

CHAPTER ONE

A BRIEF INTRODUCTION TO PROCESS MINING

1. Introduction:

Nowadays data is a big oil field, and this is illustrated by the incredible amounts of data that we are collecting and its corresponding values. For that reason, a lot of companies are racing each other to get this information and this big data's value.

In this project, we illustrate how process mining use event data to discover process, check compliance, analyze bottlenecks, compare process variants, and suggest improvements.

2. Bridging the gap between Process science and data science:

Process mining is one of the essential ingredients of data science, and we notice the absence of the process perspective in many big data initiatives and data science curricula, so the process mining is used to bridge the gap between data science and process science.

2.1 process science:

It is the domain that gathers knowledge from information technology and knowledge from management sciences to improve and run operational processes, and he Consists of several disciplines:

stochastic:

The word stochastic is an adjective in English that describes something that was randomly determined. The word first appeared in English to describe a mathematical object called a stochastic process, but now in mathematics the terms stochastic process and *random process* are considered interchangeable. The word, with its current definition meaning random, came from German, but it originally came from Greek $\sigma\tau\acute{o}\chi\omicron\varsigma$ (*stokhos*), meaning 'aim, guess' The term *stochastic* is used in many different fields, particularly where stochastic or random processes are used to represent systems or phenomena that seem to change in a random way. Examples of such fields include the physical sciences such as biology, chemistry neuroscience, and physics as well as technology and engineering fields such as image processing, signal processing, information theory , computer science , cryptography and telecommunications. It is also used in finance, due to seemingly random changes in financial markets. [19]

Optimization:

Techniques aim to provide a “best” alternative (cheapest or fastest) from a large or even

infinite set of alternatives. Consider, for example, the following question: Given a list of cities and the distances between each pair of cities, what is a shortest possible route that visits each city exactly once and returns to the origin city? Numerous optimization techniques have been developed to answer such questions as efficient as possible. Well-known approaches include Linear Programming (LP), Integer Linear Programming (ILP), constraint satisfaction, and dynamic programming. [1]

Operations management & research deals with the design:

control and management of products, processes, services and supply chains. Operations Research (OR) tends to focus on the analysis of mathematical models. Operations Management (OM) is closer to industrial engineering and business administration [2].

Business process management:

Is a discipline in operations management that uses various methods to discover, model, analyze, measure, improve, optimize, and automate business processes, BPM focuses on improving corporate performance by managing business processes. Any combination of methods used to manage a company's business processes is BPM. Processes can be structured and repeatable or unstructured and variable. Though not required, enabling technologies are often used with BPM.

It can be differentiated from program management in that program management is concerned with managing a group of inter-dependent projects. From another viewpoint, process management includes program management. In project management, process management is the use of a repeatable process to improve the outcome of the project.

As an approach, BPM sees processes as important assets of an organization that must be understood, managed, and developed to announce and deliver value-added products and services to clients or customers. This approach closely resembles other total quality management or continual improvement process methodologies. ISO 9000 promotes the process approach to managing an organization. Promotes the adoption of a process approach when developing, implementing and improving the effectiveness of a quality management system, to enhance customer satisfaction by meeting customer requirements.

BPM proponents also claim that this approach can be supported, or enabled, through technology. As such, many BPM articles and scholars frequently discuss BPM from one of two viewpoints: people and/or technology. [10]

Process mining:

Process mining is a family of techniques in the field of process management that support the analysis of business processes based on event logs. During process mining, specialized data

mining algorithms are applied to event log data in order to identify trends, patterns and details contained in event logs recorded by an information system. Process mining aims to improve process efficiency and understanding of processes. Process mining is also known as *Automated Business Process Discovery* (ABPD). However, in academic literature the term *Automated Business Process Discovery* is used in a narrower sense to refer specifically to techniques that take as input an event log and produce as output a business process model. [16] The term Process Mining is used in a broader setting to refer not only to techniques for discovering process models, but also techniques for business process conformance and performance analysis based on event logs.

Business process improvement:

Is an umbrella term for a variety of approaches aiming at process improvement? Examples are Total Quality Management (TQM), Kaizen, (Lean) Six Sigma, Theory of Constraints (TOC), and Business Process Reengineering (BPR). Note that most of the ingredients ultimately aim at process improvement, thus making the term business process improvement rather unspecific. One could argue that the whole of process science aims to improve processes.

