, *la gestion de l'incertitude* et *l'intégration/communnication IoT-BPM via une architecture de bout-en-bout - IoDEP*.

G.1 Introduction

La quantité de données produites et consommées augmente chaque année suite à la progression des avancées technologiques et des capacités de stockage. Les dispositifs et capteurs de l'IoT ont déjà généré 13,6 zettaoctets de données durant l'année 2019, et un total de 59 zettaoctets de données a été consommé vers à la fin de l'année 2020 et que ce chiffre passera à 150 zettaoctets d'ici à 2025 [28]. Avec la pandémie de SRAS-CoV-2, cette croissance s'est accélérée en raison de l'utilisation de moyens de communication digitalisés dans le monde entier (réunions, cours, etc.).

Tous les processus déployés dans chaque organisation continuent de générer une grande quantité de données (fichiers journaux, journaux d'événements), et ces données doivent être exploitées et traitées afin d'extraire le maximum de connaissances, qui peuvent ensuite être utilisées pour améliorer les processus métier. Les données et les événements deviennent une source d'information omniprésente. Il n'est pas surprenant que la science des données s'étende à tous les domaines.

Les données représentent l'ingrédient clé des améliorations et des changements de paradigme dans tous les secteurs. Toutefois, avec l'omniprésence de l'internet des objets, nous devrions ajouter les événements à ces ingrédients. En fait, les données et les événements sont en train de devenir le principal moteur de l'avancement des nouvelles technologies. Les données ont un impact énorme sur les processus et les décisions, mais aussi sur la façon dont les choses se déroulent dans les entreprises et au niveau des applications. Comme le souligne Peter Norvig, informaticien américain et directeur de recherche chez Google, "plus de données vaut mieux que des algorithmes plus performants". C'est pourquoi la science des données, en tant qu'approche multidisciplinaire, est présente dans toutes les industries et tous les domaines à forte intensité de données. Avec l'abondance des données, les milieux de la recherche universitaire et industrielle sont plus désireux que jamais d'exploiter les données de manière plus prédictive, voire proactive, plutôt que de se contenter d'un niveau descriptif. L'idée sous-jacente est d'aider l'intelligence humaine en détectant les situations intéressantes à chaque étape du processus et en prenant les bonnes décisions et actions. Le même raisonnement peut être appliqué aux événements, étant donné que les organisations sont confrontées à une abondance d'événements et à l'omniprésence de dispositifs et de capteurs IoT. C'est là que notre travail s'avère utile, car nous proposons une architecture intégrée qui relie les données aux événements et l'IoT au BPM en utilisant l'apprentissage automatique (pour l'analyse des données) et le CEP

28. https://firstsiteguide.com/big-data-stats/

(pour la gestion des événements).

L'objectif principal de cette thèse est de proposer une méthodologie de bout-en-bout basée sur les données et les événements pour augmenter les processus métier dans l'environnement IoT. Afin d'approfondir et de trouver des solutions aux défis que nous avons évoqués dans cette section, il convient de poser quelques questions afin de définir notre problème de façon plus concrète. La figure suivante (voir Figure G.1) résume les questions de recherche que nous avons posées au cours de cette thèse.

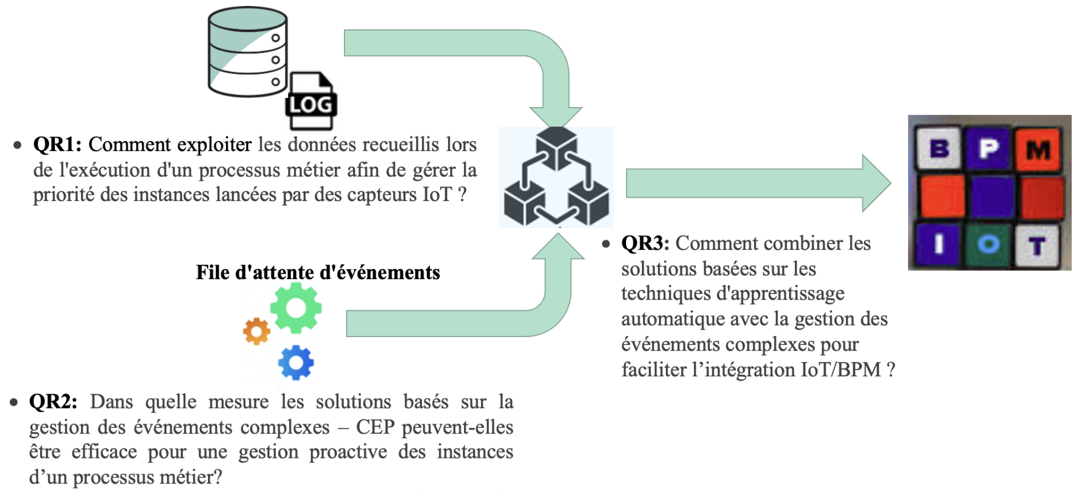

Figure G.1 – Questions de recherche

G.2 Gestion augmentée des processus métier dans un environnement IoT

Les processus métiers peuvent être vus comme une collection d'activités exécutées et coordonnées afin de produire un résultat bien spécifique, répondant aux besoins d'un client (interne et/ou externe). La gestion des processus métiers (Business process management - BPM) est un domaine de recherche très actif dans la discipline des systèmes d'informations. Il a pour objectif l'amélioration continue de l'efficacité et la performance des processus d'une entreprise, par le biais des méthodes, des techniques et des outils qu'il offre afin d'appuyer la conception, la mise en œuvre, la gestion, l'analyse, et l'automatisation, dans la mesure du possible, des processus métier, et donc gagner en termes d'agilité, de flexibilité et de performance.

Même si plusieurs méthodes d'amélioration des processus métier (Business Process Improvement - BPI) sont disponibles dans la littérature, les organisations rencontrent toujours des difficultés pour les appliquer efficacement. Ces difficultés peuvent être justifiées par le fait que les méthodes BPI existantes ne répondent pas à toutes les exigences

récentes des organisations et ne s'adaptent pas aux progrès réalisés, ces dernières années, dans plusieurs domaines tels que l'intelligence artificielle, les techniques d'analyse des données, l'apprentissage automatique, le process mining et le traitement des (flux) évènements, etc.
En outre, avec cette nouvelle ère de digitalisation et l'essor de plusieurs nouvelles technologies telles que le Big Data, l'Internet des objets (IoT), le Cloud Computing, etc, les organisations sont confrontées à de nouveaux facteurs et défis redéfinissant le marché et qui génèrent de réels changements dans le BPM traditionnel. Parmi ces nouveaux défis on trouve la quantité de données et d'évènements, provenant, avec une très grande vélocité, de différentes sources hétérogènes (des interactions internes ou externes de l'entreprise, IoT, etc). Ces données doivent être bien analysées et exploitées afin d'en extraire, des résultats à forte valeur ajoutée qui peuvent aider l'entreprise dans son processus de prise de décision. Cependant, les outils traditionnels proposés par la méthode du management des processus métiers présentent différentes limites concernant le traitement, la fouille et l'analyse des données et l'exploitation des résultats de ces analyses en temps réel.

La nature interdisciplinaire du BPM est un facteur clé qui favorise les perspectives d'amélioration dans ce domaine. L'objectif de ce travail de thèse est de proposer de nouvelles approches pour augmenter les processus métier, en s'appuyant principalement sur l'analyse des données, les algorithmes d'apprentissage automatique et le traitement des évènements complexes, afin d'exploiter les données et évènements génères par l'exécution des processus métier et de trouver des moyens d'améliorer ces processus sous différents angles tels que l'ordonnancement des instances et la gestion des évènements dans un environnement IoT.

L'IoT est en train de devenir une zone d'innovations technologiques et de promesses de développement économique pour de nombreux industries et services. Ce nouveau changement de paradigme affecte toutes les couches de l'architecture d'entreprise, de l'infrastructure au métier. Le Business Process Management (BPM) est un domaine parmi d'autres qui est affecté par cette nouvelle technologie. Pour faire face à l'explosion des données et des évènements résultant, entre autres, de l'IoT, les processus d'analyse de données combinés aux techniques de traitement des évènements, examinent de grands ensembles de données pour découvrir des modèles cachés, des corrélations inconnues entre les évènements collectés, soit à un niveau très technique (détection des incidents/anomalies, maintenance prédictive), soit au niveau métier (préférences des clients, tendances du marché, opportunités de revenus) pour fournir une meilleure efficacité opérationnelle, un meilleur service client et des avantages concurrentiels sur les organisations rivales.

Afin de capitaliser la valeur commerciale des données et des évènements génères par les capteurs de IoT, IoT, Data Analytics, Traitement des Évènements et BPM doivent se rencontrer au milieu.

G.3 Résumé des Contributions

L'amélioration des processus métier représente une étape cruciale dans le cycle de vie de chaque processus métier. L'objectif principal de cette démarche d'amélioration est d'aider les organisations à améliorer au moins un aspect de leurs processus métier (modèles de processus, métriques de performance, planification, etc.) Dans la littérature, nous trouvons plusieurs méthodologies pour l'amélioration des processus métier. Nous avons commencé cette thèse par une revue de la littérature des méthodologies BPI existantes les plus importantes. Nous avons analysé chacune d'entre elles et présenté leurs principaux avantages et leurs limites critiques. Ensuite, nous nous sommes concentrés sur les différentes métriques et indicateurs utilisés pour mesurer la performance des processus métier.

Puisque l'intérêt central de notre travail de recherche est le rôle des données et des événements dans l'amélioration des processus métier, nous avons mis en évidence l'utilisation de la science des données et la gestion des événements comme outil pour améliorer et renforcer les différents aspects d'un processus métier.

Comme nous l'avons vu précédemment, l'objectif principal de cette thèse est de proposer une méthodologie de bout-en-bout basée sur les données et les événements pour augmenter la gestion des processus métier dans un environnement IoT. Les principales contributions de cette thèse seront détaillés dans la suite de cette section.

G.3.1 Contribution 1 : Amélioration des processus métier via un ordonnancement des instances basé sur la priorité

L'objectif de cette première contribution était d'améliorer l'aspect ordonnancement des instances de processus en fonction de leur priorité, en analysant les journaux d'événements historiques des exécutions passées du processus métier. Nous avons utilisé des algorithmes d'apprentissage automatique non supervisés pour le clustering, afin d'estimer le niveau de criticité des événements IoT entrants, puis de gérer la priorité de ces événements pour réaliser un ordonnancement proactif basé sur la priorité des instances de processus métier lancées par ces événements.

Cette approche est caractérisée par deux phases (Voir Figure G.2). Dans la première phase, la phase d'apprentissage, un modèle de clustering est construit en exploitant les données contenues dans le log d'événements ou le jeu de données des événements d'exécution. Des clusters de priorité sont déterminés à l'aide de ce modèle de clustering. Ces clusters sont utilisés pour déterminer les niveaux de priorité des instances. Nous calculons également le score de Fiabilité de chaque ressource humaine (qui va exécuter l'instance en question) à ce niveau. Dans la deuxième phase, la phase d'exécution ou run-time (Voir Figure G.2), le modèle appris est exploité afin d'obtenir une estimation du niveau de priorité lié à chaque instance candidate qui sera exécutée par une ressource humaine. Les instances candidates sont le résultat de l'étape de filtrage, où chaque instance qui ne représente pas un incident ou qui ne nécessite pas l'intervention d'une ressource humaine est éliminée. L'étape la plus importante de cette phase est le matching d'une instance de processus avec une ressource humaine appropriée, sur la base du niveau de priorité de l'instance et du score de fiabilité de la ressource.

