Dipartimento di Impresa e Management

Corso di Laurea Triennale in Management and Computer Science

Cattedra di Digital Business and Workplace Technology

Unlocking Operational Excellence: The Synergy of Digital Twins and Process Mining

Prof. Paolo Spagnoletti
RELATORE

Penelope Aprea
Matr. 272221
CANDIDATO

Anno Accademico 2023/24

Table of Contents

Abstract

This thesis explores the integration of Digital Twins (DTs) and Process Mining (PM) to enhance decision-making and operational efficiency in modern organizations. The study examines the definitions, historical development, and applications of DTs. It also delves into the role of PM in extracting and analyzing event data to uncover inefficiencies and optimize processes. The research employs a systematic literature review and a practical instantiation using Celonis on a multinational company's Purchase-to-Pay (P2P) process. Findings demonstrate that PM significantly supports DT creation by providing real-time data, enhancing simulation capabilities, and improving synchronization. This synergy offers profound benefits in increased productivity, cost reduction, and continuous improvement. This integration is becoming an essential tool in the digital transformation of business processes.

Introduction

In recent years, digitalization and digital transformation have emerged as one of the key aspects of change in organizations across various industries. This transformation entails the adoption of digital technologies in all aspects of the business, from administration to production [52]. Nowadays, organizations have the opportunity to develop and improve all their processes through various technological advancements, including but not limited to artificial intelligence, could computing, big data analytics, and the Internet of Things (IoT).

Every industry is constantly evolving, with disruptive technologies, new entrants, and new consumer preferences. The competition is set to always grow so organizations must adapt and innovate constantly to remain relevant and competitive. Innovation must be embraced by organizations to survive in today’s business landscape. Automation and data-driven decision-making are revolutionizing how organizations operate by reducing manual intervention, optimizing processes, and increasing operational efficiency.

In the rapidly evolving landscape of modern business, the integration of digital technologies has become paramount for organizations striving to maintain competitiveness and drive innovation. Central to this transformation are concepts such as Digital Twins and process mining, which offer powerful tools for enhancing operational efficiency and strategic decision-making.

The concept of the "Digital Twin" emerges as a disruptive innovation with the potential to radically revolutionize the operational management of organizations. A Digital Twin is a virtual replica of a process, product, or physical system that accurately reflects the state and behavior of its real-world counterpart. Through the use of real-time data, sensors, and analytics algorithms, Digital Twins enable organizations to monitor, simulate, and optimize their operational processes with an unprecedented level of precision and detail.

In this context, Digital Twins are a strategically significant tool that can redefine processes and operations to achieve operational excellence. This means pursuing efficiency, quality, and continuous improvement in all aspects of organizational operations.

The aim of this thesis is to answer the following research question: *"How can process mining support the creation of digital twins?"*. This will be achieved through an analysis of the definitions, history, applications, advantages and challenges in implementation of Digital Twins, as well as a practical study using process mining with Celonis. The final goal is to provide an in-depth analysis and insight into the best practices for the implementation of Digital Twins with the support of process mining.

1. Digital Twins

Digital Twins represent a paradigm shift in how organizations conceptualize and manage their assets and processes. This section provides a comprehensive overview, starting with a historical perspective and a precise definition of Digital Twins. It delves into the challenges and opportunities associated with the implementation of Digital Twins, shedding light on the complexities organizations face in leveraging this technology effectively. Additionally, the section examines the diverse applications of Digital Twins across industries, showcasing their versatility and transformative potential.

1.1. Historical Overview

In recent years, the concept of Digital Twin (DT) has gained massive popularity but its origin is not entirely new. The concept of DT dates back to 2002 in relation to the Product Lifecycle Management (PLM) at the University of Michigan. The proposed model, then referred to as "*Mirrored Space Model*" [3], has three components: real space, virtual space, and a linking mechanism for the flow of data between the two. In 1991, David Gelernter imagined a similar model, called "*Mirror Worlds*" [4], in which a software is used to mimic reality from information given as input from the physical world. Additionally, in 2003, Främling, K. et al. proposed as a solution to the inefficiency of transfer of information via paper for PLM "*an agent-based architecture where each product item has a corresponding 'virtual counterpart' or agent associated with it. Agents provide services for their physical counterparts."* [5]. In 2006, the name of Grieves's conceptual model was changed to "*Information Mirroring Model*" [2]. The model emphasized that the linking between the two spaces is bidirectional and that there are more virtual spaces for a single real space in order to be able to explore alternate ideas (Figure 1.1). At the time, DT had no practical applications due to limitations of technologies. The first mention of the name 'Digital Twin' (DT) appeared in NASA's draft version of the technological roadmap in 2010 where DT was also referred to as "*Virtual Digital Fleet Leader*". NASA was the first association to coin a definition for DT, it was described as "*an integrated multi-physics, multi-scale, probabilistic simulation of a vehicle or system that uses the best available physical models, sensor updates, fleet history, etc., to mirror the life of its flying twin.*" [6]. The timeline of the evolution of DT can be seen in Figure 1.2.

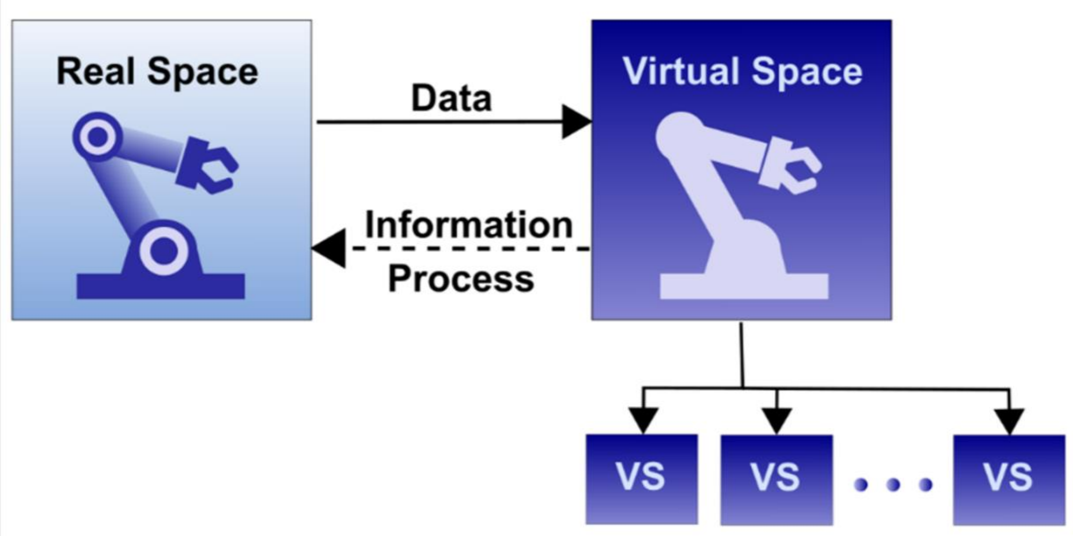

Figure 1.1. Mirrored Space Model/Information Mirroring Model as proposed by Micheal W. Grieves. Source: Singh, M. et al. (2021)

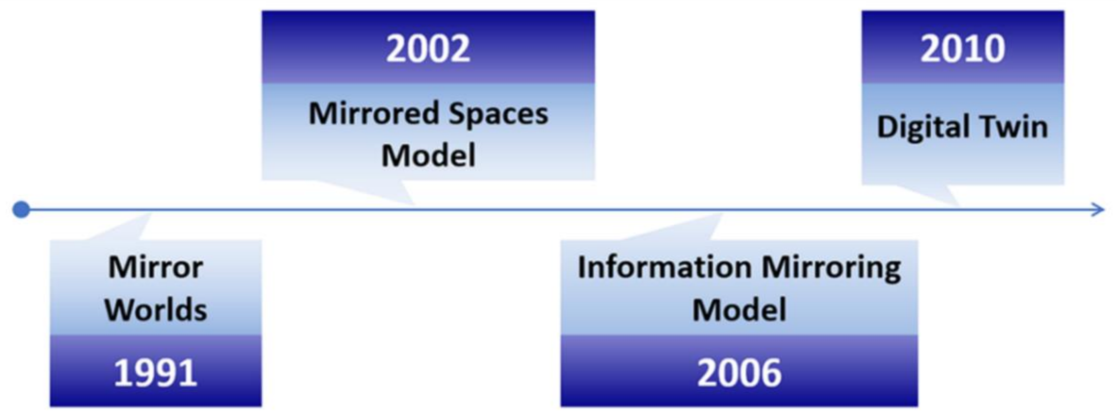

Figure 1.2. Timeline of evolution of Digital Twin. Source: Singh, M. et al. (2021)

1.2. Definition of Digital Twin

There is a myriad of definitions of DT in the literature, so there is no consensus of what can actually be described as DT. Since the first definition published by NASA, each author described DTs based on their different applications. At first, the definitions described DTs only with respect to the aeronautics/aerospace industry and then to manufacturing. However, since DTs' application has become broader, also its definition is not restricted to industrial products anymore but it can define complex biological systems, as trees and humans. Most definitions share the concept of bidirectionality of the transfer of data from the physical counterpart to the virtual one since a DT is a virtual representation of a physical object. The data shared includes environmental data, historical data, quantitative and qualitative data, and real-time data. This last type of data is the most important one in the development of a DT. A DT can be used to observe the performance of a system and its predictive maintenance because it used data to reflect the real world at any point in time [7]. The real-time data is essential to have dynamic models and simulations, which means they can change and make new predictions adapting to the new information.

As stated before, there is no unique definition of DT making it possible for academics and businesses to use the term DT loosely, creating confusion between different terminologies. In order to mitigate the confusion around different terminologies, Singh, M. et al. (2021) [1] proposes a definition that "can be applied irrespective of the industry or its application: '*A Digital Twin is a dynamic and self-evolving digital/virtual model or simulation of a real-life subject or object (part, machine, process, human, etc.) representing the exact state of its physical twin at any given point of time via exchanging the real-time data as well as keeping the historical data. It is not just the Digital Twin which mimics its physical twin but any changes in the Digital Twin are mimicked by the physical twin too.*'"

1.3. Characteristics and Types of Digital Twins

Digital Twins (DTs) can be classified into different types according to their distinctive properties, however all DTs have few characteristics in common such as high-fidelity (DTs should be an almost identical copy of its physical counterpart), dynamic (DTs should be able to change continuously to adapt to the changes in the real world), self-evolving (DTs should evolve in real time with the help of the data collected), identifiable (every physical asset needs to have its own DT), and hierarchical (DTs can be seen as a series of integrated submodels, the final product is composed by different parts and components that all have their own DT model).