Process automation & workflow management focuses on the development:

of information systems supporting operational business processes including the routing and distribution of work. Workflow Management (WFM) systems are model-driven, i.e., a process model suffices to configure the information system and run the process. As a result, a process can be changed by modifying the corresponding process model [2].

Formal methods & concurrency theory:

in computer science, specifically software engineering and hardware engineering, formal methods are a particular kind of mathematically based techniques for the specification, development and verification of software and hardware systems. The use of formal methods for software and hardware design is motivated by the expectation that, as in other engineering disciplines, performing appropriate mathematical analysis can contribute to the reliability and robustness of a design. Formal methods are best described as the application of a fairly broad variety of theoretical computer science fundamentals, in particular logic calculi, formal languages, automata theory, discrete event dynamic system and program semantics, but also type systems and algebraic data types to problems in software and hardware specification and verification. [14].

2.2 Data science:**Data mining:**

Data mining is the process of discovering patterns in large data sets involving methods at

the intersection of machine learning, statistics, and database systems. It is an essential process where intelligent methods are applied to extract data patterns. It is an interdisciplinary subfield of computer science. The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use. Aside from the raw analysis step, it involves database and data management aspects, data pre-processing, model and inference considerations, interestingness metrics, complexity considerations, post-processing of discovered structures, visualization, and online updating. Data mining is the analysis step of the "knowledge discovery in databases" process, or KDD. The term is a misnomer, because the goal is the extraction of patterns and knowledge from large amounts of data, not the extraction (*mining*) of data itself. [16] It also is a buzzword and is frequently applied to any form of large-scale data or information processing (collection, extraction, warehousing, analysis, and statistics) as well as any application of computer decision support system, including artificial intelligence, machine learning, and business intelligence.

The actual data mining task is the semi-automatic or automatic analysis of large quantities of data to extract previously unknown, interesting patterns such as groups of data records (cluster analysis), unusual records (anomaly detection), and dependencies (association rule mining, sequential pattern mining). This usually involves using database techniques such as spatial indices. These patterns can then be seen as a kind of summary of the input data, and may be used in further analysis or, for example, in machine learning and predictive analytics. For example, the data mining step might identify multiple groups in the data, which can then be used to obtain more accurate prediction results by a decision support system. Neither the data collection, data preparation, [16]nor result interpretation and reporting is part of the data mining step, but do belong to the overall KDD process as additional steps, be used in creating new hypotheses to test against the larger data populations.

Algorithm:

In mathematics and computer science, an algorithm/'ælgərɪðəm/ is an unambiguous specification of how to solve a class of problems. Algorithms can perform calculation, data processing and automated reasoning tasks. An algorithm is an effective method that can be expressed within a finite amount of space and time and in a well-defined formal language for calculating a function. Starting from an initial state and initial input (perhaps empty), the instructions describe a computation that, when executed, proceeds through a finite number of well-defined successive states, eventually producing "output" and terminating at a final ending state. The transition from one state to the next is not necessarily deterministic; some algorithms, known as randomized algorithms, incorporate random input. The concept of *algorithm* has

existed for centuries and the use of the concept can be ascribed to Greek mathematicians, e.g. the sieve of Eratosthenes and Euclid's algorithm; the term *algorithm* itself derives from the 9th Century mathematician Muhammad ibn Mūsā al-Khwārizmī, Latinized 'Algorithmic'. A partial formalization of what would become the modern notion of *algorithm* began with attempts to solve the Entscheidungs problem (the "decision problem") posed by David Hilbert in 1928. Subsequent formalizations were framed as attempts to define "effective calculability" or "effective method"; those formalizations included the Kleene recursive functions of 1930, 1934 and 1935, Alonzo Church's lambda calculus of 1936, Emil post "Formulation 1" of 1936, and Alan Turing's Turing machines of 1936–7 and 1939.

Database:

A database is an organized collection of data. A relational database, more restrictively, is a collection of schemas, tables, queries, reports, views, and other elements. Database designers typically organize the data to model aspects of reality in a way that supports processes requiring information, such as (for example) modelling the availability of rooms in hotels in a way that supports finding a hotel with vacancies. Database-management system (DBMS) is a software application that interacts with end-users, other applications, and the database itself to capture and analyze data. A general-purpose DBMS allows the definition, creation, querying, update, and administration of databases.[12] A database is not generally portable across different DBMSs, but different DBMSs can interoperate by using standards such as SQL and ODBC or JDBC to allow a single application to work with more than one DBMS. Computer scientists may classify database-management systems according to the database models that they support; the most popular database systems since the 1980s have all supported the relational model - generally associated with the SQL language. Sometimes a DBMS is loosely referred to as a "database".