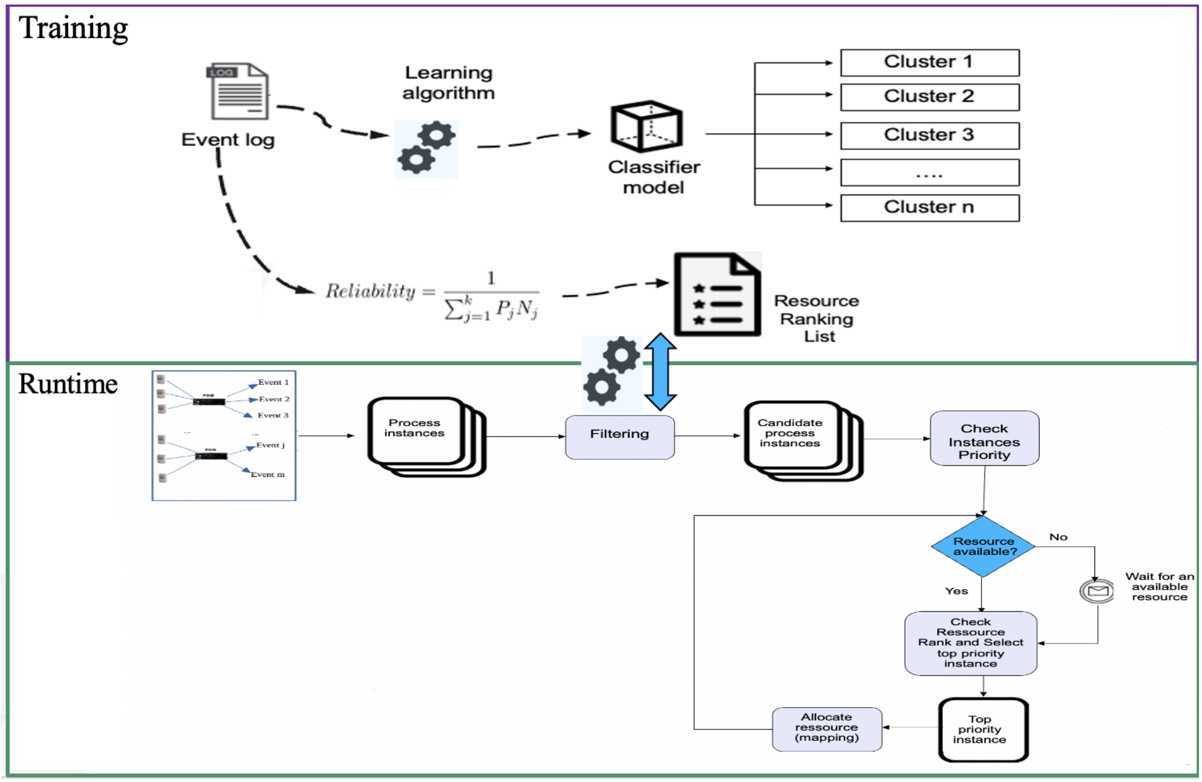

Figure G.2 – Aperçu de l'approche proposée

L'implémentation de cette approche dans notre étude de cas a prouvé son efficacité. Cependant, comme notre approche est principalement basée sur le clustering, nous avons rencontré des difficultés pour déterminer la priorité des événements qui appartiennent au même cluster. Nous avons essayé de gérer ce degré d'incertitude en intégrant la logique floue.

G.3.2 Contribution 2 : Gestion de l'incertitude dans les processus métier par la logique floue

La gestion de l'incertitude concernant le niveau de priorité des événements d'un même cluster (et donc qui partagent les mêmes caractéristiques - voir Figure G.3) est l'objectif principal de notre deuxième contribution. Nous avons commencé par définir l'incertitude et expliquer le raisonnement derrière le choix de la logique floue.

Cette approche est basée sur l'intégration d'un système d'inférence floue afin de déterminer le niveau de criticité des événements générés par un objet connecté d'une part et le niveau de priorité des instances de processus métier déclenchées par ces événements d'autre part. Le principal avantage de cette approche est que la logique floue nous permet de gérer le degré d'incertitude concernant le niveau de criticité/priorité des événements générés par certaines sources qui peuvent avoir les mêmes caractéristiques.

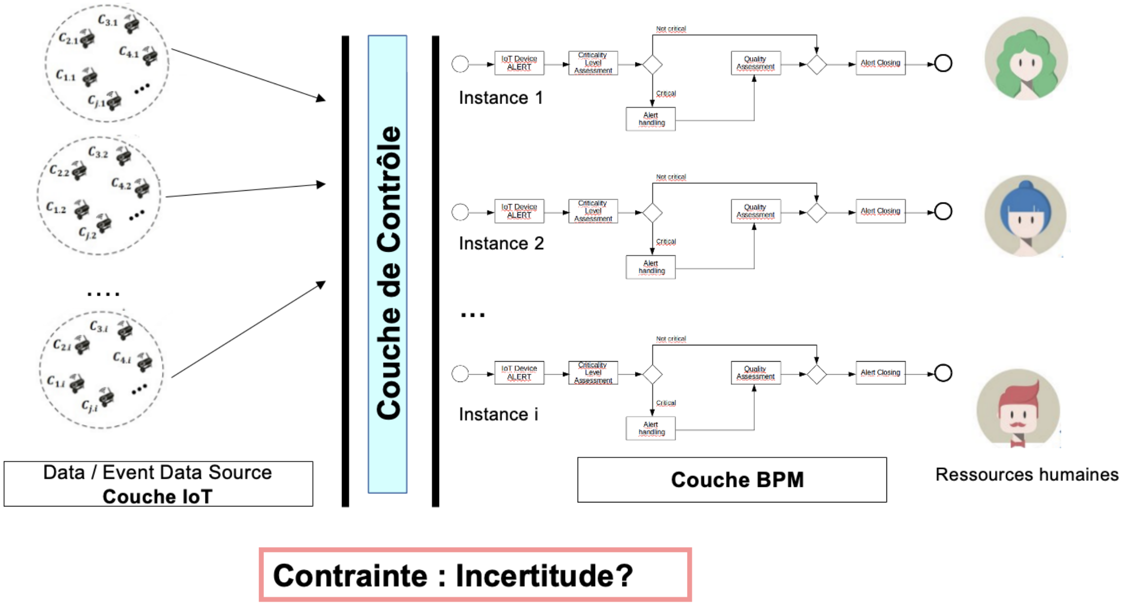

Figure G.3 – Contrainte d'incertitude liée à la gestion des priorités

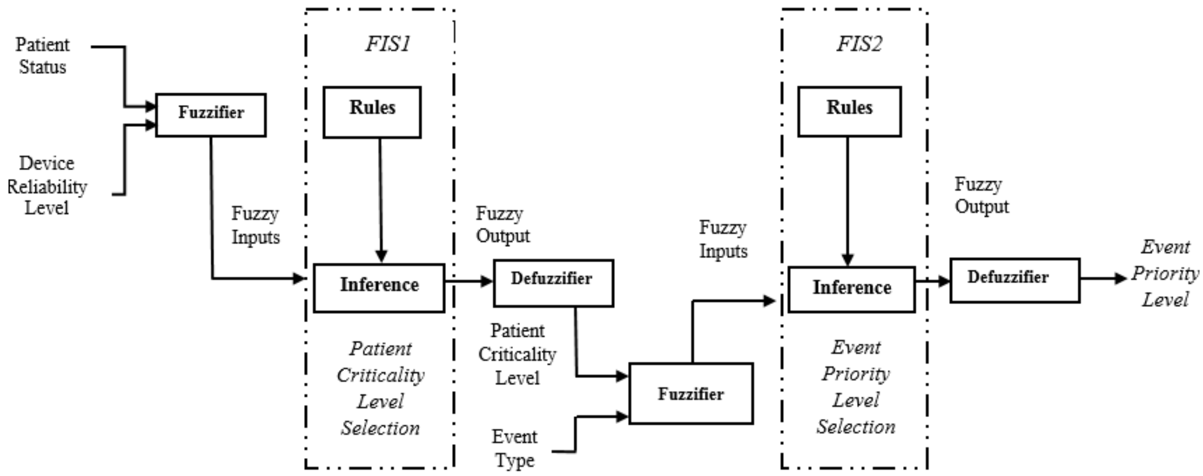

Figure G.4 – Système d'inférence floue pour la gestion d'incertitude

Notre système d'inférence floue (Fuzzy Inference System - FIS) est composé de deux systèmes d'inférence floue en série (FIS 1 et FIS 2) voir Figure G.4. Le premier étant destiné à la détermination du niveau de criticité de l'évènement et le second à la détermination du niveau de priorité de l'instance.

Si on zoom sur la Figure G.4, on va voir que chaque système comprend quatre étapes importantes : la fuzzification, les règles If- Then, l'agrégation et la défuzzification. Les

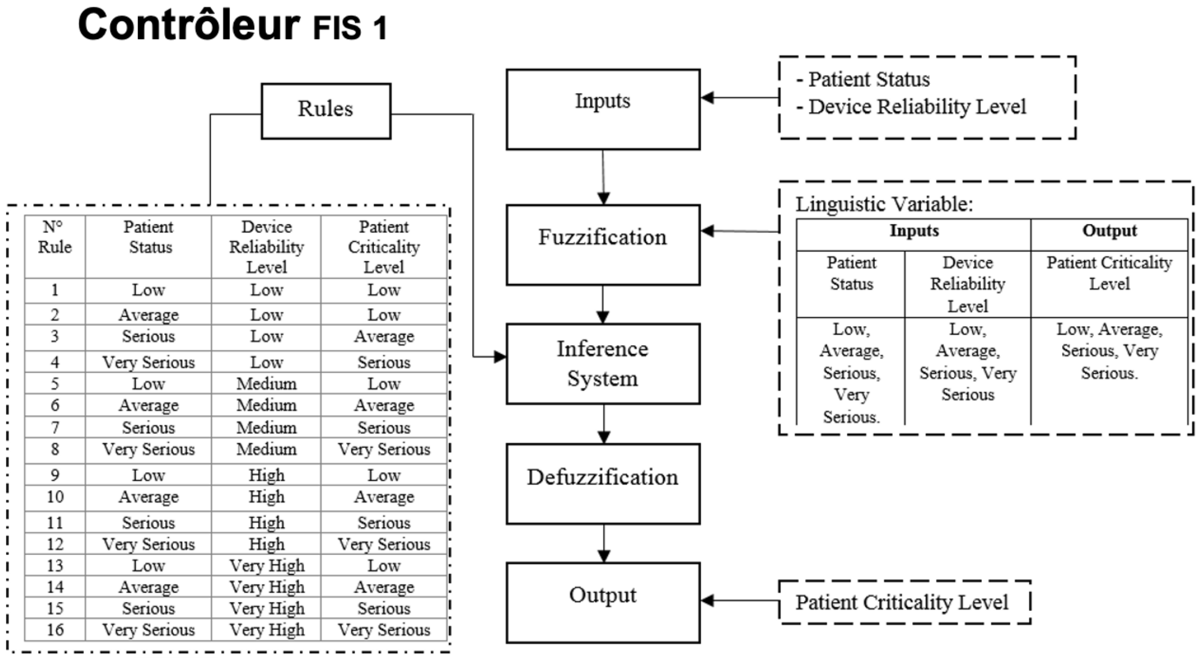

Figure G.5 – FIS1: Détermination du niveau de criticité d'un événement

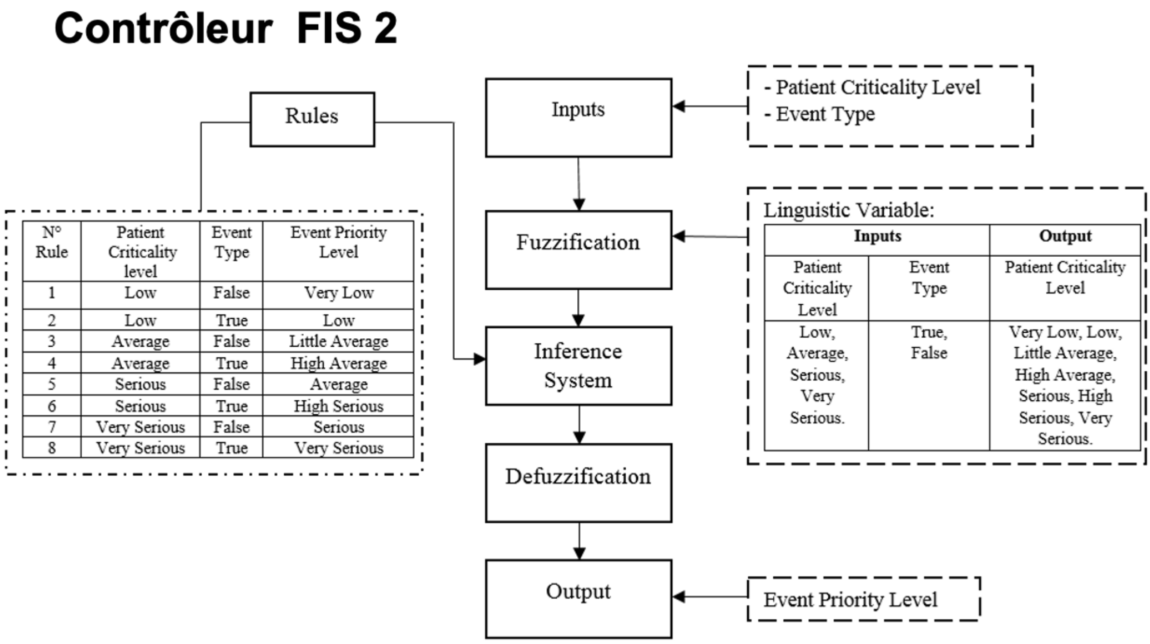

Figure G.6 – FIS2: Détermination du niveau de priorité d'une instance

figures (G.5 et G.6) résument les étapes d'un système d'inférence floue et décrivent les différents blocs du FIS1 pour la détermination du niveau de criticité de l'évènement et du

FIS2 pour la détermination du niveau de priorité de l'instance, respectivement.