There are different criteria to classify DTs, the terms Digital Model, Digital Shadow, and Digital Twin are often used as synonyms. However, their definitions depend on the level of data integration between the physical and the digital counterpart, some representations are fully connected to their physical object and integrated with real-time data exchange, while others are modelled manually [8]. The three subcategories designed by Kritzinger, W. et al. (2018) [8] are shown in Figure 1.3, Figure 1.4, and Figure 1.5.

Digital Model: This is a digital representation that might include a comprehensive description of the physical object, but there is no use of automated data exchange. The data exchange between the physical and digital object happens manually. The changes in the state of the digital object are not seen in the physical one directly, and vice versa. Some examples of Digital Models include simulation models of planned factories or mathematical models of new products.

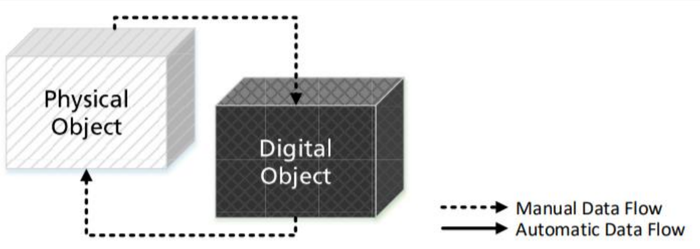

Figure 1.3. Data Flow in a Digital Model. Source: Kritzinger, W. et al. (2018)

Digital Shadow: This is a Digital Model with an automated one-way data flow. The data exchange from the physical object to the digital one is automatic, so it happens in real-time. However, this does not happen the other way around.

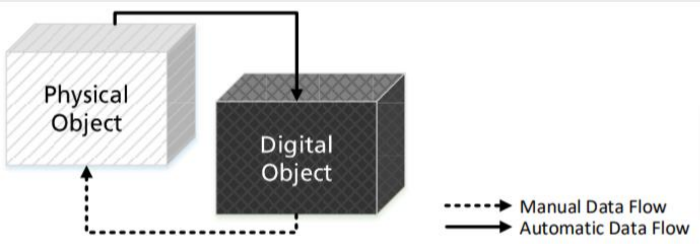

Figure 1.4. Data Flow in a Digital Shadow. Source: Kritzinger, W. et al. (2018)

Digital Twin: The data flow happens automatically in both directions. In this case, the digital object can be the controlling instance of the physical one. Furthermore, there can be other physical or digital objects that can affect the state of the digital object. Therefore, any changes to the physical object are reflected in the digital one, and vice versa.

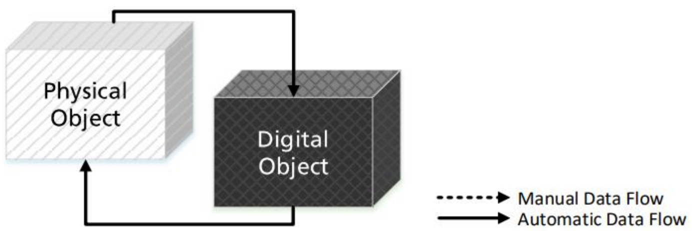

Figure 1.5. Data Flow in a Digital Twin. Source: Kritzinger, W. et al. (2018)

1.4. Advantages and Challenges in the Implementation of Digital Twins

The implementation of Digital Twins entails both advantages and challenges. Some of the advantages reported for DT are:

Cost-effective: As illustrated by Figure 1.6, the advantage to use DTs in regards to costs is clear. It is assumed that the costs are equal today, and if these costs are projected in the future, it can be noted how they diverge. Physical costs increase due to the effect of the inflation, while virtual costs exponentially decrease [9]. This is due the fact that DTs involve mostly virtual resources, and the cost of prototyping decreases with time. In addition, with traditional prototyping, performing destructive tests is highly costly because the prototype is completely destroyed, making the initial costs moot. On the other hand, when conducting destructive tests using DTs, products can be recreated and adapted without any additional material cost.

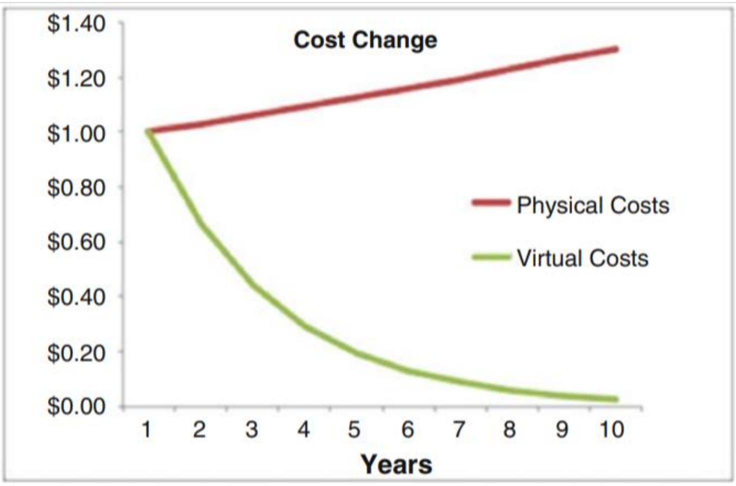

Figure 1.6. Physical costs vs Virtual costs. Source: Grieves, M., & Vickers, J. (2017)

Time-effective: The process of prototyping and re-designing is made easier and faster by using DT technology, because the simulations allow the investigation of different scenarios. During the design stage of a prototype, repeated simulation tests are required in order to ensure the desired performance of a product. Since a DT exists through the whole life-cycle of a physical object, it evolves with the prototype and can record the data of the product and the influence of the environment [10]. Additionally, through the real-time data, accurate prediction of the actual performance of the physical product can be made, and if needed, defects can be timely solved.

Waste Reduction: Using DT technology in the simulations and tests for a product or system significantly reduces wastage. Prototypes are studied and changed in a virtual environment, not only saving on material wastage but also reducing development costs and time to market.

Safer than Physical Counterpart: Due to the capability of a DT to access remotely its physical counterpart and its predictive nature, the risk of accidents and hazardous failures is extremely reduced in industries such as mining, and oil and gas. However, this is not the only example of increase safety when considering DT. During Covid-19, DTs where used to monitor the patients in order to reduce the frequency of in-person monitoring, thus reducing the risk of infection and improving the safety for employees. According to a recent Gartner study, "*one-third of mid-to-large-size companies that implemented IoT will have implemented at least one Digital Twin associated with a COVID-19-motivated use case.*" [11].

Problem Prediction: DTs can be used to predict errors and problems for future states of their physical counterpart. This ability can be used to plan the system accordingly. Since DTs have real-time flow of data, they can predict problems at each stage of the product lifecycle. It is harder to predict component failures with conventional methods in products that are very complex, so using DTs is beneficial for this kind of products that are composed by multiple parts, multiple materials, and complex structure [10].

While the advantages of implementing a technology like DTs are various, there are just as many challenges in their development. The level of difficulty of the implementation depends on the scale and the complexity of the DT, but there are some common barriers. DT technology has great potential but it is in its infancy stage and the complications that arise are several. The most common are the following:

Novelty of Technology: DT technology is still at the beginning of its development, meaning that is very limited even though it has great potential. In order to make a DT, several technologies are employed, such as AI, IoT, big data, machine learning, 3D simulations, and cloud computing. There is a need for in-depth research in these technologies in order to improve the infrastructures to implement DT [10]. In addition, there is a lack of case studies regarding either the implementation of DT in business activities or realistic estimation on the costs of the implementation [12].

Time and Cost: One of the major challenges in implementing DTs is the high cost associated with their implementation. Developing simulation software that are high-performing so that the events in the DT are faster than the ones in the real world is an extremely complicated process. This process requires huge amount of computational power and it is remarkably time-consuming, making DTs an expensive investment [13]. Gartner analyst Marc Halpern showed some concern when talking about DTs the 2018 PDT Europe conference in Gothenburg, Sweden, saying that "*it will take longer and will be more resource-consuming than anyone can imagine to get these solutions in place.*" [14]. It is crucial for all industries to analyze the benefits over the costs before implementing DTs.

Data Related Issues: One of the biggest issues that arises from the implementation of DTs is about the data, in particular about the ownership of the data, privacy, and confidentiality and transparency. Beyond the complexity of the technology, data ownership is influenced also by company policies and the mindset of people and society about the topic. The lack of policies and information sharing can result in data silos in different departments of the same organization that can lead to issues with the value chain [15]. In addition, the quality of the data is extremely important, it needs to be cleaned and sorted. In order to implement an effective DT, there must not be inconsistencies in the data otherwise there will be a performance risk in the DT due to the corrupted or missing data [21].

Trust and Expectations: Two sides of the same coin, lack of trust and high expectations, both negative and positive, on everything that regards this new technology present equally challenging concerns [21]. Both problems arise from a lack of understanding and come both from an organization point of view and that of the user. DTs are a fairly new technology and there is a great need to develop adequate foundations. In order to overcome these issues, this technology needs to be further discussed and validated, which means verifying that DTs are performing as expected.

1.5. Applications of Digital Twins

In this section, the different applications of Digital Twins (DTs) are discussed. Their role in the revolution of decision-making, in the optimization of operational efficiencies, and in the development of prediction-based maintenance strategies is examined. This will provide a detailed analysis of their deployment across different domains.

DTs can be considered the enabling layer in healthcare for building smart environments shaped as mirror worlds – a digital layer operated by software agents coupled with a physical environment that has an exact digital counterpart in the mirror [16]. Figure 1.7 shows a conceptual representation of an agent-based Digital Twin as mirror world for the healthcare context.