Predictive analytics :

Predictive analytics encompasses a variety of statistical techniques from predictive modelling, machine learning, and data mining that analyze current and historical facts to make predictions about future or otherwise unknown events.

In business, predictive models exploit patterns found in historical and transactional data to identify risks and opportunities. Models capture relationships among many factors to allow assessment of risk or potential associated with a particular set of conditions, guiding decision making for candidate transactions. The defining functional effect of these technical approaches is that predictive analytics provides a predictive score (probability) for each individual (customer, employee, healthcare patient, product SKU, vehicle, component, machine, or other

organizational unit) in order to determine, inform, or influence organizational processes that pertain across large numbers of individuals, such as in marketing, credit risk assessment, fraud detection, manufacturing, healthcare, and government operations including law enforcement. Predictive analytics is used in actuarial science, marketing, financial services, insurance, telecommunications, retail, travel, mobility, healthcare, child protection, pharmaceuticals, planning and other fields. One of the best-known applications is credit scoring, which is used throughout financial services. Scoring models process a customer's credit history, loan application, customer data, etc., in order to rank-order individuals by their likelihood of making future credit payments on time.[2]

Machine learning :

is a subset of artificial intelligence in the field of computer science that often uses statistical techniques to give computers the ability to "learn" (i.e., progressively improve performance on a specific task) with data, without being explicitly programmed.

The name *machine learning* was coined in 1959 by Arthur Samuel. Evolved from the study of pattern recognition and computational learning theory in artificial intelligence, machine learning explores the study and construction of algorithms that can learn from and make predictions on data – such algorithms overcome following strictly static program instructions by making data-driven predictions or decisions, through building a model from sample inputs. Machine learning is employed in a range of computing tasks where designing and programming explicit algorithms with good performance is difficult or infeasible; example applications include email filtering, detection of network intruders or malicious insiders working towards a data breach, optical character recognition (OCR) and computer vision. Machine learning is closely related to (and often overlaps with) computational statistics, which also focuses on prediction-making through the use of computers. It has strong ties to mathematical optimization, which delivers methods, theory and application domains to the field. Machine learning is sometimes conflated with data mining, where the latter subfield focuses more on exploratory data analysis and is known as unsupervised learning.

distributed system :

is a model in which components located on networked computers communicate and coordinate their actions by passing messages.^[1] The components interact with each other in order to achieve a common goal. Three significant characteristics of distributed systems are: concurrency of components, lack of a global clock, and independent failure of components. Examples of distributed systems vary from SOA-based systems to massively multiplayer online games to peer-to-peer applications. Computer that runs in a distributed system is called a

distributed program, and distributed programming is the process of writing such programs. There are many alternatives for the message passing mechanism, including pure HTTP, RPC-like connectors and message queues.

A goal and challenge pursued by some computer scientists and practitioners in distributed systems is location transparency; however, this goal has fallen out of favor in industry, as distributed systems are different from conventional non-distributed systems, and the differences, such as network partitions, partial system failures, and partial upgrades, cannot simply be "papered over" by attempts at "transparency" (see CAP theorem). *Distributed computing* also refers to the use of distributed systems to solve computational problems. In *distributed computing*, a problem is divided into many tasks, each of which is solved by one or more computers, which communicate with each other by message passing.

business model :

Describes the rationale of how an organization creates, delivers, and captures value, in economic, social, cultural or other contexts. The process of business-model construction forms a part of strategy. In theory and practice, the term *business model* is used for a broad range of informal and formal descriptions to represent core aspects of a business, including purpose, business process, target customers, offerings, strategies, infrastructure, organizational structures, sourcing, trading practices, and operational processes and policies including culture. Process mining brings together traditional model-based process analysis and data-centric analysis techniques. As shown in Fig 1.1, *process mining can be viewed as the link between data science and process science*. Process mining seeks the confrontation between event data (i.e., observed behavior) and process models (hand- made or discovered automatically). Mainstream data science approaches tend to be process agnostic. Data mining, statistics and machine learning techniques do not consider end-to-end process models. Process science approaches are process- centric, but often focus on modeling rather than learning from event data. The unique positioning of process mining, as sketched in Fig , makes it a powerful tool to exploit the growing availability of data for improving end-to-end processes. Process mining only recently emerged as a sub discipline of both data science and process science, but the corresponding techniques can be applied to any type of operational processes (organizations and systems). Example applications include: analyzing treatment processes in hospitals, improving customer service processes in a multinational corporation, understanding the browsing behavior of customers .