Les résultats obtenus, après l'implémentation de cette approche, montrent des améliorations prometteuses concernant le niveau de criticité/priorité des événements générés par certaines sources qui peuvent avoir les mêmes caractéristiques.

G.3.3 Contribution 3 : Architecture IoDEP: Intégration IoT/BPM vers un BPM augmenté

Comme troisième contribution, nous avons proposé une architecture IoT-BPM de bout-en-bout. Nous avons nommé cette architecture IoDEP (IoT-Data-Event-Process). L'idée derrière cette architecture est de gérer les données et les événements en même temps via une approche d'intégration qui inclut quatre concepts :

- IoT pour détecter l'environnement,
- CEP pour détecter les situations d'intérêt via une analyse en temps réel d'un flux d'évènements,
- Machine Learning pour analyser nos données, y trouver des modèles, puis faire des prédictions, afin de faciliter la prise de décision,
- BPM pour gérer nos instances de processus métier.

Cette architecture permet une meilleur communication/intégration IoT-BPM et une gestion proactive de bout-en-bout des incidents, en respectant toutes les étapes du processus de gestion des incidents à savoir : la Collecte, le Filtrage, la Détection/Identification, l'Enregistrement, le Traitement, la Catégorisation, la Priorisation, l'Investigation et diagnostic et la Résolution.

Cette approche a été partiellement mise en œuvre.

G.4 Résumé des Perspectives

Les conclusions tirées dans cette thèse ouvrent des perspectives à court et moyen terme. Certaines améliorations possibles pour notre travail ont été identifiées et sont listées ci-après:

- Améliorer la gestion proactive des processus métier via l'exploration du prédictive process mining. Il s'agit d'un domaine de la fouille de processus (Process mining) qui se concentre sur la prédiction à l'avance de certains aspects de l'exécution d'un processus en cours (inachevé), tels que : Le temps d'exécution d'une tâche ou d'une instance, La prochaine activité, ou Le résultat (outcome), etc. L'objectif de cette perspective est de combiner l'utilisation de l'exploration prédictive de processus sur des instances en cours d'exécution et des techniques d'apprentissage automatique (la première contribution de cette thèse) sur des données historiques (event logs, data set, etc.), (Voir Figure G.7) pour prédire la priorité des activités ou instances en cours d'exécution pour avoir plus de flexibilité (en cas d'échec d'exécution par exemple), et aussi pour gérer l'allocation des ressources (matching basée sur la priorité des ressources disponibles, et savoir quand intervenir en cas d'échec d'exécution ou en cas d'urgence, etc).

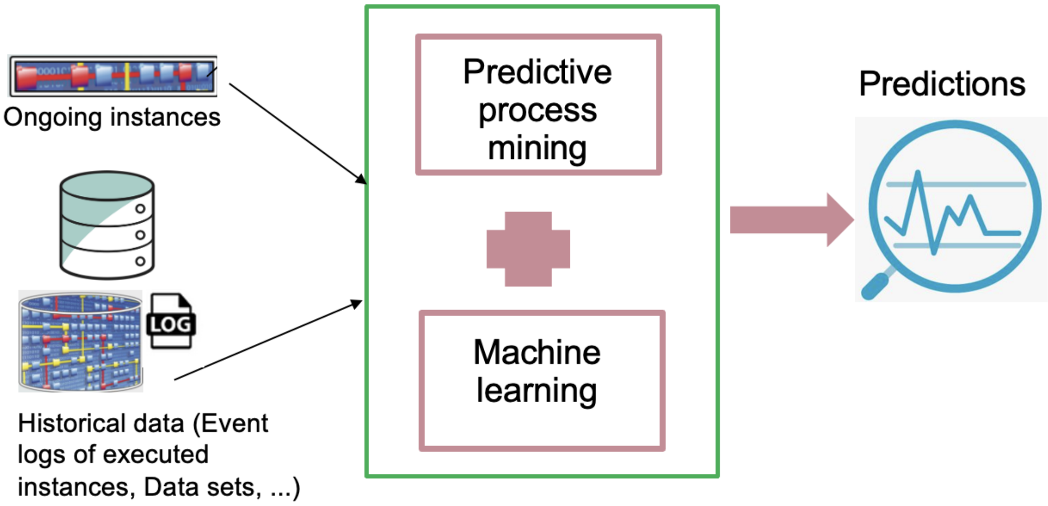

Figure G.7 – Aperçu de la perspective liée à l'exploration prédictive des processus

- Nous avons appliqué nos approches sur un ensemble de données et de journaux d'événements statiques. Cependant, afin de réduire le délai entre le moment où l'événement se produit dans le monde réel (et où l'instance est lancée) et le moment où l'information utile est extraite et utilisée, nous proposons, comme perspective, d'utiliser le Streaming process mining. L'objectif de cette technique est de traiter un flux de données, au lieu d'un log statique d'événements, pour extraire des informations pertinentes sur les processus en cours.

- Améliorer l'aspect du traitement des événements par l'apprentissage et la génération automatiques de règles, afin d'éviter la spécification manuelle des règles et de disposer d'un plus grand nombre de règles couvrant plus de cas possibles. Cette étape facilitera la transition vers une gestion proactive/prédictive des événements et des instances.

Bibliography

[1] Christian Janiesch, Agnes Koschmider, Massimo Mecella, Barbara Weber, Andrea Burattin, Claudio Di Ciccio, Giancarlo Fortino, Avigdor Gal, Udo Kannengiesser, Francesco Leotta, et al. The internet of things meets business process management: a manifesto. *IEEE Systems, Man, and Cybernetics Magazine*, 6(4):34–44, 2020.

[2] DENIS Canty. Architecting for dynamic analytics in iot. `https://dzone.com/articles/architecting-for-dynamic-analytics-in-iot`, 2014. [Online; accessed July-2021].

[3] Marlon Dumas, Marcello La Rosa, Jan Mendling, and Hajo A Reijers. Fundamentals of business process management.

[4] Remco M Dijkman, Marlon Dumas, and Chun Ouyang. Semantics and analysis of business process models in bpmn. *Information and Software technology*, 50(12):1281–1294, 2008.

[5] Ayodeji Adesina and Derek Molloy. A business process management based virtual learning environment-customised learning paths. In *CSEDU (1)*, pages 365–368, 2011.

[6] Usama Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth. From data mining to knowledge discovery in databases. *AI magazine*, 17(3):37–37, 1996.

[7] Tamim Saleh Erik Lenhard and David Opolon. How to Avoid the Big Bad Data Trap. `https://www.bcg.com/fr-fr/publications/2015/big-data-digital-economy-how-to-avoid-big-data-trap`, 2015. [Online; accessed 01-September-2021].

[8] Gert Janssenswillen, Marijke Swennen, Benoıt Depaire, Mieke Jans, and Koen Vanhoof. Enabling event-data analysis in r.

[9] Wil van der Aalst. *Process Mining: Data Science in Action*, volume 1. Springer, 2016.

[10] Reihaneh H Hariri, Erik M Fredericks, and Kate M Bowers. Uncertainty in big data analytics: survey, opportunities, and challenges. *Journal of Big Data*, 6(1):1–16, 2019.

[11] Juan Boubeta-Puig, Gregorio Díaz, Hermenegilda Macià, Valentín Valero, and Guadalupe Ortiz. Medit4cep-cpn: An approach for complex event processing modeling by prioritized colored petri nets. *Information Systems*, 81:267–289, 2019.

[12] Sandra Lusk, Staci Paley, and Andrew Spanyi. The evolution of business process management as a professional discipline. *BP Trends*, 20:1–9, 2005.

[13] Stephen George and Arnold Weimerskirch. *Total quality management: Strategies and techniques proven at today's most successful companies*. John Wiley & Sons, 1994.

[14] Mikel J Harry and Richard R Schroeder. *Six Sigma: The breakthrough management strategy revolutionizing the world's top corporations*. Crown Pub, 2005.

[15] Michael Hammer. Reengineering work: don't automate, obliterate. *Harvard business review*, 68(4):104–112, 1990.

[16] Martyn A Ould. *Business processes: modelling and analysis for re-engineering and improvement*. Wiley, 1995.

[17] Mary Sumner. *Enterprise resource planning*. Pearson Education, 2007.

[18] Howard Smith and Peter Fingar. Business process management: the third wave (vol. 1): Meghan, 2003.

[19] Naghmeh Niknejad, Waidah Ismail, Imran Ghani, Behzad Nazari, Mahadi Bahari, et al. Understanding service-oriented architecture (soa): A systematic literature review and directions for further investigation. *Information Systems*, 91:101491, 2020.

[20] Alan C Maltz, Aaron J Shenhar, and Richard R Reilly. Beyond the balanced scorecard:: Refining the search for organizational success measures. *Long range planning*, 36(2):187–204, 2003.

[21] E Mahmoodzadeh, Sh Jalalinia, and F Nekui Yazdi. A business process outsourcing framework based on business process management and knowledge management. *Business Process Management Journal*, 2009.

[22] Wil MP Van der Aalst, Martin Bichler, and Armin Heinzl. Robotic process automation, 2018.

[23] Marlon Dumas, Fabiana Fournier, Lior Limonad, Andrea Marrella, Marco Montali, Jana-Rebecca Rehse, Rafael Accorsi, Diego Calvanese, Giuseppe De Giacomo, Dirk Fahland, et al. Augmented business process management systems: A research manifesto. *arXiv preprint arXiv:2201.12855*, 2022.

[24] Stephan A Fahrenkrog-Petersen, Niek Tax, Irene Teinemaa, Marlon Dumas, Massimiliano de Leoni, Fabrizio Maria Maggi, and Matthias Weidlich. Fire now, fire later: alarm-based systems for prescriptive process monitoring. *Knowledge and Information Systems*, 64(2):559–587, 2022.

[25] Lilin Rao, Chunxiao Fan, Yuexin Wu, Xiaoying Zhang, and Hai Li. A self-adapting dynamic service management platform for internet of things. In *LISS 2013*, pages 783–791. Springer, 2015.

[26] Raef Mousheimish. *Combining the Internet of things, complex event processing, and time series classification for a proactive business process management*. PhD thesis, Université Paris-Saclay, 2017.

[27] Wil MP Van der Aalst. Business process management: a comprehensive survey. *International Scholarly Research Notices*, 2013, 2013.

[28] Rouven Poll, Artem Polyvyanyy, Michael Rosemann, Maximilian Röglinger, and Lea Rupprecht. Process forecasting: Towards proactive business process management. In *International Conference on Business Process Management*, pages 496–512. Springer, 2018.

[29] Abir Ismaili-Alaoui, Khalid Benali, Karim Baïna, and Jamal Baïna. Business process instances scheduling with human resources based on event priority determination. In *International Conference on Big Data, Cloud and Applications*, pages 118–130. Springer, 2018.