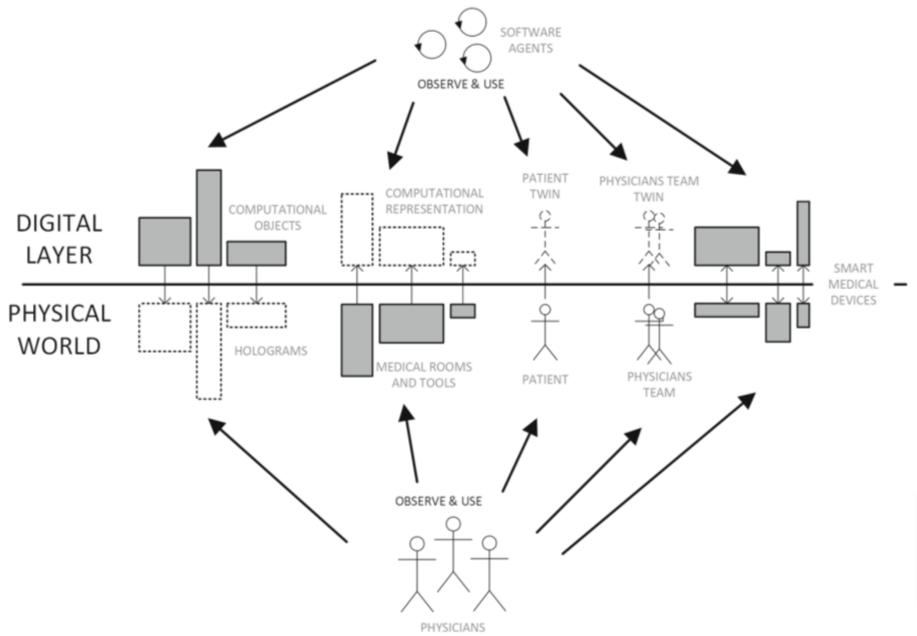

Figure 1.7. Agent-based Digital Twins as mirror worlds – a conceptual representation for the healthcare context. Source: Croatti, A. et al. (2020)

In this context, any relevant asset (e.g. patients, a vital sign monitoring device, or even structures like hospitals, with rooms, medics, etc.) could have a digital counterpart modelled in the environment that can be perceived by software agents and acted upon. This makes it possible to have simulations of environments specific to their needs, which can be useful to support agent decision making. For example, an agent can simulate the effect that a planned action is going to have according to the state of the twin, before executing it. Furthermore, combining DTs with AI algorithms allows to make more accurate predictions that are beneficial the ongoing treatment of a patient. In addition, DTs in the healthcare sector are used to plan and perform surgical procedures and to simulate the effects of certain drugs. A possible future application could be a DT of a human that can give real-time analysis of the body. The use of DTs in healthcare is still in its infancy. Nevertheless, the potential of this technology is huge [21].

DTs are also extremely used in the manufacturing industry, as there is a continuous research to increase efficiency [17]. Increasing digitalization in every stage of manufacturing gives manufacturers the potential to reach a new level of productivity. Figure 1.8 describes the notion of a Digital Twin in the lifecycle of a product, where the information created in each stage is made available to subsequent stages.

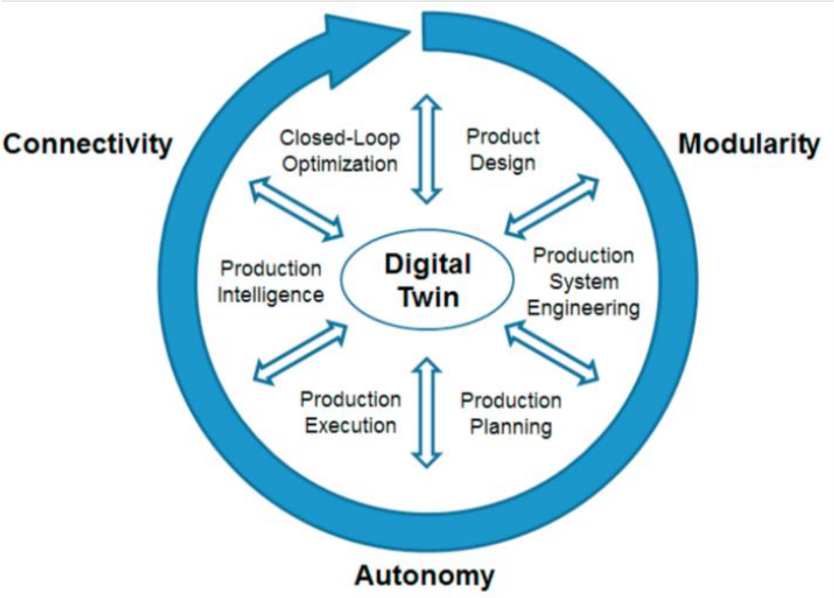

Figure 1.8. Notion of the Digital Twin in the lifecycle of a product. Source: Rosen, R. et al. (2015)

In this context, DTs can be a tool in production, programming, control, and can also assist with maintenance. Additionally, DTs can be also integrated to analyze a machine lifecycle, handling information flow through the different phases. This makes it possible to manage information more efficiently and have a complete visualization of a machine's condition [18]. Bambura, R. et al. (2020) [19], demonstrated the feasibility of an application of DTs in the real state of a manufacturing plant. Based on the data analysis given by the DT, they were able to detect the bottlenecks in the production line, and, as a result, they managed to increase productivity. Thus proving that DTs provide new perspective both on the production line and its management. The use of DTs coupled with AI algorithms has the potential to exponentially increase accuracy as the machine can hold large amount of data, which is needed for the analysis of the performance and for predictions [21].

The advancement of DTs is providing an important tool in the development of smart cities. Ivanov, S. et al. (2020) proposes a definition of a DT of a city: "*The Digital Twin of a city is a system of interconnected Digital Twins, representing certain aspects of the* *functioning and development of the urban environment.*" [22]. It is also stated that the key aspect to the functioning of a DT of a city is a continuous flow of data generated by different sources. In Figure 1.9, it is highlighted an example of the sources, which include information about traffic flow, both public and private, information about the physical environment, data collected from outdoors cameras and data from open sources, such as data on meteorological condition or open reporting information of business entities.

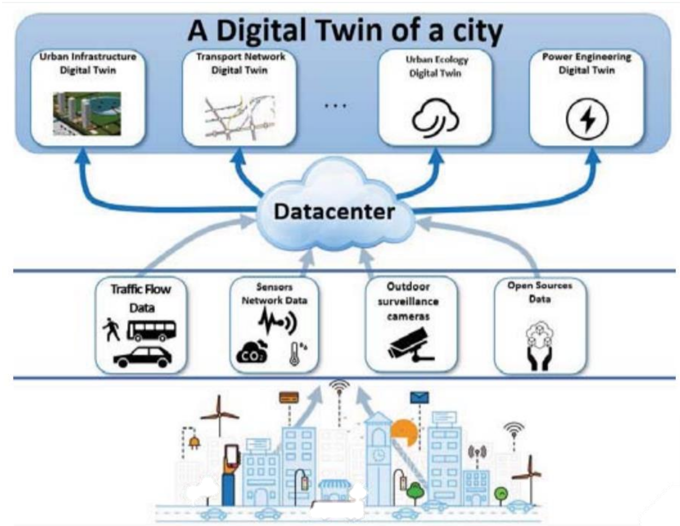

Figure 1.9. Example of the interface's appearance. Source: Ivanov, S. et al. (2020)

The use of DTs in smart cities gives great benefits non only in planning but also in energy savings [21]. In the first instance, sensors, that are imbedded in services and infrastructures within a smart city and controlled by IoT devices, provide the data to plan and develop current smart cities and also help with the development of new smart cities. Secondly, the data collected by the sensors gives also an insight on how the utilities are being used and distributed.

Lastly, the domain of agriculture is explored. DTs are becoming a pivotal tool for improving productivity, sustainability, and decision-making processes. DTs can create virtual replicas of the assets, such as crops, livestock, and farming operations. This technology can be integrated to monitor crops, by using data from different sources, such as sensors, satellite, and weather stations [20]. In addition, DTs can be also used for prediction for soil management, by modelling its characteristics and conditions. This way farmers can optimize and adapt the irrigation and fertilization. Finally, DTs can be used to model entire farming operations, in order to streamline processes and reduce costs.

2. Process Mining in Digital Twins

Process mining is fundamental in the creation of Digital Twins. It offers insights into the inner workings of organizational processes and can solve the challenges described in the previous section. This subsection begins by elucidating the concept of process mining and its fundamental principles. It then explores the pivotal role of process mining in analyzing business processes, emphasizing its capacity to uncover inefficiencies, bottlenecks, and opportunities for optimization. Furthermore, the practical application of process mining is discussed, highlighting its capabilities in facilitating data-driven process analysis and improvement initiatives. Finally, the findings obtained from a systematic literature review are analyzed, highlighting the different benefits of the applications of process mining in the creation of Digital Twins.

2.1. Definition and Concept of Process Mining

The origins of process mining can be traced back to the late ‘90s and early 2000s, when it emerged as a solution to the need to improve traditional business process management (BPM) techniques. In 2020, Gartner defines Process Mining (PM) as “*a technique designed to discover, monitor and improve real processes (i.e., not assumed processes) by extracting readily available knowledge from the event logs of information systems.*” [23]. There are various facets of PM including automated process discovery, conformance checking, social network/organizational mining, case prediction, automated construction of simulations models, and many more.