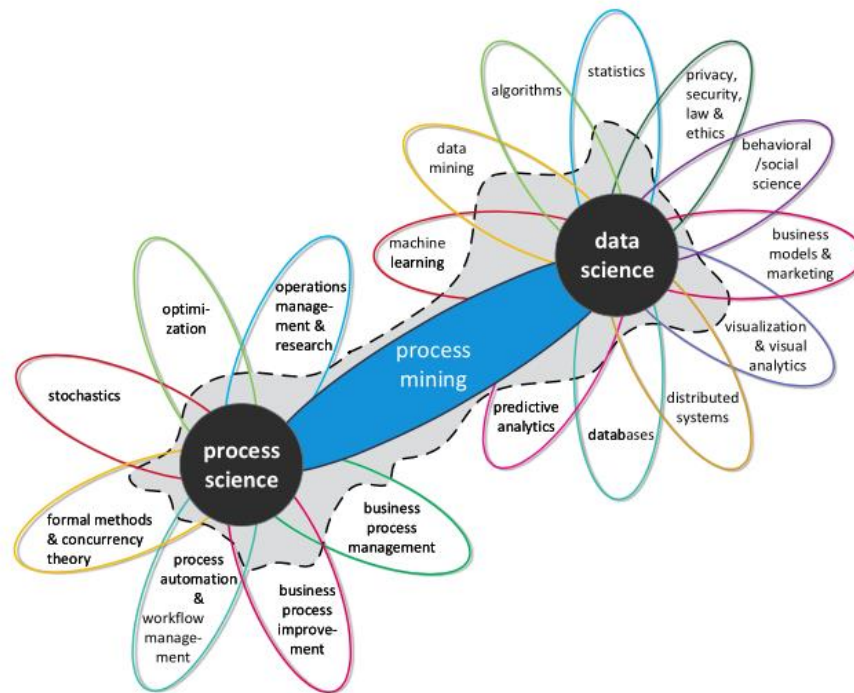


fig 1.1 : Process mining as the bridge between data science and process science[1]

3. process mining to improve the process in companies:

process mining is both data-driven and process-centric using a combination of event data and process models as a wide range of conformance questions (why are the case late, where are the real bottlenecks and which resources are overloaded) and performance questions (why are these cases late, where are the real bottlenecks), We show how to use Prom tool to answer these questions in chapters 4 .

4. Common Questions:

And This is a set of questions that managers usually have about processes in organizations:

1. What is the most frequent path in my process?
2. How is the distribution of all cases over the different paths through the process?
3. How compliant are the cases (i.e. process instances) with the deployed process models? Where are the problems? How frequent is the (non-)compliance?
4. What are the routing probabilities for each split task (XOR or OR split/join points)?
5. What is the average/minimum/maximum throughput time of cases?
6. Which paths take too much time on average? How many cases follow these routings? What are the critical sub-paths for these paths?
7. What is the average service time for each task?
8. How much time was spent between any two tasks in the process model?
9. How are the cases actually being executed?

10. What are the business rules in the process model?
11. Are the rules indeed being obeyed?
12. How many people are involved in a case?
13. What is the communication structure and dependencies among people?
14. How many transfers happen from one role to another role?
15. Who are important people in the communication flow? (the most frequent flow)
16. Who subcontract work to whom?
17. Who work on the same tasks?
18. To what extent do my cases comply with my process model?
19. What are the routing probabilities in my process?
20. What are the throughput times of my cases?
21. What are the service times for my tasks?
22. When will a case be completed?
23. How much time was spent between any two tasks in my process?
24. What are the business rules in my process, and are they being obeyed?
25. How many of my people are typically involved in a case?
26. Which people are central in my organization?