[30] Monica Anastassiu, Flavia Maria Santoro, Jan Recker, and Michael Rosemann. The quest for organizational flexibility: driving changes in business processes through the identification of relevant context. *Business Process Management Journal*, 22(4):763–790, 2016.

[31] Federico Bergenti, Giovanni Caire, and Danilo Gotta. Interactive workflows with wade. In *2012 IEEE 21st International Workshop on Enabling Technologies: Infrastructure for Collaborative Enterprises*, pages 10–15. IEEE, 2012.

[32] Irene Barba, Barbara Weber, Carmelo Del Valle, and Andrés Jiménez-Ramírez. User recommendations for the optimized execution of business processes. *Data & Knowledge Engineering*, 86:61–84, 2013.

[33] Mukhammad Andri Setiawan, Shazia Sadiq, and Ryan Kirkman. Facilitating business process improvement through personalized recommendation. In *International Conference on Business Information Systems*, pages 136–147. Springer, 2011.

[34] Lila Rao, Gunjan Mansingh, and Kweku-Muata Osei-Bryson. Building ontology based knowledge maps to assist business process re-engineering. *Decision Support Systems*, 52(3):577–589, 2012.

[35] Ravendra Singh, Krist V Gernaey, and Rafiqul Gani. An ontological knowledge-based system for the selection of process monitoring and analysis tools. *Computers & chemical engineering*, 34(7):1137–1154, 2010.

[36] Mohammad Khanbabaei, Farzad Movahedi Sobhani, Mahmood Alborzi, and Reza Radfar. Developing an integrated framework for using data mining techniques and ontology concepts for process improvement. *Journal of Systems and Software*, 137:78–95, 2018.

[37] Wil Van Der Aalst, Arya Adriansyah, Ana Karla Alves De Medeiros, Franco Arcieri, Thomas Baier, Tobias Blickle, Jagadeesh Chandra Bose, Peter Van Den Brand, Ronald Brandtjen, Joos Buijs, et al. Process mining manifesto. In *International Conference on Business Process Management*, pages 169–194. Springer, 2011.

[38] Santiago Aguirre, Carlos Parra, and Jorge Alvarado. Combination of process mining and simulation techniques for business process redesign: a methodological approach. In *International Symposium on Data-Driven Process Discovery and Analysis*, pages 24–43. Springer, 2012.

[39] Sebastian Decker. Data-driven business process improvement: An illustrative case study about the impacts and success factors of business process mining, 2019.

[40] David Redlich and Wasif Gilani. Event-driven process-centric performance prediction via simulation. In *International Conference on Business Process Management*, pages 473–478. Springer, 2011.

[41] Alaaeddine Yousfi, Adrian De Freitas, Anind K Dey, and Rajaa Saidi. The use of ubiquitous computing for business process improvement. *IEEE Transactions on Services Computing*, 9(4):621–632, 2015.

[42] Alaaeddine Yousfi, Kimon Batoulis, and Mathias Weske. Achieving business process improvement via ubiquitous decision-aware business processes. *ACM Transactions on Internet Technology (TOIT)*, 19(1):1–19, 2019.

[43] Christoph Stoiber. Exploiting internet of things for business process improvement. 2021.

[44] Maciel M Queiroz, Samuel Fosso Wamba, Marcio C Machado, and Renato Telles. Smart production systems drivers for business process management improvement: An integrative framework. *Business Process Management Journal*, 2020.

[45] Sylvain Cherrier and Varun Deshpande. From bpm to iot. In *International Conference on Business Process Management*, pages 310–318. Springer, 2017.

[46] Kevin Ashton et al. That 'internet of things' thing. *RFID journal*, 22(7):97–114, 2009.

[47] Anthony Furness. A framework model for the internet of things. *GRIFS/CASAGRAS: Hong Kong, China*, 2008.

[48] Jeremy Rifkin. *The zero marginal cost society: The internet of things, the collaborative commons, and the eclipse of capitalism.* St. Martin's Press, 2014.

[49] Enrico Natalizio. *Mobility as a Primitive to Improve Communications in Wireless Networks.* PhD thesis, UTC Compiègne, 2017.

[50] Robert Wehlitz, Ingo Rößner, and Bogdan Franczyk. Integrating smart devices as business process resources–concept and software prototype. In *International Conference on Service-Oriented Computing*, pages 252–257. Springer, 2017.

[51] Marlon Dumas, Wil M Van der Aalst, and Arthur H Ter Hofstede. *Process-aware information systems: bridging people and software through process technology*. John Wiley & Sons, 2005.

[52] Diimitrios Georgakopoulos, Mark Hornick, and Amit Sheth. An overview of workflow management: From process modeling to workflow automation infrastructure. *Distributed and parallel Databases*, 3(2):119–153, 1995.

[53] Christoph Stoiber and Stefan Schönig. Event-driven business process management enhancing iot–a systematic literature review and development of research agendas. 2021.

[54] Imen Graja, Slim Kallel, Nawal Guermouche, and Ahmed Hadj Kacem. Bpmn4cps: A bpmn extension for modeling cyber-physical systems. In *2016 IEEE 25th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE)*, pages 152–157. IEEE, 2016.

[55] Sonja Meyer, Andreas Ruppen, and Carsten Magerkurth. Internet of things-aware process modeling: integrating iot devices as business process resources. In *International conference on advanced information systems engineering*, pages 84–98. Springer, 2013.

[56] Stefan Schönig, Lars Ackermann, Stefan Jablonski, and Andreas Ermer. An integrated architecture for iot-aware business process execution. In *Enterprise, business-process and information systems modeling*, pages 19–34. Springer, 2018.

[57] Abir Ismaili-Alaoui, Karim Baïna, Khalid Benali, and Jamal Baïna. Towards Smart Incident Management Under Human Resource Constraints for an IoT-BPM Hybrid Architecture. In *International Conference on Web Services*, pages 457–471. Springer, 2018.

[58] Stefan Schönig, Lars Ackermann, and Stefan Jablonski. Internet of things meets bpm: A conceptual integration framework. In *SIMULTECH*, pages 307–314, 2018.

[59] Zakaria Maamar, Ejub Kajan, Ikbel Guidara, Leyla Moctar-M'Baba, and Mohamed Sellami. Bridging the gap between business processes and iot. In *Proceedings of the 24th Symposium on International Database Engineering & Applications*, pages 1–10, 2020.

[60] Christian Friedow, Maximilian Völker, and Marcin Hewelt. Integrating iot devices into business processes. In *International conference on advanced information systems engineering*, pages 265–277. Springer, 2018.

[61] Rongjia Song, Ying Wang, Weiping Cui, Jan Vanthienen, and Lei Huang. Towards improving context interpretation in the iot paradigm: a solution to integrate context information in process models. In *Proceedings of the 2018 2nd International Conference on Management Engineering, Software Engineering and Service Sciences*, pages 223–228. ACM, 2018.

[62] Stefan Schönig, Lars Ackermann, Stefan Jablonski, and Andreas Ermer. Iot meets bpm: a bidirectional communication architecture for iot-aware process execution. *Software and Systems Modeling*, 19(6):1443–1459, 2020.

[63] Christian Sturm, Stefan Schönig, and Claudio Di Ciccio. Distributed multi-perspective declare discovery. In *BPM (Demos)*, 2017.

[64] Dave Evans. The internet of things: How the next evolution of the internet is changing everything. *CISCO white paper*, 1(2011):1–11, 2011.

[65] Zahrah A Almusaylim and Noor Zaman. A review on smart home present state and challenges: linked to context-awareness internet of things (iot). *Wireless networks*, 25(6):3193–3204, 2019.

[66] Carlos Soto. Bpm and iot: The signal of true disruption. *Redmond Media Group*, 2015.

[67] Simon Tjoa, Stefan Jakoubi, Gernot Goluch, Gerhard Kitzler, Sigrun Goluch, and Gerald Quirchmayr. A formal approach enabling risk-aware business process modeling and simulation. *IEEE Transactions on Services Computing*, 4(2):153–166, 2011.

[68] M. Rijmenam. Why the 3v's are not sufficient to describe big data - datafloq. `https://datafloq.com/read/3vs-sufficient-describe-big-data/166`. [Online; accessed July-2021].

[69] Chii Chang, Satish Narayana Srirama, and Rajkumar Buyya. Mobile cloud business process management system for the internet of things: a survey. *ACM Computing Surveys (CSUR)*, 49(4):70, 2017.

[70] Jayavardhana Gubbi, Rajkumar Buyya, Slaven Marusic, and Marimuthu Palaniswami. Internet of things (iot): A vision, architectural elements, and future directions. *Future generation computer systems*, 29(7):1645–1660, 2013.

[71] Daniela Grigori. *Eléments de flexibilité des systèmes de workflow pour la définition et l'exécution de procédés coopératifs*. PhD thesis, Université Henri Poincaré-Nancy 1, 2001.

[72] HJ Harrington. P. american society for quality control quality, business process improvement: the breakthrough strategy for total quality, productivity, and competitiveness, 1991.

[73] Karim Baïna, Khalid Benali, and Claude Godart. DISCOBOLE: A service architecture for interconnecting workflow processes. *Comput. Ind.*, 57(8-9):768–777, 2006.

[74] Wil MP van Der Aalst. *Workflow patterns. Encyclopedia of Database Systems, pages 3557–3558*. 2009.

[75] Ruopeng Lu and Shazia Sadiq. A survey of comparative business process modeling approaches. In *International conference on business information systems*, pages 82–94. Springer, 2007.

[76] Ryan KL Ko, Stephen SG Lee, and Eng Wah Lee. Business process management (bpm) standards: a survey. *Business Process Management Journal*, 2009.

[77] OMG. v2.3 omg. unified modelling language.

[78] OMG. business process model and notation (bpmn) 2.0.

[79] Wil MP Van Der Aalst and Arthur HM Ter Hofstede. Yawl: yet another workflow language. *Information systems*, 30(4):245–275, 2005.

[80] August-Wilhelm Scheer. *ARIS—business process modeling*. Springer Science & Business Media, 2000.

[81] Boudewijn F van Dongen, Wil MP Van der Aalst, and Henricus MW Verbeek. Verification of epcs: Using reduction rules and petri nets. In *International Conference on Advanced Information Systems Engineering*, pages 372–386. Springer, 2005.

[82] Richard J Mayer, Christopher P Menzel, Michael K Painter, Paula S Dewitte, Thomas Blinn, and Benjamin Perakath. Information integration for concurrent engineering (iice) idef3 process description capture method report. Technical report, KNOWLEDGE BASED SYSTEMS INC COLLEGE STATION TX, 1995.

[83] David Hollingsworth and UK Hampshire. Workflow management coalition: The workflow reference model. *Document Number TC00-1003*, 19(16):224, 1995.

[84] Stephen A White. Introduction to bpmn. *Ibm Cooperation*, 2(0):0, 2004.

[85] Michael Zur Muehlen and Jan Recker. How much language is enough? theoretical and practical use of the business process modeling notation. In *Seminal Contributions to Information Systems Engineering*, pages 429–443. Springer, 2013.

[86] Joseph M Juran and A Blanton Godfrey. *The quality control process*. McGraw-Hill, 1999.

[87] Joachim Van den Bergh, Sara Thijs, and Stijn Viaene. *Transforming Through Processes: Leading Voices on BPM, People and Technology*. Springer, 2014.

[88] Florian Forster. The idea behind business process improvement: toward a business process improvement pattern framework. *BP Trends*, 2006(April):1–14, 2006.

[89] H James Harrington. *Business process improvement: The breakthrough strategy for total quality, productivity, and competitiveness*. McGraw Hill Professional, 1991.

[90] H James Harrington, Erik KC Esseling, and Harm Van Nimwegen. *Business process improvement workbook: documentation, analysis, design, and management of business process improvement*. McGraw Hill Professional, 1997.

[91] V Daniel Hunt. *Process mapping: how to reengineer your business processes*. John Wiley & Sons, 1996.

[92] Ben Graham. The roots of the business process mappingi. *BP Trends*, pages 1–6, 2008.

[93] Joe Paradiso. The essential process: the quick and painless path to successful process mapping. *Industrial engineer*, 35(4):46–49, 2003.