Process Mining can be seen as an intersection of data science and process science (Figure 2.1) [24]. Data science can be described as “*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.*” [25]. On the other hand, vom Brocke, J. et al. (2021) [26] defines process science as “*the interdisciplinary study of continuous change. By process, we mean a coherent series of changes that unfold over time and occur at multiple levels.*”

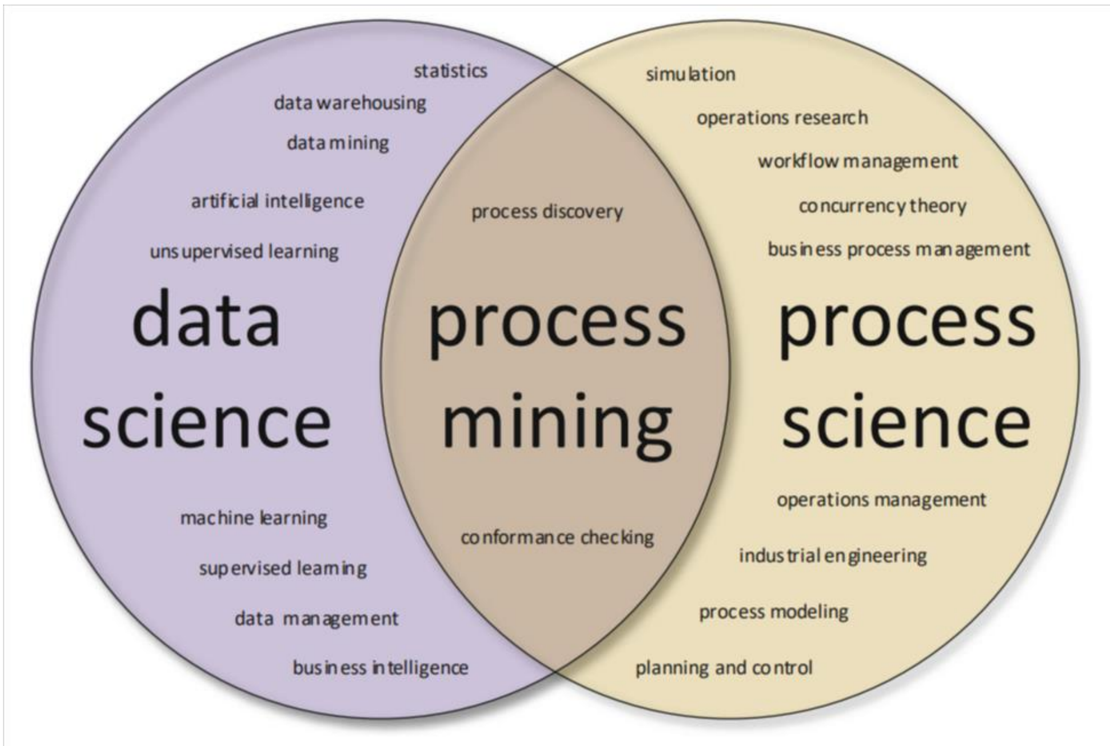

Figure 2.1. Process mining as an intersection of data science and process science. Source: van der Aalst, W.M.P. (2022)

From these definitions and from Figure 2.1, PM is viewed as a bridge between data science and process science. PM aims to tackle the challenge of exploiting event data in a meaningful way. For example, PM can be used to identify bottlenecks, anticipate problems and recommend countermeasures, provide insights, and streamline processes [27]. PM techniques assume as possible the storage of events in sequential order such that each event is related to a particular case and refers to an activity.

There are three basic types of PM: discovery, conformance, and enhancement. In the first technique, which is the most prominent, an event log is taken and a model is produced without the use of any a-priori information. On the other hand, the conformance checking technique utilizes an already existing process model and it compares it to an event log of the same process, in order to check if the model conforms to reality, and vice versa. Finally, the enhancement technique is used to change or extend an already existing model. An existing process model is improved using information about the actual process contained in an event log. Figure 2.2 shows the positioning of the three types of PM.

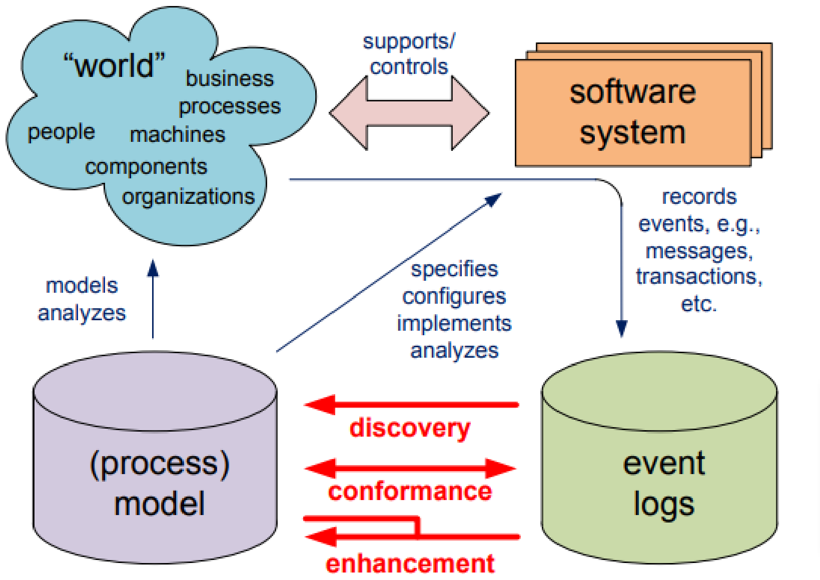

Figure 2.2. Positioning of the three types of process mining. Source: van der Aalst, W.M.P. et al. (2011) [28]

2.2. Role of Process Mining in the Analysis of Business Processes

In today's business landscape, Process Mining is a business process intelligence (BPI) technology that is receiving a lot of interest both in research and in practice. Badakhshan, P. et al. [29] conducted a study in 2022 aimed at highlighting how organizations use PM to create business value. They conducted a series of interviews with stakeholders from eight internationally operating companies. PM can be used to inform managerial actions by visualizing and measuring, though the use of digital trace data, the performance of business processes. Figure 2.3 depicts the results of the study. It shows three main categories: Process Mining Features, Process Mining Affordances, and Business Values.

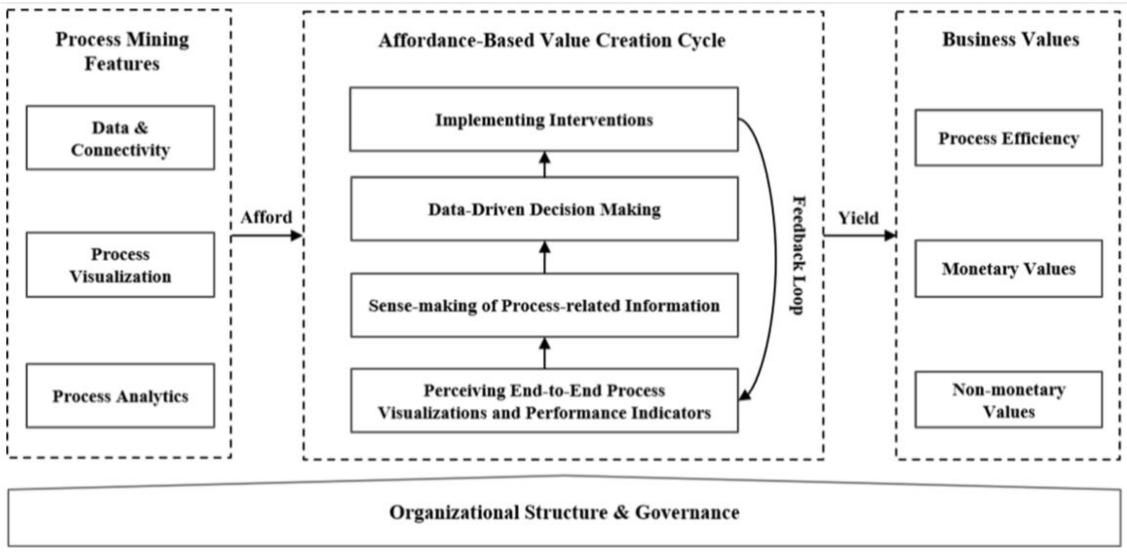

Figure 2.3. Process Mining Value Model. Source: Badakshan, P. et al. (2022)

The first category includes the technological characteristics used for value creation. This was found to have three sub-categories: Data & Connectivity, which shows how PM makes it more accessible to collect process-related data, Process Visualization, which translates process data into end-to-end process diagrams, and Process Analytics, which includes the calculations of KPIs (key performance indicators), rule-based alerts and recommendations, and the comparison of the process to a predefined model. One of the interviewee stated that "*with process mining, we monitor what is changing, why it is changing and where, and receive the list of violations and where they are happening, as well as receiving recommendations on what to do with the detected violations*" showing how these features allow for setting process-related rules for triggering process activities.

Secondly, through the PM features, stakeholders have a set of affordances to enact in order to create business value. First, they are able to perceive end-to-end process visualization and performance indicators. PM makes it possible to have insights into process behaviors not only to specialized personnel but also to people not too keen on process analysis and data analytics. In addition, PM affords a sense-making of process-related information. The key aspect needed to turn the representations in PM software into actionable information is the knowledge of both the technical and business aspects, which helps with the detection of actual bottlenecks and to differentiate them from noise. Furthermore, PM affords data-driven decision making. PM can be used to leverage knowledge across the organization. Finally, PM affords organizations to improve their business processes and organizational performance through the design and implementations of targeted interventions.

The third and final category developed by the study is Business Values, which makes the distinction between value obtained by Process Efficiency (reduce process costs, time etc.), Monetary Values (optimize working capital), and Non-Monetary Values (increase customer satisfaction). The model defined by Badakhshan, P. et al. (2022) [29] includes also the category Organizational Structure and Governance. This category is pivotal for the adoption of PM in business organizations. In order to use PM technology for value creation, it is needed an in-depth analysis of cultural aspects and administrative work. The interviews revealed that, before implementing this technology, employees needed to be informed and trained. To achieve a sustained organizational acceptance and usage, this technology needs to be trusted and this can be done by increasing awareness.

2.3. Applications of Process Mining

Process Mining can be applied across a plethora of industries. The most common are: healthcare (patient treatment, or processes of a hospital), information and communication technology (ICT – software development, telecommunication companies), manufacturing (industrial activities), and finance (investments, risk analysis and mitigation) [30].