5. Process mining :

The goal of process mining is to extract information (like process models) from event logs. Typically, process mining assumes that it is possible to record events such that each event refers to an activity (a step in the process) and is related to a particular case (a process instance).

Furthermore, additional data stored in the log (like the performer of the event, the timestamp of the event, or data elements recorded with the event) can be used. The omnipresence of event logs is an important enabler of process mining: Analysis of run-time behavior is only possible if events are recorded. Fortunately, all kinds of information systems provide the necessary detailed logs, like classical workflow management systems (Staffware), ERP systems (SAP), case handling systems (FLOWer), PDM systems (Windchill), CRM systems (Microsoft Dynamics CRM), middleware (IBM WebSphere), and hospital information systems (Chipsoft). Also, all kinds of embedded systems increasingly log events, like medical systems (X-ray machines), mobile phones, car entertainment systems, production systems (e.g., wafer steppers), copiers, and sensor networks. Process mining has emerged as a way to analyze systems and their actual use based on the event logs they produce. Unlike classical data mining,

The focus of process mining is on concurrent processes instead of on static or mainly

sequential structures. Note that commercial "Business Intelligence"(BI) tools are not doing any process mining: They typically look at aggregate data (frequencies, averages, utilization, service levels). Unlike BI tools, process mining looks "inside the process" (causal dependencies, bottlenecks) and at a very refined level. In a hospital context, BI tools focus on performance indicators such as the number of knee operations, the length of waiting lists, and the success rate of surgery, where process mining is more concerned with the paths followed by individual patients and whether certain procedures are followed or not.

6.business process management life cycle :

To illustrate how process mining works we need first to describe the so-called BPM life-cycle :

the design phase, a process is designed. This model is transformed , shows, process models play a dominant role in the (re)design and configuration/implementation phases, whereas data plays a dominant role in the enactment/monitoring and diagnosis/requirements phases. The figure also lists the different ways in which process models are used (as identified in Sect. Until recently, there were few connections between the data produced while executing the process and the actual process design. In fact, in most organizations the diagnosis/requirements phase is not supported in a systematic and continuous manner. Only severe problems or major external changes will trigger another iteration of the life-cycle, and factual information about the current process is not actively used in redesign decisions. Process mining offers the possibility to truly "close" the BPM life-cycle. Data recorded by information systems can be used to provide a better view on the actual processes, i.e., deviations can be analyzed and the quality of models can be improved .Process mining is a relative young research discipline that sits between machine learning and data mining on the one hand and process modeling and analysis on the other hand. The idea of process mining is to discover, monitor and improve real processes (i.e., not assumed processes) by extracting knowledge from event logs readily available in today's systems.[1]

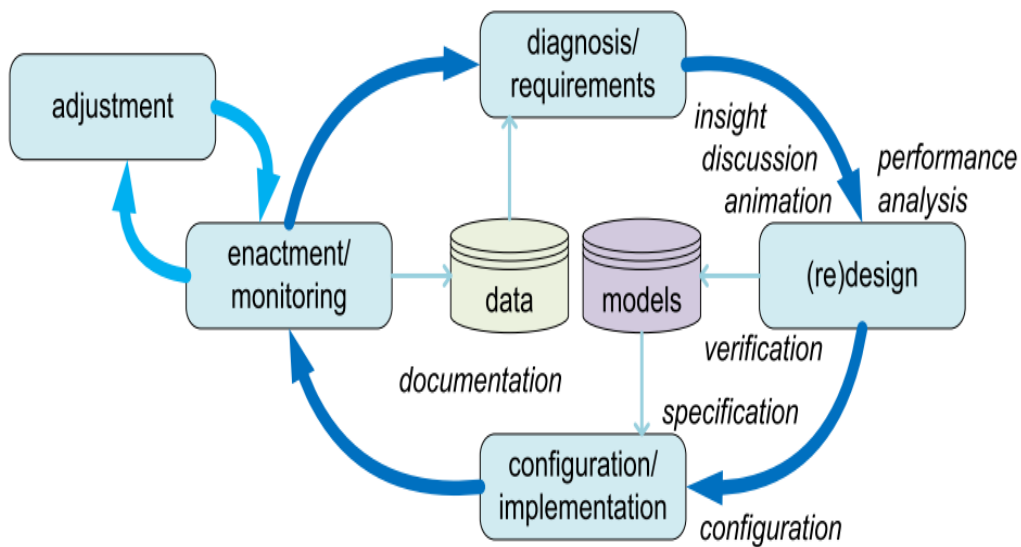


fig 1.2 The BPM life-cycle showing the different uses of process models .