[94] Bill Curtis, Marc I Kellner, and Jim Over. Process modeling. *Communications of the ACM*, 35(9):75–90, 1992.

[95] Leidy Klotz, Michael Horman, Henry H Bi, and John Bechtel. The impact of process mapping on transparency. *International Journal of Productivity and Performance Management*, 2008.

[96] Matthias Holweg. The genealogy of lean production. *Journal of operations management*, 25(2):420–437, 2007.

[97] James Womak, Daniel T Jones, and Daniel Roos. The machine that changed the world. *New York: Rawson Associates*, 1990.

[98] Richard J Schonberger. Japanese production management: An evolution—with mixed success. *Journal of Operations Management*, 25(2):403–419, 2007.

[99] James P Womack and Daniel T Jones. Lean thinking—banish waste and create wealth in your corporation. *Journal of the Operational Research Society*, 48(11):1148–1148, 1997.

[100] Javier Santos, Richard A Wysk, and Jose M Torres. *Improving production with lean thinking*. John Wiley & Sons, 2014.

[101] Natalie J Sayer and Bruce Williams. *Lean for dummies*. John Wiley & Sons, 2012.

[102] Michael Dudbridge. *Handbook of lean manufacturing in the food industry*. John Wiley & Sons, 2011.

[103] W Edwards Deming. *Out of the Crisis*. MIT press, 2018.

[104] Colin Morgan and Stephen Murgatroyd. *Total quality management in the public sector: An international perspective.* McGraw-Hill Education (UK), 1994.

[105] Mohamed Zairi. *Total quality management for engineers.* Elsevier, 1991.

[106] James R Evans. Total quality management. *Infor*, 40(4):364, 2002.

[107] Edward Sallis. *Total quality management in education.* Routledge, 2014.

[108] James F Chang. *Business process management systems: strategy and implementation.* Auerbach Publications, 2016.

[109] Bhimaraya A Metri. Total quality transportation through deming's 14 points. *Journal of Public Transportation*, 9(4):3, 2006.

[110] Euclides A Coimbra. *Kaizen in logistics and supply chains.* McGraw-Hill Education New York, NY, 2013.

[111] Sami Al Smadi. Kaizen strategy and the drive for competitiveness: challenges and opportunities. *Competitiveness Review: An International Business Journal*, 2009.

[112] Larisa Dragolea and Denisa Cotîrlea. Benchmarking-a valid strategy for the long term? *Annales Universitatis Apulensis: Series Oeconomica*, 11(2):813, 2009.

[113] R Cross and A Iqbal. The rank xerox experience: benchmarking ten years on. In *Benchmarking—Theory and practice*, pages 3–10. Springer, 1995.

[114] Arthur Thompson, Margaret Peteraf, John Gamble, Alonzo J Strickland III, and Arun Kumar Jain. *Crafting & executing strategy 19/e: The quest for competitive advantage: Concepts and cases.* McGraw-Hill Education, 2013.

[115] Robert C Camp, Robert C Camp, et al. *Business process benchmarking: finding and implementing best practices.* Asq Press, 1995.

[116] Tim Stapenhurst. *The benchmarking book.* Routledge, 2009.

[117] Asbjørn Rolstadås. *Performance management: A business process benchmarking approach.* Springer Science & Business Media, 2012.

[118] Graeme Knowles. *Profit From Six Sigma.* Bookboon, 2012.

[119] K Muralidharan. Six sigma for organizational excellence. *Springer Proceedings of the Institution of Mechanical Engineers*, 203(B1):43–50, 2015.

[120] Thomas Pyzdek and Paul Keller. *Six sigma handbook.* McGraw-Hill Education, 2014.

[121] Michael George. *Lean Six Sigma pour les services.* Maxima, 2013.

[122] Zeynep Tuğçe Şimşit, Noyan Sebla Günay, and Özalp Vayvay. Theory of constraints: A literature review. *Procedia-Social and Behavioral Sciences*, 150:930–936, 2014.

[123] Eliyahu M Goldratt and Jeff Cox. *The goal: a process of ongoing improvement.* Routledge, 2016.

[124] Shams-ur Rahman. Theory of constraints: a review of the philosophy and its applications. *International Journal of Operations & Production Management*, 1998.

[125] Mark Woeppel. *Manufacturer's guide to implementing the theory of constraints.* CRC Press, 2000.

[126] Thomas B McMullen Jr. *Introduction to the theory of constraints (TOC) management system.* CRC Press, 1998.

[127] James F Cox III, CIRM CFPIM, and John G Schleier Jr. *Theory of constraints handbook.* McGraw-Hill Education, 2010.

[128] Kevin J Watson, John H Blackstone, and Stanley C Gardiner. The evolution of a management philosophy: The theory of constraints. *Journal of operations Management*, 25(2):387–402, 2007.

[129] David Hoyle. *Automotive quality systems handbook.* Elsevier, 2000.

[130] Jaideep Motwani, Ashok Kumar, and Chun Hung Cheng. A roadmap to implementing iso 9000. *International Journal of Quality & Reliability Management*, 1996.

[131] Hung W Chung. *Understanding quality assurance in construction: a practical guide to ISO 9000 for contractors.* Routledge, 2002.

[132] David Hoyle. *ISO 9000: 2000: an AZ guide.* Routledge, 2012.

[133] Bill Self, Greg Roche, and Nigel Hill. *Customer satisfaction measurement for ISO 9000: 2000.* Routledge, 2007.

[134] Alex Douglas, Shirley Coleman, and Richard Oddy. The case for iso 9000. *The TQM magazine*, 2003.

[135] Perry Lawrence Johnson. *ISO 9000: Meeting the new international standards.* McGraw-Hill New York, 1993.

[136] Samia M Siha and Germaine H Saad. Business process improvement: empirical assessment and extensions. *Business Process Management Journal*, 2008.

[137] Sanjay Mohapatra. Business process reengineering: a consolidated approach to different models. In *Business Process Reengineering*, pages 1–38. Springer, 2013.

[138] Te Fu Chen. *Implementing New Business Models in For-Profit and Non-Profit Organizations: Technologies and Applications: Technologies and Applications.* IGI Global, 2010.

[139] Paul Harmon. *Business process change: a manager's guide to improving, redesigning, and automating processes.* Morgan Kaufmann, 2003.

[140] A Tanswell. Business restructuring-the key to radical change. *Professional Engineering*, 6:24–24, 1993.

[141] Steven Gross, Monika Malinova, and Jan Mendling. Navigating through the maze of business process change methods. In *Proceedings of the 52nd Hawaii international conference on system sciences*, 2019.

[142] Thomas H Davenport, James E Short, et al. The new industrial engineering: information technology and business process redesign. 1990.

[143] Varun Grover and William J Kettinger. *Business process change: Concepts, methods, and technologies.* IGI Global, 1995.

[144] Marlon Dumas, Marcello La Rosa, Jan Mendling, Hajo A Reijers, et al. *Fundamentals of business process management*, volume 1. Springer, 2013.

[145] Paul Savory and John Olson. Guidelines for using process mapping to aid improvement efforts. *Industrial and Management Systems Engineering Faculty Publications*, page 58, 2001.

[146] Abby Ghobadian and Simon Speller. Gurus of quality: a framework for comparison. *Total Quality Management*, 5(3):53–70, 1994.

[147] Abby Ghobadian and David N Gallear. Total quality management in smes. *Omega*, 24(1):83–106, 1996.

[148] Michal Marton and Iveta Paulová. Applying the theory of constraints in the course of process improvement. *Vedecké Práce Materiálovotechnologickej Fakulty Slovenskej Technickej Univerzity v Bratislave so Sídlom v Trnave*, 18(29):71, 2010.

[149] Fadwa Yahya, Khouloud Boukadi, and Hanêne Ben-Abdallah. Improving the quality of business process models: Lesson learned from the state of the art. *Business Process Management Journal*, 2019.

[150] Samira Si-Said Cherfi, Sarah Ayad, and Isabelle Comyn-Wattiau. Improving business process model quality using domain ontologies. *Journal on Data Semantics*, 2(2-3):75–87, 2013.

[151] Matthias Born, Christian Brelage, Ivan Markovic, Daniel Pfeiffer, and Ingo Weber. Auto-completion for executable business process models. In *International Conference on Business Process Management*, pages 510–515. Springer, 2008.

[152] Sarah Ayad. *Business Process Models Quality: Evaluation and Improvement*. PhD thesis, Paris, CNAM, 2013.

[153] Jörg Becker, Michael Rosemann, and Christoph Von Uthmann. Guidelines of business process modeling. In *Business process management*, pages 30–49. Springer, 2000.

[154] Jan Mendling, Hajo A Reijers, and Wil MP van der Aalst. Seven process modeling guidelines (7pmg). *Information and Software Technology*, 52(2):127–136, 2010.

[155] Henrik Leopold, Jan Mendling, and Oliver Günther. Learning from quality issues of bpmn models from industry. *IEEE software*, 33(4):26–33, 2015.

[156] Ana Ivanchikj, Vincenzo Ferme, and Cesare Pautasso. On the performance overhead of bpmn modeling practices. In *International conference on business process management*, pages 216–232. Springer, 2017.

[157] Wiem Khlif, Nourchène Elleuch Ben Ayed, and Hanêne Ben-Abdallah. Evares: A quality-driven refactoring method for business process models. In *ICEIS (3)*, pages 409–416, 2017.

[158] María Fernández-Ropero, Ricardo Pérez-Castillo, Ismael Caballero, and Mario Piattini. Quality-driven business process refactoring. In *International Conference on Business Information Systems (ICBIS 2012)*, pages 960–966. Citeseer, 2012.

[159] Barbara Weber, Manfred Reichert, Jan Mendling, and Hajo A Reijers. Refactoring large process model repositories. *Computers in industry*, 62(5):467–486, 2011.

[160] Barbara Weber and Manfred Reichert. Refactoring process models in large process repositories. In *International Conference on Advanced Information Systems Engineering*, pages 124–139. Springer, 2008.

[161] Marcello La Rosa, Arthur HM Ter Hofstede, Petia Wohed, Hajo A Reijers, Jan Mendling, and Wil MP Van der Aalst. Managing process model complexity via concrete syntax modifications. *IEEE Transactions on Industrial Informatics*, 7(2):255–265, 2011.

[162] Batini Carlo, Stefano Ceri, and Navathe Sham. *Conceptual database design: an entity-relationship approach*. Benjamin/Cummings, 1992.

[163] Sophie Dupuy-Chessa, Kathia Marçal de Oliveira, and Samira Si-Said Cherfi. Qualité des modèles: retours d'expériences. In *32ème congrès Inforsid'2014*, pages 363–378, 2014.

[164] Odd Ivar Lindland, Guttorm Sindre, and Arne Solvberg. Understanding quality in conceptual modeling. *IEEE software*, 11(2):42–49, 1994.

[165] John Krogstie. Quality of business process models. In *Quality in Business Process Modeling*, pages 53–102. Springer, 2016.

[166] Djida Bahloul. *Une approche hybride de gestion des connaissances basée sur les ontologies*. PhD thesis, Université Paul Sabatier, 2006.

[167] Tom Gruber. Ontology., 2018.

[168] William F Opdyke. Refactoring object-oriented frameworks. 1992.

[169] Martin Fowler. *Refactoring: improving the design of existing code*. Addison-Wesley Professional, 2018.

[170] Wiem Khlif, Hanene Ben-Abdallah, and Nourchène Elleuch Ben Ayed. A methodology for the semantic and structural restructuring of bpmn models. *Business Process Management Journal*, 2017.

[171] Wil MP van der Aalst, AK Alves De Medeiros, and AJMM Weijters. Process equivalence: Comparing two process models based on observed behavior. In *International conference on business process management*, pages 129–144. Springer, 2006.

[172] Daniel L. Moody and Graeme G. Shanks. What makes a good data model? a framework for evaluating and improving the quality of entity relationship models. *Australian Computer Journal*, 30(3):97–110, 1998.

[173] John Krogstie, Odd Ivar Lindland, and Guttorm Sindre. Towards a deeper understanding of quality in requirements engineering. In *International Conference on Advanced Information Systems Engineering*, pages 82–95. Springer, 1995.