Over the last years, the healthcare sector has been using PM in various ways to improve processes and enhance efficiency by analyzing the data generated within the healthcare systems. Every patient is different from another, with its unique set of complications. In order to decide the right treatment there is a complex decision-making process. PM is used to identify differences between groups of pathway, which are highly flexible, and help to decide if and where changes need to be made [31]. Disease trajectory modeling is also a possible application of PM. This process involves the mapping of the progress of a disease over time and the comparison of its evolution depending on patient attributes. A challenge in this application is found in the development of a clear model easy to understand and in the comparison of these models with clinical guidelines.

PM is vastly used in the ICT sector, in particular it is used to reach goals such as process improvement, process evaluation, risk management, increased performance maintenance, and compliance check in operation management [30]. An example of a possible way to employ PM in this area was carried out by Lemos, A.M. et al. (2011) [32]. This study showed that PM can be used to improve software process maturity level. The authors examined the potential of employing PM in evaluating the Quantitative Project Management (QPM) process area and discovered that it effectively facilitates the establishment of an infrastructure for Statistical Process Control (SPC) at higher maturity levels. Furthermore, Gupta, M. (2014) [33] suggested using process mining to develop execution-time process models, identify and eliminate deviations, enhance the capabilities of existing software engineering tools, and increase stakeholders' awareness of the process. This approach aims to better understand interaction patterns and boost the efficiency of software development.

The manufacturing sector also employs PM in different cases thanks to the increasing presence of information systems such as enterprise resource planning (ERP), manufacturing execution systems (MES), and supervisory control and data acquisition

(SCADA) [30]. In 2014, Karray, M.-H. et al. [34] developed a system called PETRA for examining maintenance processes. This system utilizes process mining to verify behaviors and to derive rules for maintenance activity execution. Additionally, it incorporates the IMAMO (Industrial Maintenance Management Ontology) to classify maintenance actions into 12 categories, which include corrective, preventive, scheduled, pre-determined, condition-based, predictive, remote, deferred, immediate, on-line, on-site, and operator. More recently, Ruschel, E. et al. (2017) [35] advanced an application that merges process mining with Bayesian networks. This integration forms predictive models that establish maintenance schedules for industrial machinery. The application is designed to support managerial decision-making by optimizing maintenance timing, thereby reducing equipment downtime and avoiding unexpected interruptions.

Lastly, PM can be applied in the finance sector for different applications such as risk analysis, insurance claim handling, analysis of ATM (Automated Teller Machines) processes, financial auditing, fraud, and is also capable of helping banks with loan approvals and the performance of contact centers [30]. In 2015, Conforti, R. et al. [36] implemented a recommendation system at a major insurance company to mitigate risks in operational processes. This system evaluates potential errors and their consequences, ensuring that users can make risk-informed decisions about processes occurring simultaneously. The system operates through two main components: (i) risk assessment – calculates the likelihood and impact of possible errors, incorporating statistical analyses of each error's severity and its effect on overall process operations –, (ii) activity allocation – determines the most suitable personnel for specific tasks based on the assessed risks using integer linear programming, thereby optimizing task allocation and minimizing overall process risks. This method is also capable to improve the efficiency of global execution time. In this context, a technique known as CMS miner, which is artifact-centric, was introduced to explore multiple perspectives within the same process. The primary objective was to uncover correlations between the behaviors of different artifacts to gain deeper insights and assess how these artifacts interact with each other. Applied to a loan process in a financial institution, the CMS miner efficiently revealed hidden information. This facilitated the identification of issues such as the average time spent on tasks and highlighted the top five most intriguing interactions between artifacts, among other findings [37].

2.4. Process Mining in the Creation and Optimization of Digital Twins

The analysis of the results of the systematic literature review indicates that process mining (PM) significantly support the creation and optimization of Digital Twins (DTs) by providing a method to extract, analyze, and utilize event log data generated during manufacturing processes. PM enhances the functionality of DTs by providing detailed, real-time data that improves synchronization, simulation accuracy, error recovery, production planning, and overall system optimization. This integration makes DTs a powerful tool for modern manufacturing. In this industry, DTs are used to model processes and machines to provide a comprehensive overview, and help with identifying potential issues and improvements. For example, a virtual model of an injection molding machine was made by using sensors that offer an insight into the machine's performance. The DT, created with the generated data, allows for detailed analysis and real-time monitoring, optimizing operations and preventing failures [38] [42].

PM facilitates the collection and integration of data generated by sensors and other sources in manufacturing, such as machines, production logs, and enterprise systems. Additionally, it is crucial for analyzing and understanding this data. These PM techniques are used for a detailed process analysis through the transformation of raw data into meaningful event logs. A DT must be trustworthy and has to effectively mirror real-world manufacturing processes. Data accuracy and reliability are ensured by data preprocessing and event log construction. The integration and consolidation of data creates a comprehensive dataset that reflects the operations and status of the physical system, which is vital for the creation of an accurate DT [38] [42] [49] [44] [51].

In order to apply PM techniques, one needs to construct event logs from sensor data. This involves data cleaning, data selection, event definition, and case definition. These steps represent accurately the physical processes and ensure reliable data for further analysis. Another benefit of PM is that it enhances the DT's simulation capabilities by providing realistic time constants and operation durations. Including detailed operation durations and resource availabilities leads to more precise are reliable simulation that enable better decision making and process optimization. DTs leverage PM to generate synthetic datasets through simulation. These datasets are crucial to enhance the DT's ability to predict and classify real-time data accurately. The predictive capabilities of a DT are improved by this synthetic data because it is used to train machine learning algorithms to recognize and respond accordingly to different types of anomalies. Furthermore, a key feature of DTs is the ability to stay synchronized with their physical counterpart. This feature can be achieved with the help pf PM because through PM techniques DTs are continuously updated with real-time data from the production system. This allows for real-time monitoring and control and the DT can remain relevant and accurate over time [39] [40] [41] [44] [45] [46] [50].

DTs enhanced by PM can dynamically generate and adjust production plans, increasing flexibility and responsiveness to changing conditions. An automated production plan can reduce downtime and improve efficiency by optimizing resource allocation and adjusting to real-time demands without user intervention and maintaining accuracy and reliability. The ability to adapt to disruptions and technological changes makes DTs a robust tool for both long-term planning and short-term decisions. Additionally, PM improves error recovery processes. DTs can use PM data to identify a different production path from a failed one. This capability improves resilience and reliability of a production system [39] [43] [47] [50].

Additionally, detailed and accurate data from PM allows monitoring many variables virtually within the DT, reducing the need for extensive physical sensors. This reduction in hardware lowers costs and maintenance efforts while maintaining high monitoring accuracy. Virtual sensors created through process mining techniques can monitor system performance and detect anomalies, providing a cost-efficient solution for comprehensive system monitoring [39].

Analyzing changing patterns in sensor data, PM can help identify frequent changes and their implications. This supports the creation of DTs able to predict and simulate different behaviors. PM also helps with identifying factors that lead to defects and inefficiencies, giving the possibility for the DT to suggest improvements and preventive measures. By understanding the root causes of problems, manufacturers can implement targeted solutions to improve process reliability and product quality [38]. PM can identify patterns indicating when a component is likely to fail. This predictive capability allows for proactive maintenance and optimization, which minimizes the downtime and extends the lifespan of equipment [42] [44] [50].

By providing a clear understanding of current workflows and predicting future trends, process mining helps managers make informed decisions regarding process adjustments, scheduling, and resource allocation. For example, PM can shed light on resource-intensive processes, allowing the optimization of resource allocation and improved efficiency [42] [44] [49] [51]. An example of a use for a DT enhanced with process mining is in supply chain management. By analyzing data from different points, PM identifies delays, inefficiencies and risks, allowing the DT to simulate different supply chain configuration and increase overall performance [44] [51].

PM is also fundamental in the context of cognitive DTs, which are DTs that leverage advanced data analytics and AI capabilities to provide deeper insights and decision making support. PM enables DTs to offer predictive and prescriptive insights by analyzing and integrating complex datasets. This enhances the cognitive capabilities and allows DTs to offer advanced analytics [39] [46].

In manufacturing there are often complex material flows, for example in assembly processes. Traditional PM techniques can have issues with this complexity due to their reliance on single parts identifiers. A solution is object-centric process mining (OCPM), which considers multiple object types and their interactions, thus enabling the generation of a complete graph model of the manufacturing system that includes non-linear material flows. OCPM also preserves the relationship between different objects throughout the production process, thus solving the "data flattening" issue. This problem involves different part identifiers and material flows that are collapsed into a single dataset and consequently lose critical information. Thanks to OCPM, it is possible to create an accurate DT that represents complex manufacturing processes without losing important details [46] [50].

Another core principal of manufacturing is continuous improvement. DTs supported by PM are an important tool to reach this goal. They provide ongoing insights into process performance and enable simulation of various scenarios. This makes it possible to have continuous support for process optimization, the DTs evolves together with the physical system so it remains relevant and effective. The ability of DTs to provide accurate simulations helps with testing strategies before the actual implementation in the real world, enabling decision-makers to have predictions of outcomes and making it possible for them to adjust their strategies accordingly. The simulated what-if scenarios help manufactures to stay flexible and agile in adapting to the constant changes in market demands and production configurations [40] [44] [45] [46] [50] [51]. Furthermore, DT models must be neither too simplistic nor overly complex. This is challenging to achieve but PM provides methods to tune the model by adjusting the detail level and removing redundant activities. The result is a more streamlined and efficient model that balances accuracy and efficiency [50].

Park, K.T. et al. (2022) [43] states that PM can be used to generate a systematic logic library that reflects functional units of agents with physical assets. Policies and decision logic are created for the DT to operate autonomously. In this context, AI and machine learning techniques are employed to learn human operators' decision-making processes, thus enabling DTs to make similar decision automatically. The data provided by PM enables human operators to visualize the changes and variations in production lines. Then they can understand and act on the DT's insights effectively.

A similar concept is described by Bano, D. et al. (2022) [48]. The user interaction part of a DT, which provides a graphical interface for data visualization, is the Digital Twin Cockpit (DTC). PM automates the generation and update of DTCs, improving usability and functionality for operators and managers. A Process-Aware Digital Twin Cockpit (PADTC) is a DTC with a stronger focus on processes that are explicitly defined. PM techniques extract live monitoring data from the event logs. This enhances the DTs' process awareness and allows for real-time monitoring and dynamic adjustments, providing flexibility and customization.