[174] Carlo Batini, L Furlani, and Enrico Nardelli. What is a good diagram? a pragmatic approach. In *Proceedings of the Fourth International Conference on Entity-Relationship Approach*, pages 312–319, 1985.

[175] Ravi Anupindi, Sunil Chopra, Sudhakar D Deshmukh, Jan A Van Mieghem, and Eitan Zemel. *Managing business process flows*, volume 400. Prentice Hall Upper Saddle River, NJ, 1999.

[176] David A Bishop. Key performance indicators: Ideation to creation. *IEEE Engineering Management Review*, 46(1):13–15, 2018.

[177] Avraham Shtub and Reuven Karni. *ERP: the dynamics of supply chain and process management*. Springer Science & Business Media, 2010.

[178] Selma Limam Mansar and Hajo A Reijers. Best practices in business process redesign: validation of a redesign framework. *Computers in industry*, 56(5):457–471, 2005.

[179] SW Cranenbroek. Changing business processes: changing the business process of renting out and reselling of accommodation units at de meeuw bouwsystemen bv. 2009.

[180] MH Jansen-Vullers, MWNC Loosschilder, PAM Kleingeld, and HA Reijers. Performance measures to evaluate the impact of best practices. In *Proceedings of Workshops and Doctoral Consortium of the 19th International Conference on Advanced Information Systems Engineering (BPMDS workshop)*, volume 1, pages 359–368. Tapir Academic Press Trondheim, 2007.

[181] John Martin. *The best practice of business*. J. Martin, 1978.

[182] Jonathan Golovin. *Achieving stretch goals: best practices in manufacturing for the new millennium*. Prentice Hall, 1997.

[183] PJArttEP Butler. A strategic framework for health promotion in darebin. *A report to the East Preston and Northcote community health centers by the center for development and innovation in health, center for development and innovation in health, Melbourne, Australia*, 1996.

[184] H. Van der Kolk N. Brand. *Workflow Analysis and Design*. Kluwer Bedrijfswetenschappen, Deventer, (in Dutch), 1995.

[185] Alaaeddine Yousfi. *Smart business process improvement for process aware information systems*. PhD thesis, 2016.

[186] Gordon Masiuk. Developing and using process performance metrics", calagary community of business process management professionals. `http://fr.slideshare.net/gmasiuk/june-21-2012-process-performance-metrics-presentation`.

[187] Susan Page. *The power of business process improvement: 10 simple steps to increase effectiveness, efficiency, and adaptability*. AMACOM Div American Mgmt Assn, 2015.

[188] Cengiz Kahraman and Seda Yanık. *Intelligent Decision Making in Quality Management: Theory and Applications*, volume 97. Springer, 2015.

[189] Gordon Masiuk. businessdictionary. `http://www.businessdictionary.com/`.

[190] Jan Mendling, Gero Decker, Richard Hull, Hajo A Reijers, and Ingo Weber. How do machine learning, robotic process automation, and blockchains affect the human factor in business process management? *Communications of the Association for Information Systems*, 43(1):19, 2018.

[191] Wolfgang Kratsch, Jonas Manderscheid, Maximilian Röglinger, and Johannes Seyfried. Machine learning in business process monitoring: a comparison of deep learning and classical approaches used for outcome prediction. *Business & Information Systems Engineering*, pages 1–16, 2020.

[192] Nijat Mehdiyev, Andreas Emrich, Björn P Stahmer, Peter Fettke, and Peter Loos. iprodict-intelligent process prediction based on big data analytics. In *BPM (Industry Track)*, pages 13–24, 2017.

[193] Michael Arias, Rodrigo Saavedra, Maira R Marques, Jorge Munoz-Gama, and Marcos Sepúlveda. Human resource allocation in business process management and process mining: A systematic mapping study. *Management Decision*, 2018.

[194] Dusanka Dakic, Darko Stefanovic, Ilija Cosic, Teodora Lolic, and Milovan Medojevic. Business process mining application: A literature review. *Annals of DAAAM & Proceedings*, 29, 2018.

[195] Mathias Kirchmer. Robotic process automation-pragmatic solution or dangerous illusion. *BTOES Insights, June*, 17, 2017.

[196] Tathagata Chakraborti, Vatche Isahagian, Rania Khalaf, Yasaman Khazaeni, Vinod Muthusamy, Yara Rizk, and Merve Unuvar. From robotic process automation to intelligent process automation. In *International Conference on Business Process Management*, pages 215–228. Springer, 2020.

[197] A Th Schreiber, Guus Schreiber, Hans Akkermans, Anjo Anjewierden, Nigel Shadbolt, Robert de Hoog, Walter Van de Velde, Bob Wielinga, R Nigel, et al. *Knowledge engineering and management: the CommonKADS methodology.* MIT press, 2000.

[198] M. Rijmenam. Why the 3V's Are Not Sufficient To Describe Big Data. `https://datafloq.com/read/3vs-sufficient-describe-big-data/166`, 2013. [Online; accessed 01-September-2021].

[199] William J Frawley, Gregory Piatetsky-Shapiro, and Christopher J Matheus. Knowledge discovery in databases: An overview. *AI magazine*, 13(3):57–57, 1992.

[200] Wil MP Van der Aalst. Data scientist: The engineer of the future. In *Enterprise interoperability VI*, pages 13–26. Springer, 2014.

[201] Philip Russom et al. Big data analytics. *TDWI best practices report, fourth quarter*, 19(4):1–34, 2011.

[202] Steven Engelhard. Visualizing Latency Part 3: Rendering Event Data. `https://www.stevenengelhardt.com/2017/12/01/visualizing-latency-part-3-rendering-event-data/`, 2017. [Online; accessed 01-September-2021].

[203] Rehan Syed, Suriadi Suriadi, Michael Adams, Wasana Bandara, Sander JJ Leemans, Chun Ouyang, Arthur HM ter Hofstede, Inge van de Weerd, Moe Thandar Wynn, and Hajo A Reijers. Robotic process automation: contemporary themes and challenges. *Computers in Industry*, 115:103162, 2020.

[204] Fredrik Milani, Luciano García-Bañuelos, and Marlon Dumas. Blockchain and business process improvement. *BPTrends newsletter (October 2016)*, 2016.

[205] Wattana Viriyasitavat, Li Da Xu, Zhuming Bi, and Assadaporn Sapsomboon. Blockchain-based business process management (bpm) framework for service composition in industry 4.0. *Journal of Intelligent Manufacturing*, 31(7):1737–1748, 2020.

[206] Le portail des ministeres economiques et financier. Participez à l'enqute sur la silver economie. `https://www.economie.gouv.fr/participez-a-lenquete-acteurs-activites-et-perspectives-a-cinq-ans-silver-economie`, 2013. [Online; accessed 19-February-2021].

[207] Charles Møller, Carsten J Maack, and Rune D Tan. What is business process management: A two stage literature review of an emerging field. *Research and practical issues of enterprise information systems II*, pages 19–31, 2007.

[208] Jiajie Xu, Chengfei Liu, and Xiaohui Zhao. Resource allocation vs. business process improvement: How they impact on each other. In *International Conference on Business Process Management*, pages 228–243. Springer, 2008.

[209] Kahina Bessai. *Gestion optimale de l'allocation des ressources pour l'ex ecution des processus dans le cadre du Cloud.* PhD thesis, Université Paris1 Panthéon-Sorbonne, 2014.

[210] R MKAOUAR HACHICHA, EM DAFAOUI, and A EL MHAMEDI. Assignment problem under competences and preferences constraints: Modelling and resolution. *IFAC Proceedings Volumes*, 45(6):1170–1176, 2012.

[211] A Volgenant. A note on the assignment problem with seniority and job priority constraints. *European journal of operational research*, 154(1):330–335, 2004.

[212] Kahina Bessai and François Charoy. Business process tasks-assignment and resource allocation in crowdsourcing context. In *2016 IEEE 2nd International Conference on Collaboration and Internet Computing (CIC)*, pages 11–18. IEEE, 2016.

[213] Luiz Fernando Bittencourt and Edmundo Roberto Mauro Madeira. Hcoc: a cost optimization algorithm for workflow scheduling in hybrid clouds. *Journal of Internet Services and Applications*, 2(3):207–227, 2011.

[214] Weiwei Chen, Rafael Ferreira Da Silva, Ewa Deelman, and Rizos Sakellariou. Using imbalance metrics to optimize task clustering in scientific workflow executions. *Future Generation Computer Systems*, 46:69–84, 2015.

[215] WenAn Tan, Yong Sun, Ling Xia Li, GuangZhen Lu, and Tong Wang. A trust service-oriented scheduling model for workflow applications in cloud computing. *IEEE Systems Journal*, 8(3):868–878, 2013.

[216] Deepak Poola, Kotagiri Ramamohanarao, and Rajkumar Buyya. Enhancing reliability of workflow execution using task replication and spot instances. *ACM Transactions on Autonomous and Adaptive Systems (TAAS)*, 10(4):1–21, 2016.

[217] Sonia Yassa, Rachid Chelouah, Hubert Kadima, and Bertrand Granado. Multi-objective approach for energy-aware workflow scheduling in cloud computing environments. *The Scientific World Journal*, 2013, 2013.

[218] Wei-Jen Wang, Yue-Shan Chang, Win-Tsung Lo, and Yi-Kang Lee. Adaptive scheduling for parallel tasks with qos satisfaction for hybrid cloud environments. *The Journal of Supercomputing*, 66(2):783–811, 2013.

[219] Nidhi Rehani and Ritu Garg. Reliability-aware workflow scheduling using monte carlo failure estimation in cloud. In *Proceedings of International Conference on Communication and Networks*, pages 139–153. Springer, 2017.

[220] Madhu Sudan Kumar, Indrajeet Gupta, and Prasanta K Jana. Forward load aware scheduling for data-intensive workflow applications in cloud system. In *2016 International Conference on Information Technology (ICIT)*, pages 93–98. IEEE, 2016.

[221] SH Niu, SK Ong, and AYC Nee. An improved intelligent water drops algorithm for achieving optimal job-shop scheduling solutions. *International Journal of Production Research*, 50(15):4192–4205, 2012.

[222] Jiajie Xu, Chengfei Liu, and Xiaohui Zhao. Resource planning for massive number of process instances. In *OTM Confederated International Conferences" On the Move to Meaningful Internet Systems"*, pages 219–236. Springer, 2009.

[223] Zahra Momenzadeh and Faramarz Safi-Esfahani. Workflow scheduling applying adaptable and dynamic fragmentation (wsadf) based on runtime conditions in cloud computing. *Future Generation Computer Systems*, 90:327–346, 2019.

[224] Hamid Arabnejad and Jorge Barbosa. Fairness resource sharing for dynamic workflow scheduling on heterogeneous systems. In *2012 IEEE 10th International Symposium on Parallel and Distributed Processing with Applications*, pages 633–639. IEEE, 2012.

[225] Wil MP van der Aalst and Kees M van Hee. Business process redesign: A petri-net-based approach. *Computers in industry*, 29(1-2):15–26, 1996.

[226] Maciej Malawski, Kamil Figiela, Marian Bubak, Ewa Deelman, and Jarek Nabrzyski. Scheduling multilevel deadline-constrained scientific workflows on clouds based on cost optimization. *Scientific Programming*, 2015, 2015.

[227] Li Liu, Miao Zhang, Rajkumar Buyya, and Qi Fan. Deadline-constrained coevolutionary genetic algorithm for scientific workflow scheduling in cloud computing. *Concurrency and Computation: Practice and Experience*, 29(5):e3942, 2017.

[228] Guoqi Xie, Gang Zeng, Yuekun Chen, Yang Bai, Zhili Zhou, Renfa Li, and Keqin Li. Minimizing redundancy to satisfy reliability requirement for a parallel application on heterogeneous service-oriented systems. *IEEE Transactions on Services Computing*, 2017.

[229] Zhongjin Li, Jidong Ge, Hongji Yang, Liguo Huang, Haiyang Hu, Hao Hu, and Bin Luo. A security and cost aware scheduling algorithm for heterogeneous tasks of scientific workflow in clouds. *Future Generation Computer Systems*, 65:140–152, 2016.

[230] Andrés García García, Ignacio Blanquer Espert, and Vicente Hernández García. Sla-driven dynamic cloud resource management. *Future Generation Computer Systems*, 31:1–11, 2014.

[231] Saurabh Kumar Garg, Adel Nadjaran Toosi, Srinivasa K Gopalaiyengar, and Rajkumar Buyya. Sla-based virtual machine management for heterogeneous workloads in a cloud datacenter. *Journal of Network and Computer Applications*, 45:108–120, 2014.

[232] Somayeh Kianpisheh, Nasrolah Moghadam Charkari, and Mehdi Kargahi. Reliability-driven scheduling of time/cost-constrained grid workflows. *Future Generation Computer Systems*, 55:1–16, 2016.

[233] Daniel Cosmin Porumbel. *Algorithmes Heuristiques et Techniques d'Apprentissage-Applications au Probleme de Coloration de Graphe*. PhD thesis, Université d'Angers, 2009.

[234] Johan Tordsson, Rubén S Montero, Rafael Moreno-Vozmediano, and Ignacio M Llorente. Cloud brokering mechanisms for optimized placement of virtual machines across multiple providers. *Future generation computer systems*, 28(2):358–367, 2012.

[235] Suraj Pandey, Linlin Wu, Siddeswara Mayura Guru, and Rajkumar Buyya. A particle swarm optimization-based heuristic for scheduling workflow applications in cloud computing environments. In *2010 24th IEEE international conference on advanced information networking and applications*, pages 400–407. IEEE, 2010.

[236] R Sathish Kumar and S Gunasekaran. Improving task scheduling in large scale cloud computing environment using artificial bee colony algorithm. *International Journal of Computer Applications*, 103(5), 2014.

[237] Pardeep Kumar and Amandeep Verma. Independent task scheduling in cloud computing by improved genetic algorithm. *International Journal of Advanced Research in Computer Science and Software Engineering*, 2(5), 2012.

[238] Jyoti Thaman and Manpreet Singh. Current perspective in task scheduling techniques in cloud computing: A review. *International Journal in Foundations of Computer Science & Technology*, 6(1):65–85, 2016.

[239] Sana Bouajaja and Najoua Dridi. A survey on human resource allocation problem and its applications. *Operational Research*, 17(2):339–369, 2017.

[240] Irfan Younas, Farzad Kamrani, Christian Schulte, and Rassul Ayani. Optimization of task assignment to collaborating agents. In *2011 IEEE Symposium on Computational Intelligence in Scheduling (SCIS)*, pages 17–24. IEEE, 2011.

[241] Walter J Gutjahr and Marion S Rauner. An aco algorithm for a dynamic regional nurse-scheduling problem in austria. *Computers & Operations Research*, 34(3):642–666, 2007.

[242] Cristina Cabanillas, José María García, Manuel Resinas, David Ruiz, Jan Mendling, and Antonio Ruiz-Cortés. Priority-based human resource allocation in business processes. In *International Conference on Service-Oriented Computing*, pages 374–388. Springer, 2013.

[243] Jason Bell. *Machine learning: hands-on for developers and technical professionals*. John Wiley & Sons, 2020.

[244] Jason Brownlee. Supervised and unsupervised machine learning algorithms. *Machine Learning Mastery*, 16(03), 2016.

[245] Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction*. MIT press, 2018.

[246] Taiwo Oladipupo Ayodele. Types of machine learning algorithms. *New advances in machine learning*, 3:19–48, 2010.

[247] Francesco Ricci, Lior Rokach, and Bracha Shapira. Introduction to recommender systems handbook. In *Recommender systems handbook*, pages 1–35. Springer, 2011.

[248] Rui Xu and Donald Wunsch. Survey of clustering algorithms. *IEEE Transactions on neural networks*, 16(3):645–678, 2005.

[249] Dongkuan Xu and Yingjie Tian. A comprehensive survey of clustering algorithms. *Annals of Data Science*, 2(2):165–193, 2015.

[250] Gerardo Beni and Jing Wang. Swarm intelligence in cellular robotic systems. In *Robots and biological systems: towards a new bionics?*, pages 703–712. Springer, 1993.

[251] Amrita Chakraborty and Arpan Kumar Kar. Swarm intelligence: A review of algorithms. *Nature-Inspired Computing and Optimization*, pages 475–494, 2017.

[252] Michalis Mavrovouniotis, Changhe Li, and Shengxiang Yang. A survey of swarm intelligence for dynamic optimization: Algorithms and applications. *Swarm and Evolutionary Computation*, 33:1–17, 2017.

[253] Mohd Nadhir Ab Wahab, Samia Nefti-Meziani, and Adham Atyabi. A comprehensive review of swarm optimization algorithms. *PloS one*, 10(5):e0122827, 2015.

[254] Hajo A Reijers, Monique H Jansen-Vullers, Michael Zur Muehlen, and Winfried Appl. Workflow management systems+ swarm intelligence= dynamic task assignment for emergency management applications. In *International Conference on Business Process Management*, pages 125–140. Springer, 2007.

[255] Kahina Bessai, Samir Youcef, Ammar Oulamara, Claude Godart, and Selmin Nurcan. Business process scheduling strategies in cloud environments with fairness metrics. In *2013 IEEE International Conference on Services Computing*, pages 519–526. IEEE, 2013.

[256] Michael Arias, Eric Rojas, Jorge Munoz-Gama, and Marcos Sepúlveda. A framework for recommending resource allocation based on process mining. In *International Conference on Business Process Management*, pages 458–470. Springer, 2016.

[257] Jean L Pierobom, Myriam R Delgado, and Celso AA Kaestner. Particle swarm optimization applied to task assignment problem. In *10th Brazilian Congress on Computational Intelligence*, 2011.

[258] Guillaume Rosinosky, Samir Youcef, and Francois Charoy. A genetic algorithm for cost-aware business processes execution in the cloud. In *International Conference on Service-Oriented Computing*, pages 198–212. Springer, 2018.

[259] Dinabandhu Bhandari, CA Murthy, and Sankar K Pal. Variance as a stopping criterion for genetic algorithms with elitist model. *Fundamenta Informaticae*, 120(2):145–164, 2012.

[260] Le portail des ministeres economiques et financier. "participez a l'enquete sur la silver economie". `https://www.economie.gouv.fr/participez-a-lenquete-acteurs-activites-et-perspectives-a-cinq-ans-silver-economie`. [Online; accessed July-2021].

[261] Kimberly A Faulkner, Mark S Redfern, Jane A Cauley, Douglas P Landsittel, Stephanie A Studenski, Caterina Rosano, Eleanor M Simonsick, Tamara B Harris, Ronald I Shorr, Hilsa N Ayonayon, et al. Multitasking: association between poorer performance and a history of recurrent falls. *Journal of the American Geriatrics Society*, 55(4):570–576, 2007.

[262] Paul A Stalenhoef, Jos PM Diederiks, Luc P de Witte, Karin H Schiricke, and Harry FJM Crebolder. Impact of gait problems and falls on functioning in independent living persons of 55 years and over: a community survey. *Patient Education and Counseling*, 36(1):23–31, 1999.

[263] Paul A Stalenhoef, Jos PM Diederiks, J André Knottnerus, Luc P de Witte, and Harry FJM Crebolder. The construction of a patient record-based risk model for recurrent falls among elderly people living in the community. *Family Practice*, 17(6):490–496, 2000.

[264] PA Stalenhoef, JPM Diederiks, JA Knottnerus, ADM Kester, and HFJM Crebolder. A risk model for the prediction of recurrent falls in community-dwelling elderly: a prospective cohort study. *Journal of clinical epidemiology*, 55(11):1088–1094, 2002.

[265] Vianda S Stel, Saskia MF Pluijm, Dorly JH Deeg, Johannes H Smit, Lex M Bouter, and Paul Lips. A classification tree for predicting recurrent falling in community-dwelling older persons. *Journal of the American Geriatrics Society*, 51(10):1356–1364, 2003.

[266] Sławomir Wawak. Preventive actions vs. risk management in iso 9001: 2015. In *9th international quality Conference (str. 3–9). Kragujevac: University of Kragujevac*, 2015.

[267] Radim Bělohlávek, Joseph W Dauben, and George J Klir. *Fuzzy logic and mathematics: a historical perspective.* Oxford University Press, 2017.

[268] Lotfi Asker Zadeh, George J Klir, and Bo Yuan. *Fuzzy sets, fuzzy logic, and fuzzy systems: selected papers*, volume 6. World Scientific, 1996.

[269] Michael Smithson. *Ignorance and uncertainty: Emerging paradigms.* Springer Science & Business Media, 2012.

[270] Michael Smithson. A vocabulary of ignorance. In *Ignorance and Uncertainty*, pages 1–13. Springer, 1989.

[271] Philippe Smets. Imperfect information: Imprecision and uncertainty. In *Uncertainty management in information systems*, pages 225–254. Springer, 1997.

[272] Didier Dubois and Henri Prade. Modelling uncertainty and inductive inference: a survey of recent non-additive probability systems. *Acta Psychologica*, 68(1-3):53–78, 1988.

[273] Patric Bosc and Henri Prade. An introduction to the fuzzy set and possibility theory-based treatment of flexible queries and uncertain or imprecise databases. In *Uncertainty management in information systems*, pages 285–324. Springer, 1997.

[274] George J Klir and B Yuan. Fuzzy sets and fuzzy logic: theory and applications prentice hall. *Upper Saddle River, NJ*, 1995.

[275] Vern R Walker. Theories of uncertainty: explaining the possible sources of error in inferences. In *The Dynamics of Judicial Proof*, pages 197–236. Springer, 2002.

[276] Herbert Alexander Simon. *Models of bounded rationality: Empirically grounded economic reason*, volume 3. MIT press, 1997.

[277] Bruce F Baird. *Managerial decisions under uncertainty: An introduction to the analysis of decision making*, volume 4. John Wiley & Sons, 1989.

[278] John D Hey. Dynamic decision making under uncertainty: An experimental study of the dynamic competitive firm. *Oxford Economic Papers*, pages 58–82, 1993.

[279] Marco Pegoraro and Wil MP van der Aalst. Mining uncertain event data in process mining. In *2019 International Conference on Process Mining (ICPM)*, pages 89–96. IEEE, 2019.

[280] Michael Koller and RA Wicklund. Communication caused by uncertainty. *Psychological Beitrage*, 35(2/3):261–273, 1993.

[281] Ward Edwards. The theory of decision making. *Psychological bulletin*, 51(4):380, 1954.

[282] Clarence W De Silva. *Intelligent control: fuzzy logic applications*. CRC press, 1995.

[283] Vandna Kamboj and Amrit Kaur. Comparison of constant sugeno-type and mamdani-type fuzzy inference system for load sensor. *International Journal of Soft Computing and Engineering*, 3(2):204–207, 2013.

[284] Katarína Valášková, Tomáš Klieštik, and Mária Mišánková. The role of fuzzy logic in decision making process. In *2014 2nd International Conference on Management Innovation and Business Innovation*, volume 44, pages 143–148, 2014.

[285] Ouafae Kasmi, Amine Baina, and Mostafa Bellafkih. Multi level integrity management in enhanced telecom operation for lte/lte-a networks. In *2017 International Conference on Wireless Networks and Mobile Communications (WINCOM)*, pages 1–6. IEEE, 2017.

[286] Tomasz Korol. A fuzzy logic model for forecasting exchange rates. *Knowledge-Based Systems*, 67:49–60, 2014.

[287] Talha Azfar, Waqas Ahmed, Abdul Haseeb, Rizwan Ahmad, Sufi Tabassum Gul, and Fu-Chun Zheng. A low complexity online controller using fuzzy logic in energy harvesting wsns. *Science China Information Sciences*, 62(4):1–10, 2019.

[288] Malay K Pakhira. A linear time-complexity k-means algorithm using cluster shifting. In *2014 International Conference on Computational Intelligence and Communication Networks*, pages 1047–1051. IEEE, 2014.

[289] Daniela Grigori, Fabio Casati, Malu Castellanos, Umeshwar Dayal, Mehmet Sayal, and Ming-Chien Shan. Business process intelligence. *Computers in industry*, 53(3):321–343, 2004.