In summary, PM is crucial for the creation and integration of DTs. It provides the necessary tools for developing accurate, real-time models, and improving simulations. PM ensures that DTs can effectively mirror and optimize real-world processes by leveraging data-driven insight, thus leading to improved reliability, adaptability, and overall efficiency.

3. Methodology

This section describes the methods employed in order to answer the research question "*How can process mining support the creation of Digital Twins?*". First of all, a systematic literature review was performed to obtain an overall picture of the use of process mining in the creation of Digital Twins in organizations. Subsequently, a process mining instantiation has been conducted on a business process in Celonis in order to provide a practical insight on the different benefits of implementing process mining for the creation of Digital Twins. Celonis is a leading process mining and execution management platform. It uses process mining to provide deep insights into business processes, which enables companies to identify inefficiencies, discover hidden opportunities, and drive continuous improvement.

3.1. Systematic Literature Review (SLR)

To pursue the objective of this study, a systematic literature review on the literature discussing the role of process mining in supporting the creation of Digital Twins was performed. Peer-reviewed articles published in top journals between 2019 to 2024 were selected from the fields of Information Systems (IS), Innovation Management (IM), Operation Management (OM), and General Management (GM). To assemble a dataset that accomplishes the needs of this study, a systematic literature review procedure was followed. The sampled articles were assembled using a "backward" and "forward" search. The backward search begins from existing reviews to identify articles discussing the use of process mining in the creation of DTs. The corpora were expanded through the Scopus search engine database, excluding books and conference proceedings; this ensures that only high-quality scholarship outputs that have undergone rigorous peer reviews are included, consistent with findings demonstrating the relationship between journal ranking and rigor. Search terms that included keywords about process mining and digital twins and their variations were adopted. The search was conducted in the title, abstract, and keywords of articles published from 2019 forward, written in English. This period was chosen based on the increasing prevalence of digital twins and process mining technologies in organizational contexts. The query is reported in Table I.

Table I Research query

The query resulted in 32 items on May 2024. Moreover, the backward search led to the identification of 2 papers. Hence, the total number of documents retrieved is 34. Pre-specified inclusion/exclusion criteria guided the document screening process, as shown in Table II.

Table I. Research query.

Key terms	Term 1	Term 2
	"digital twin*" AND	"process mining"
Synonyms	"virtual replica" OR	"process analysis" OR
	"digital clone" OR	"process discovery" OR
	"digital model" OR	"business process mining"
	"digital shadow" OR	
	"virtual duplicate" OR	
	"digital mirror"	

Table II Exclusion and inclusion criteria

A total of 34 article titles and abstracts were screened. The final corpus is composed of 14 research articles, the details of which are displayed in Table III.

Table II. Exclusion and inclusion criteria.

No	Exclusion Criteria	Inclusion Criteria
1	Articles that do not offer a detailed description of the research context.	Articles must provide a comprehensive overview of the empirical study context, specifically within organizations using process mining and Digital Twins.
2	Articles not pertinent to the phenomenon of process mining and Digital Twins, such as those focused solely on technological aspects without organizational context.	Articles must primarily address the use and impact of process mining on the creation and functionality of Digital Twins.

Table III Complete review process

Table III. Complete review process.

Stage	Database Query
Identification	Records identified from: • Journals OM (n = 17) • Journals IS (n = 13) • Journals GM (n = 2) • Journals IM (n = 2)
Screening	Records screened (n = 34)
	Records excluded (n = 20)
	Records excluded: • No context provided (n = 9) • Not pertinent (n =8) • Other reasons (n =3)
Analysis	Studies included in the review (n = 14)

3.2. Process Mining Instantiation

In order to conduct a process mining instantiation to demonstrate how process mining can be useful in the creation of Digital Twins, the dataset from the BPI Challenge 2019 was taken into consideration [53]. The data, collected over approximately one year, describes the Purchase-to-Pay (P2P) process of a multinational company operating from the Netherlands in the area of paints and coatings. The dataset contains over 1.5 million events across over 250 thousand cases. Each case represents a purchase order with its associated items. Each event includes details such as order ID, item ID, vendor information, document types, timestamp, and various flags that indicate some specific condition. In the data, each purchase order contains one or more line items, for which there are four types of flows: (i) 3-way matching, invoice after goods receipt (GR) – the value of the GR has to be matched against the value of an invoice receipt message and the value put during the creation of the data, (ii) 3-way matching, invoice before goods receipt – GR message is required but GR-based invoice is not required, (iii) 2-way matching – the value of the invoice has to match the value at creation but a separate GR message is not required, (iv) consignment – there are not invoices on purchase order level because it is completely handled in a separate process. Since the data was collected in 2018, any event with a timestamp dated before January 1st 2018 was excluded from the analysis.

4. Findings and Implications

This section describes the results found form the analyses described in the previous chapter. First, there are the findings obtained by the process mining instantiation, which explore the chosen dataset by implementing process discovery, conformance checking, and process enhancement. Afterwards, the implications of the process mining instantiation are explored to make a stronger argument in favor of the use of process mining in the implementation of Digital Twins.

4.1. Findings of Process Mining Instantiation

The first step in the analysis is Process Discovery. This involves the inspection of the main activities, the detection of bottlenecks, the identification of deviations and variations in the process. In Celonis, the *Process Explorer* visualization outlined the primary steps in the procurement process and highlighted some key activities such as *Create Purchase Order Item, Vendor Creates Invoice, Record Goods Receipt, Record Invoice Receipt, and Clear Invoice*. The majority of cases follow this sequence, which indicates a structured procurement process. Then, with the *Variant Explorer*, the most common variant was identified. The variant covered 20% of cases, not only indicating significant process standardization, but also suggests room for improvement in order to reduce variation. Figure 4.1 displays the number of Purchase Order (PO) items and their net value on a monthly basis. There is a general increase in PO items with a peak in June 2018. The net value generally mirrors the items trend, with a notable decline at the end of 2018. This graph provides valuable insights into the value and volume trends of purchase orders over time and can help identifying periods of high and low productivity.

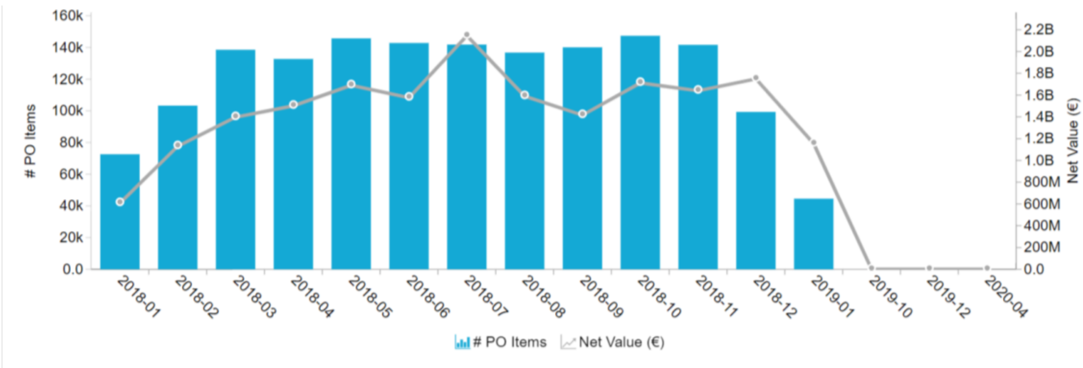

Figure 4.1. Purchase Order Items and Value by Month. Source: Personal elaboration, Celonis

Afterwards, a dashboard containing the *Process Overview* was used to get a more complete view of the process. It highlighted that the average throughput time was of 67 days, suggesting potential inefficiencies or delays in the process. Bottlenecks significantly increase the overall throughput time and need to be addressed to streamline the process. The most common bottlenecks detected were:

- **Record Invoice Receipt → Clear Invoice**: Duration of 44 work days, affecting 53% of cases.
- **Remove Payment Block → Clear Invoice**: Duration of 36 days, affecting 20% of cases.
- **Records Goods Receipt → Record Invoice Receipt**: Duration of 22 workdays, affecting 44% of cases.

Additionally, some activities can indicate areas where process deviations occur most frequently. In this process, they include: *Remove Payment Block* (22% of cases), *Create Purchase Requisition Item* (19% of cases), *Receive Order Confirmation* (13% of cases), and *Change Quantity* (7% of cases). Finally, the *"happy path"* describes the flow from the most frequent starting activity to the most frequent ending activity. The one detected in this process, shown in Figure 4.2, covers 20.01% of cases and has a throughput time of 70 days, which is slightly higher than the average.

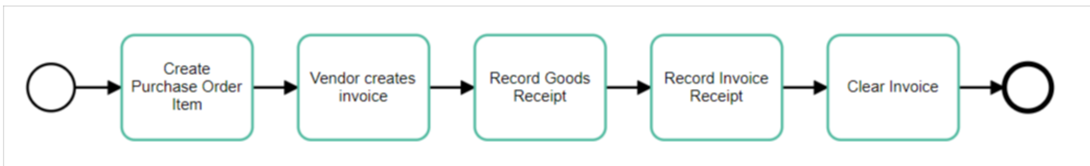

Figure 4.2. Happy Path/Ideal Model. Source: Personal elaboration, Celonis

Secondly, a Process Performance Checking was performed to compare the discovered process model with the ideal process model. Deviations and non-conformance issues were identified due to their relevance in the development of an accurate Digital Twin. The ideal model is the happy path provided in Figure 4.2 so the conforming cases are 50.3 thousand and the non-conforming cases cover the remaining 80%. Figure 4.3 shows the conformance history graph, which displays a declining in conformance over the year, starting from 26.36% and decreasing towards 0%. This trend suggests that the deviations are increasing and that there is potential process degradation over time.