[290] Bernd Schwegmann, Martin Matzner, and Christian Janiesch. A method and tool for predictive event-driven process analytics. In *Wirtschaftsinformatik*, page 46. Citeseer, 2013.

[291] Werner Schmidt. Business activity monitoring (bam). In *Business Intelligence and Performance Management*, pages 229–242. Springer, 2013.

[292] Opher Etzion and Peter Niblett. Event processing in action, manning publications, 2011.

[293] Rainer von Ammon, Opher Etzion, Heiko Ludwig, Adrian Paschke, and Nenad Stojanovic. Introduction to the second international workshop on event-driven business process management (edbpm09). In *International Conference on Business Process Management*, pages 345–346. Springer, 2009.

[294] Christian Janiesch, Martin Matzner, and Oliver Müller. Beyond process monitoring: a proof-of-concept of event-driven business activity management. *Business Process Management Journal*, 18(4):625–643, 2012.

[295] Deng Bo, Ding Kun, and Zhang Xiaoyi. A high performance enterprise service bus platform for complex event processing. In *2008 Seventh International Conference on Grid and Cooperative Computing*, pages 577–582. IEEE, 2008.

[296] David Luckham. Soa, eda, bpm and cep are all complementary. *Retrieved May*, 20:2011, 2007.

[297] Antonio Estruch and José Antonio Heredia Álvaro. Event-driven manufacturing process management approach. In *International Conference on Business Process Management*, pages 120–133. Springer, 2012.

[298] Rainer von Ammon, Christoph Emmersberger, Florian Springer, and Christian Wolff. Event-driven business process management and its practical application taking the example of dhl. 2008.

[299] Claire Costello, Owen Molloy, and K Hinkelmann. Towards a semantic framework for business activity monitoring and management. In *AAAI spring symposium: AI meets business rules and process management*, pages 17–27, 2008.

[300] David Luckham. *The power of events*, volume 204. Addison-Wesley Reading, 2002.

[301] K Danielsson and G Trotta. Key requirements for event-driven bpm and soa.

[302] Ling Liu and M Tamer Özsu. *Encyclopedia of database systems*, volume 6. Springer New York, NY, USA:, 2009.

[303] Amina Jarraya, Nathan Ramoly, Amel Bouzeghoub, Khedija Arour, Amel Borgi, and Béatrice Finance. A fuzzy semantic cep model for situation identification in smart homes. In *Proceedings of the Twenty-second European Conference on Artificial Intelligence*, pages 1678–1679, 2016.

[304] Ronny Seiger, Francesca Zerbato, Andrea Burattin, Luciano García-Bañuelos, and Barbara Weber. Towards iot-driven process event log generation for conformance checking in smart factories. In *2020 IEEE 24th International Enterprise Distributed Object Computing Workshop (EDOCW)*, pages 20–26. IEEE, 2020.

[305] Wen Yao, Chao-Hsien Chu, and Zang Li. Leveraging complex event processing for smart hospitals using rfid. *Journal of Network and Computer Applications*, 34(3):799–810, 2011.

[306] Raef Mousheimish, Yehia Taher, Karine Zeitouni, and Michel Dubus. Pact-art: Enrichment, data mining, and complex event processing in the internet of cultural things. In *2016 12th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*, pages 476–483. IEEE.

[307] David Luckham. A short history of complex event processing. *Part 1: Beginnings. 1-a-short-history-of-cep-part-1. pdf*, 2007.

[308] D Robins. Complex event processing. In *Second International Workshop on Education Technology and Computer Science. Wuhan*, pages 1–10. Citeseer, 2010.

[309] Darko Anicic, Paul Fodor, Sebastian Rudolph, Roland Stühmer, Nenad Stojanovic, and Rudi Studer. Etalis: Rule-based reasoning in event processing. In *Reasoning in event-based distributed systems*, pages 99–124. Springer, 2011.

[310] Gianpaolo Cugola and Alessandro Margara. Processing flows of information: From data stream to complex event processing. *ACM Computing Surveys (CSUR)*, 44(3):15, 2012.

[311] David S Rosenblum and Alexander L Wolf. *A design framework for Internet-scale event observation and notification*, volume 22. ACM, 1997.

[312] Alessandro Margara, Gianpaolo Cugola, and Giordano Tamburrelli. Learning from the past: automated rule generation for complex event processing. In *Proceedings of the 8th ACM International Conference on Distributed Event-Based Systems*, pages 47–58. ACM, 2014.

[313] Brenda M Michelson. Event-driven architecture overview. *Patricia Seybold Group*, 2(12):10–1571, 2006.

[314] Bertha Mazon-Olivo, Dixys Hernández-Rojas, José Maza-Salinas, and Alberto Pan. Rules engine and complex event processor in the context of internet of things for precision agriculture. *Computers and electronics in agriculture*, 154:347–360, 2018.

[315] Andrew Eisenberg and Jim Melton. Sql: 1999, formerly known as sql3. *ACM Sigmod record*, 28(1):131–138, 1999.

[316] Nijat Mehdiyev, Julian Krumeich, David Enke, Dirk Werth, and Peter Loos. Determination of rule patterns in complex event processing using machine learning techniques. *Procedia Computer Science*, 61:395–401, 2015.

[317] Christopher Mutschler and Michael Philippsen. Learning event detection rules with noise hidden markov models. In *Adaptive Hardware and Systems (AHS), 2012 NASA/ESA Conference on*, pages 159–166. IEEE, 2012.

[318] Sinan Sen, Nenad Stojanovic, and Ljiljana Stojanovic. An approach for iterative event pattern recommendation. In *Proceedings of the Fourth ACM International Conference on Distributed Event-Based Systems*, pages 196–205. ACM, 2010.

[319] Julia Pielmeier, Stefan Braunreuther, and Gunther Reinhart. Approach for defining rules in the context of complex event processing. *Procedia CIRP*, 67:8–12, 2018.

[320] David C Luckham. *Event processing for business: organizing the real-time enterprise*. John Wiley & Sons, 2011.

[321] Jeff kaplan, when internet of things meets big data. `https://www.informationweek.com/big-data/big-data-analytics/when-internet-of-things-meets-big-data/a/d-id/1298137?` Accessed: 2021-08-30.

[322] UCISA ITIL. Introducing service operation. `https://www.ucisa.ac.uk`. [Online; accessed January-2021].

[323] J.Kaplan. When internet of things meets big data. `https://www.informationweek.com/big-data/big-data-analytics/when-internet-of-things-meets-big-data/a/d-id/1298137?`,. [Online; accessed July-2021].

[324] Jesús Gutiérrez, Víctor Rodríguez, and Sergio Martin. Comprehensive review of vision-based fall detection systems. *Sensors*, 21(3):947, 2021.

[325] Amal El Kaid, Karim Baïna, and Jamal Baïna. Reduce false positive alerts for elderly person fall video-detection algorithm by convolutional neural network model. *Procedia computer science*, 148:2–11, 2019.

[326] Yuxi Chen, Xiangbo Kong, Lin Meng, and Hiroyuki Tomiyama. An edge computing based fall detection system for elderly persons. *Procedia Computer Science*, 174:9–14, 2020.

[327] Cisco. "cisco fog computing solutions: Unleash the power of the internet of things", syst. inc., san jose, ca, usa, white paper, 2015. `https://www.cisco.com/c/dam/en_us/solutions/trends/iot/docs/computing-solutions.pdf`. [Online; accessed July-2021].

[328] Winsystems. "cloud, fog and edge computing – what's the difference?". `https://www.winsystems.com/cloud-fog-and-edge-computing-whats-the-difference/`. [Online; accessed July-2021].

[329] Qi Zhang, Lu Cheng, and Raouf Boutaba. Cloud computing: state-of-the-art and research challenges. *Journal of internet services and applications*, 1(1):7–18, 2010.

[330] Peter Mell, Tim Grance, et al. The nist definition of cloud computing. 2011.

[331] Peter Kairouz, H Brendan McMahan, Brendan Avent, Aurélien Bellet, Mehdi Bennis, Arjun Nitin Bhagoji, Kallista Bonawitz, Zachary Charles, Graham Cormode, Rachel Cummings, et al. Advances and open problems in federated learning. *arXiv preprint arXiv:1912.04977*, 2019.

[332] Chiara Di Francescomarino and Chiara Ghidini. Predictive process monitoring.

[333] Andrea Burattin. Streaming process mining. *Process Mining Handbook*, 349, 2022.

[334] John A Hartigan and Manchek A Wong. Algorithm as 136: A k-means clustering algorithm. *Journal of the royal statistical society. series c (applied statistics)*, 28(1):100–108, 1979.

[335] Robert L Thorndike. Who belongs in the family? *Psychometrika*, 18(4):267–276, 1953.

[336] Peter J Rousseeuw. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20:53–65, 1987.

[337] Robert Tibshirani, Guenther Walther, and Trevor Hastie. Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2):411–423, 2001.

[338] Kamran Khan, Saif Ur Rehman, Kamran Aziz, Simon Fong, and Sababady Sarasvady. Dbscan: Past, present and future. In *The fifth international conference on the applications of digital information and web technologies (ICADIWT 2014)*, pages 232–238. IEEE, 2014.

[339] Timothy J Ross. *Fuzzy logic with engineering applications*. John Wiley & Sons, 2005.

[340] Lior Rokach. A survey of clustering algorithms. In *Data mining and knowledge discovery handbook*, pages 269–298. Springer, 2009.

[341] James C Bezdek. *Pattern recognition with fuzzy objective function algorithms*. Springer Science & Business Media, 2013.

[342] Fernando García Alcalde. Fuzzy technology for gene ontology and transcription factor binding sites. *A PhD Dissertation*, 2011.

[343] Keinosuke Fukunaga and Larry Hostetler. The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE Transactions on information theory*, 21(1):32–40, 1975.

[344] Dorin Comaniciu, Visvanathan Ramesh, and Peter Meer. Kernel-based object tracking. *IEEE Transactions on pattern analysis and machine intelligence*, 25(5):564–577, 2003.

[345] Wenbing Tao, Hai Jin, and Yimin Zhang. Color image segmentation based on mean shift and normalized cuts. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 37(5):1382–1389, 2007.

[346] Damir Demirović. An implementation of the mean shift algorithm. *Image Processing On Line*, 9:251–268, 2019.

[347] Md Shamsur Rahim, Khoi Anh Nguyen, Rodney Anthony Stewart, Tanvir Ahmed, Damien Giurco, and Michael Blumenstein. A clustering solution for analyzing residential water consumption patterns. *Knowledge-Based Systems*, 233:107522, 2021.

[348] Oncel Tuzel, Fatih Porikli, and Peter Meer. Kernel methods for weakly supervised mean shift clustering. In *2009 IEEE 12th International Conference on Computer Vision*, pages 48–55. IEEE, 2009.

[349] Miguel A Carreira-Perpinán. A review of mean-shift algorithms for clustering. *arXiv preprint arXiv:1503.00687*, 2015.

[350] Andrew Y Ng, Michael I Jordan, and Yair Weiss. On spectral clustering: Analysis and an algorithm. In *Advances in neural information processing systems*, pages 849–856, 2002.

[351] Francis Bach and Michael Jordan. Learning spectral clustering. *Advances in neural information processing systems*, 16(2):305–312, 2004.

[352] Fei Gao. *Distributed approximate spectral clustering for large-scale datasets*. PhD thesis, Applied Science: School of Computing Science, 2011.

[353] Hadley Wickham. *Advanced r*. CRC press, 2019.

[354] Hadley Wickham and Garrett Grolemund. *R for data science: import, tidy, transform, visualize, and model data*. " O'Reilly Media, Inc.", 2016.

[355] Hadley Wickham, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, Alex Hayes, Lionel Henry, Jim Hester, et al. Welcome to the tidyverse. *Journal of open source software*, 4(43):1686, 2019.

[356] Nadia Rahmah and Imas Sukaesih Sitanggang. Determination of optimal epsilon (eps) value on dbscan algorithm to clustering data on peatland hotspots in sumatra.

In *IOP conference series: earth and environmental science*, volume 31, page 012012. IOP Publishing, 2016.