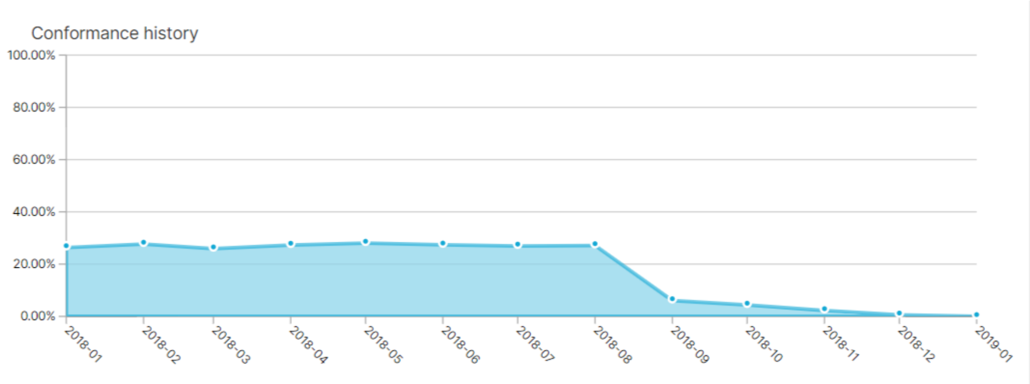

Figure 4.3. Conformance History graph. Source: Personal elaboration, Celonis

The throughput time of the violating cases is shorter by 1.3 days than the one of the conforming cases, possibly indicating shortcuts or omitted steps. On the other hand, the violating cases have more steps, indicating unnecessary or additional activities. The most common violations are:

- **Create Purchase Order Item → Record Goods Receipt** (42%): Increases steps per case by 2.1, and throughput time by 1 day.
- **Remove Payment Block** (22%): Increases steps per case by 2.1, and throughput time by 19 days.
- **Create Purchase Requisition Item** (19%): Decreases throughput time by 23 days but steps are increased by 0.8.
- **Receive Order Confirmation** (13%): Increases steps by 1.2, and throughput time by 1 day.
- **Change Quantity** (7%): Increases steps by 2.1, and throughput time by 22 days.
- **Change Price** (4%): Increases steps by 3.4, and throughput time by 24 days.

Lastly, Process Enhancement was implemented through the identification of areas in need of process improvement and some recommendations for optimization were provided.

- **Deviation Management**: The high occurrence of some deviations can be managed by establishing stricter controls and standard operating procedures (SOPs) or by regularly training staff on compliance to reduce the deviations.
- **Bottleneck Reduction**: The significant delays in the process can be addressed by implementing automation and clear guidelines for invoice processing.

- **Throughput Time Improvement**: The average throughput time can be optimized by eliminating steps that do not add value to the workflow and by implementing automation to speed up the process.
- **Conformance Improvement**: The relatively low conformance percentage of cases can be improved by regularly monitoring and conducting process audits to ensure adherence to the ideal process model and identify areas of non-compliance.

4.2. Implications

The results of the process mining instantiation clearly show the importance of process mining in the creation of Digital Twins. By providing a real-world dataset that can be used to model, simulate, and optimize business processes, it is possible to understand the complexities of a process, identify bottlenecks, and propose enhancement to streamline operations. The final objective is to leverage the insights given by process mining to create an accurate and dynamic digital representation of a company's process.

Event log data generated during manufacturing processes can be utilized, extracted, and analyzed with PM. In order to develop a DT, this capability of PM is vital because it allows for the transformation of raw data into meaningful event logs that represent real-world operations. In the analysis conducted using the BPI Challenge 2019 dataset, it was revealed that PM can shed light on the primary steps in a procurement process, detect bottlenecks, and identify deviations. A DT should be as accurate as possible when mirroring a physical process, so these insights are essential to ensure the reliability and effectiveness of the digital counterpart.

One of the core strengths of DTs is their ability to enable real-time monitoring and control by staying synchronized with their physical counterparts. PM techniques can be used to facilitate this ability by updating the DT with real-time data from production systems in a continuous manner. The DT can remain relevant and accurate over time thanks to this dynamic synchronization. Additionally, this allows for real-time adjustments and decision-making. Through the analysis in Celonis, it was highlighted how PM is able to identify the most frequent process paths and deviations in order to provide for real-time updates and continuous process monitoring.

Furthermore, by providing realistic time constants and operation durations derived from actual process data, PM is able to enhance the simulation capabilities of DTs. These simulations reflect the real-world almost perfectly, allowing for more accurate predictions and better decision-making. In the analysis, the most common process variants were identified and analyzed for efficiency. This revealed specific areas that needed improvement, showing how this ability to predict and simulate outcomes based on real-world data helps to prevent failures and optimize operations.

Another critical contribution of PM to the development of DTs is the ability to identify bottlenecks and deviations. The Celonis analysis drew attention to significant bottlenecks in the procurement process, such as delays between recording an invoice receipt and clearing the invoice. Companies are able to significantly reduce the throughput time and enhance efficiency by addressing the bottlenecks through automation and process optimization. Moreover, the DT should remain aligned with the ideal process model and this can be accomplished by establishing stricter controls and standard operating procedures.

DTs supported by PM are powerful tools for continuous improvement. The flexibility of a company and its responsiveness to changing conditions is extremely increased by the ability to dynamically generate and adjust production plans based on real-time data. In Chapter 2, the systematic literature review highlighted the fact that PM enables DTs to offer both predictive and prescriptive insights. These are crucial to make informed decisions regarding process adjustments, scheduling, and resource allocation. The practical analysis using Celonis demonstrated how PM could identify and address process variations and inefficiencies, thus providing a clear pathway for continuous process enhancements.

In addition, the analysis illustrated how detailed process data could be used to enhance the decision-making capabilities of DTs, enabling them to identify and address process inefficiencies in a proactive manner. This is corroborated by the SLR, which defined the role of Cognitive DTs that improve their utility in managing and optimizing manufacturing processes.

In summary, the integration of PM with DTs creates a powerful cooperation that makes it possible to enhance the accuracy, adaptability, and efficiency of business processes. PM provides detailed, real-time insights, thus ensuring that DTs can effectively mirror operations and optimize processes. This leads to significant improvements in cost efficiency, productivity, and overall system performance. The combined approach of these two technologies demonstrated the critical role of PM in the digital transformation of manufacturing and business processes.

Conclusion

The application of Digital Twins in organizations is increasingly developing and the employment of process mining to support their creation and implementation is proving highly beneficial. This thesis aims to contribute to the discussion surrounding the integration of Digital Twins and process mining to enhance organizational processes. The motivation for this study comes from the growing need for businesses to leverage advanced digital technologies in order to maintain competitiveness and operational excellence. In particular, it is compelling to analyze the intersection between these two technologies and their novel approach to operational management, which offers significant improvements in decision-making accuracy and productivity.

The main objective of this study was to explore how process mining can be of aid in the development of Digital Twins. The research question carried out in the study was "*How can process mining support the creation of Digital Twins?*" and the key aspects analyzed included the definitions of the two technologies and their history and applications, as well as the benefits of integrating Digital Twins with process mining, and a comparison of this integration with the existing literature.

In order to carry out a detailed and comprehensive analysis, the methodology included a systematic literature review (SLR) and a practical process mining instantiation using the Celonis platform. First, with the SLR, the existing research on PM implemented DTs was gathered and analyzed in order to get a deeper understanding of the theoretical groundwork. Afterwards, the practical instantiation involved the analysis of event log data through detailed process discovery, conformance checking, and process enhancement. The objective of the employment of these methods was to uncover process inefficiencies and to demonstrate how real-time data can be leveraged to create dynamic and accurate Digital Twins. The integration of both these two approaches ensured a robust examination of the practical and theoretical aspects of the research question.

The empirical context chosen for this study focused on the Purchase-to-Pay (P2P) process of a multinational company. This particular case was chosen to demonstrate the practical application and benefits of integrating process mining with Digital Twins. The complexity and richness of the available data allowed for a comprehensive application of process mining techniques. Through the selection of this empirical context, significant insights on the practical benefits of integrating process mining with Digital Twins were achieved.

The collection of the findings revealed that process mining can significantly enhance the creation and functionality of Digital Twins by providing real-time data, improving accuracy, synchronization, and overall system optimization. This research contributes to the literature by highlighting the complementary nature of Digital Twins with process mining. While Digital Twins alone are able to provide comprehensive monitoring and optimization capabilities, the integration with process mining enhances these benefits and ensures data reliability and accuracy.

In conclusion, this thesis underlines the importance of the integration of Digital Twins with process mining in order to unlock new levels of operational excellence. By integrating these technologies, organizations can not only improve efficiency and productivity levels, but also gain strategic tools for continuous improvement and innovation. Moreover, with this integration, companies can harness the full potential of digital transformation to enhance their manufacturing and business processes.

Bibliography

1. Singh, M., Fuenmayor, E., Hinchy, E. P., Qiao, Y., Murray, N., & Devine, D. (2021). Digital twin: Origin to future. Applied System Innovation, 4(36).
2. Grieves, M. (2016). Origins of the Digital Twin concept.
3. Grieves, M. W. (2005). Product lifecycle management: The new paradigm for enterprises. International Journal of Product Development, 2(1), 71-84.
4. Gelernter, D. (1993). Mirror worlds: Or: The day software puts the universe in a shoebox. How it will happen and what it will mean. Oxford University Press.
5. Främling, K., Holmström, J., Ala-Risku, T., & Kärkkäinen, M. (2003). Product agents for handling information about physical objects. Report from the Laboratory of Information Processing Science, Series B, TKO-B 153.
6. Shafto, M., Conroy, M., Doyle, R., Glaessgen, E., Kemp, C., LeMoigne, J., & Wang, L. (2010). Draft modeling, simulation, information technology & processing roadmap. Technology Area, 11.
7. Madni, A. M., Madni, C. C., & Lucero, S. D. (2019). Leveraging digital twin technology in model-based systems engineering. Systems, 7(1), 7.
8. Kritzinger, W., Karner, M., Traar, G., Henjes, J., & Sihn, W. (2018). Digital twin in manufacturing: A categorical literature review and classification. IFAC-PapersOnLine, 51(1), 1016–1022.
9. Grieves, M., & Vickers, J. (2017). Digital twin: Mitigating unpredictable, undesirable emergent behavior in complex systems. In K. Fischer & M. Varga (Eds.), Transdisciplinary perspectives on complex systems (pp. 85–113). Springer.
10. Tao, F., Cheng, J., Qi, Q., Zhang, M., Zhang, H., & Sui, F. (2018). Digital twin-driven product design, manufacturing and service with big data. International Journal of Advanced Manufacturing Technology, 94(9-12), 3563–3576.
11. Gartner Survey Reveals 47% of Organizations Will Increase Investments in IoT Despite the Impact of COVID-19. (2020). Available online: https://www.gartner.com/en/newsroom/press-releases/2020-10-29-gartner-survey-reveals-47-percent-of-organizations-will-increase-investments-in-iot-despite-the-impact-of-covid-19-, consulted on April 19th 2024.

12. Simchenko, N., Tsohla, S., & Chyvatkin, P. (2019). IoT & digital twins concept integration effects on supply chain strategy: Challenges and effects. International Journal of Supply Chain Management, 8(3), 803.
13. Gabor, T., Belzner, L., Kiermeier, M., Beck, M. T., & Neitz, A. (2016). A simulation-based architecture for smart cyber-physical systems. In Proceedings of the 2016 IEEE International Conference on Autonomic Computing (ICAC) (pp. 374–379). Wuerzburg, Germany, 17-22 July 2016.
14. Ogewell, V. (2018). Digital Twins: Beware of Naive Faith in Simplicity. Available online: https://www.engineering.com/PLMERP/ArticleID/16272/Digital-Twins-Beware-of-Naive-Faith-in-Simplicity.aspx, consulted on April 19th 2024.
15. Singh, S., Shehab, E., Higgins, N., Fowler, K., Tomiyama, T., & Fowler, C. (2018). Challenges of digital twin in high value manufacturing (SAE Technical Paper No. 0148-7191). SAE International.
16. Croatti, A., Gabellini, M., Montagna, S., & Ricci, A. (2020). On the integration of agents and digital twins in healthcare. Journal of Medical Systems, 44(161).
17. Rosen, R., von Wichert, G., Lo, G., & Bettenhausen, K. D. (2015). About the importance of autonomy and digital twins for the future of manufacturing. IFAC-PapersOnLine, 48(3), 567–572.
18. Lee, J., Lapira, E., Yang, S., & Kao, H. (2013). Predictive manufacturing system - Trends of next-generation production systems. IFAC Proceedings Volumes, 46(7), 150–156.
19. Bambura, R., Šolc, M., Dado, M., & Rakyta, M. (2020). Implementation of digital twin for engine block manufacturing processes. Applied Sciences, 10(18), 6578.
20. Pylianidis, C., Osinga, S., & Athanasiadis, I. N. (2021). Introducing digital twins to agriculture. Computers and Electronics in Agriculture, 184, 105942.
21. Fuller, A., Fan, Z., Day, C., & Barlow, C. (2020). Digital twin: Enabling technologies, challenges and open research. IEEE Access, 8, 108952–108971.
22. Ivanov, S., Nikolskaya, K., Radchenko, G., Sokolinsky, L., & Zymbler, M. (2020). Digital twins of city: Concept overview. In 2020 Global Smart Industry Conference (GloSIC) (pp. 178–186). Chelyabinsk, Russia.

23. Gartner Process Mining. (2020). Available online: https://www.gartner.com/en/information-technology/glossary/process-mining#:~:text=Process%20mining%20is%20a%20technique,models%20from%20an%20event%20log, consulted on April 22nd 2024.
24. van der Aalst, W. M. P. (2022). Process mining: A 360 degree overview. In W. M. P. van der Aalst & J. Carmona (Eds.), Process mining handbook. Lecture Notes in Business Information Processing (Vol. 448). Springer, Cham.
25. van der Aalst, W. M. P. (2016). Process mining: Data science in action. Springer, Heidelberg.
26. vom Brocke, J., Mendling, J., Wieringa, R. J., Eggert, A., Meyer, A., Hu, B., & Rosemann, M. (2021). Process science: The interdisciplinary study of continuous change. SSRN.
27. van der Aalst, W. M. P., Adriansyah, A., de Medeiros, A. K. A., Gunther, C. W., Rozinat, A., Rubin, V., & van Zelst, S. (2012). Process mining manifesto. In F. Daniel, K. Barkaoui, & S. Dustdar (Eds.), Business process management workshops. BPM 2011. Lecture Notes in Business Information Processing (Vol. 99, pp. 169-194). Springer, Berlin, Heidelberg.
28. van der Aalst, W. M. P. (2011). Process mining: Discovery, conformance and enhancement of business processes. Springer, Berlin.
29. Badakhshan, P., Wurm, B., Grisold, T., Geyer-Klingeberg, J., Mendling, J., & vom Brocke, J. (2022). Creating business value with process mining. The Journal of Strategic Information Systems, 31(4), 101745.
30. Garcia, C. d. S., Meincheim, A., Faria Junior, E. R., Dallagassa, M. R., Sato, D. M. V., Carvalho, D. R., Santos, E. A. P., & Scalabrin, E. E. (2019). Process mining techniques and applications – A systematic mapping study. Expert Systems with Applications, 133, 260-295.
31. Munoz-Gama, J., Martin, N., Fernandez-Llatas, C., Johnson, O. A., Sepúlveda, M., Helm, E., & Bernini, C. (2022). Process mining for healthcare: Characteristics and challenges. Journal of Biomedical Informatics, 127, 104001.
32. Lemos, A. M., Sabino, C. C., Lima, R. M. F., & Oliveira, C. A. L. (2011). Using process mining in software development process management: A case study. In

Proceedings of the 2011 IEEE International Conference on Systems, Man, and Cybernetics (pp. 1181-1186). Anchorage, AK, USA.

33. Gupta, M. (2014). Nirikshan: process mining software repositories to identify inefficiencies, imperfections, and enhance existing process capabilities. Companion proceedings of the 36th international conference on software engineering - ICSE companion 2014.
34. Karray, M.-H., Chebel-Morello, B., & Zerhouni, N. (2014). PETRA: Process evolution using a TRAce-based system on a maintenance platform. Knowledge-Based Systems, 68, 21–39.
35. Ruschel, E., Santos, E. A. P., & de Freitas Rocha Loures, E. (2017). Mining shop-floor data for preventive maintenance management: Integrating probabilistic and predictive models. Procedia Manufacturing, 11, 1127–1134.
36. Conforti, R., de Leoni, M., Rosa, M. L., van der Aalst, W. M., & ter Hofstede, A. H. (2015). A recommendation system for predicting risks across multiple business process instances. Decision Support Systems, 69, 1–19.
37. van Eck, M. L., Sidorova, N., & van der Aalst, W. M. (2017). Guided interaction exploration in artifact-centric process models. 2017 IEEE 19th conference on business informatics (CBI).
38. Bayomie, D., Revoredo, K., Bachhofner, S., Kurniawan, K., Kiesling, E., & Mendling, J. (2022). Analyzing Manufacturing Process By Enabling Process Mining on Sensor Data. PoEM'2022 Workshops and Models at Work Papers.
39. Novák, P., & Vyskočil, J. (2022). Digitalized Automation Engineering of Industry 4.0 Production Systems and Their Tight Cooperation with Digital Twins. Processes, 10(2), 404.
40. Schumacher, M., Buschermöhle, R., Haak, L., Höfinghoff, M., Seipolt, A., & Korn, G. (2023). Enhancing Digital Twins for Production through Process Mining Techniques: A Literature Review.
41. Dorofeev, A., Kurganov, V., Filippova, N., Petrov, A., Zakharov, D., & Iarkov, S. (2024). Improving Transportation Management Systems (TMSs) Based on the Concept of Digital Twins of an Organization. Applied Sciences, 14(1330).

42. Yang, M., Moon, J., Jeong, J., Sin, S., & Kim, J. (2022). "A Novel Embedding Model Based on a Transition System for Building Industry-Collaborative Digital Twin." Applied Sciences, 12(2), 553.
43. Park, K. T., Lee, S. H., & Noh, S. D. (2022). Information fusion and systematic logic library generation methods for self-configuration of autonomous digital twin. Journal of Intelligent Manufacturing, 33, 2409–2439.
44. Lv, Z. (2023). Digital Twins in Industry 5.0. Research, 6, Article 0071.
45. De Benedictis, A., Flammini, F., Mazzocca, N., Somma, A., & Vitale, F. (2023). Digital Twins for Anomaly Detection in the Industrial Internet of Things: Conceptual Architecture and Proof-of-Concept. IEEE Transactions on Industrial Informatics, 19(12), 11553-11563.
46. Lugaresi, G., & Matta, A. (2023). Automated digital twin generation of manufacturing systems with complex material flows: graph model completion. Computers in Industry, 151, 103977.
47. Kumbhar, M., Ng, A. H. C., & Bandaru, S. (2023). A digital twin based framework for detection, diagnosis, and improvement of throughput bottlenecks. Journal of Manufacturing Systems, 66, 92-106.
48. Bano, D., Michael, J., Rumpe, B., Varga, S., & Weske, M. (2022). Process-aware digital twin cockpit synthesis from event logs. Journal of Computer Languages, 70, 101121.
49. Friederich, J., Francis, D.P., Lazarova-Molnar, S., & Mohamed, N. (2022). A framework for data-driven digital twins of smart manufacturing systems. Computers in Industry, 136, 103586.
50. Lugaresi, G., & Matta, A. (2021). Automated manufacturing system discovery and digital twin generation. Journal of Manufacturing Systems, 59, 51-66.
51. Pan, Y., & Zhang, L. (2021). A BIM-data mining integrated digital twin framework for advanced project management. Automation in Construction, 124, 103564.
52. Vial, G. (2019). Understanding digital transformation: A review and a research agenda. The Journal of Strategic Information Systems, 28(2), 118-144.
53. van Dongen, B.F. (2019). Dataset BPI Challenge 2019. 4TU.Centre for Research Data. https://doi.org/10.4121/uuid:d06aff4b-79f0-45e6-8ec8-e19730c248f1