7 Main applications of process mining :

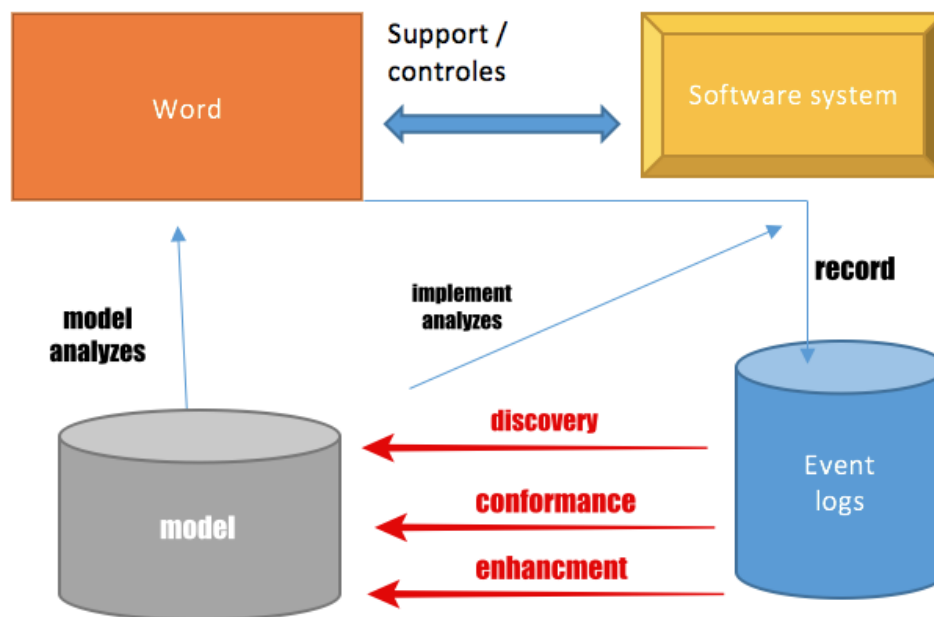


fig 1.3 Positioning of the three main types of process mining: discovery, conformance, and enhancement .

fig 1.3 illustrates the three types of process mining. The first type is discovery.

A discovery technique takes an event log and produces some kind of process model that describes the observed behavior without using any a-priori information. If the event log contains information about resources, one can also discover resource-related models, e.g., a social network showing how people work together in an organization . based on

an event log some model is constructed. For example using the α -algorithm a process model can be discovered based on low-level events. The second type of process mining is conformance checking, is the comparison of an existing process model to an event log of the same process, to check if the model corresponds to reality (as recorded in the log), or vice versa, if the implementation process is in line with the model. This makes it possible to check if the (company) rules have been respected or not. Finally, the third step is Business Process enhancement & improvement. Business Process enhancement & improvement: State-of-the-art conformance checking techniques create so-called *alignments* between the observed event data and the process model. In an alignment each case is mapped onto the closest path in the model, also in case of deviations and non-determinism. As a result, event logs can always be replayed on process models. This holds for both discovered processes and hand-made models. Next to showing deviations, alignments can also be used to reveal bottlenecks using timestamps of events. Hence, process mining provides novel forms of performance analyses that combine event data with process-centric visualizations.[2]

8-Advantages of process mining:

Objectivity: the facts do not lie; and as Event Logs reflect how a process has been implemented, the Process Discovery - the actual representation of the Claes and Poels Benefits implementation of the process - is very objective. Models reconstructed using Process Mining are not influenced by a perception that is often subjective, unlike models obtained through interviews or workshops.

Speed: Even if you still need a minimal understanding of the context in advance (only to know which Event Logs to use), it is obvious that process representation through Process Mining speeds up mapping and modeling activities. **Less effort (more efficiency):** Process discovery through process mining increases efficiency, as there is no need for more interviews or workshops with business people.

Full state: The analysis of relevant event logs makes it easier to distinguish the main process from other variants or exceptions. You will at least be able to spot variants and exceptions that you might not know without Process Mining.

Transparency: going hand in hand with objectivity, extracting the process also makes it possible to zoom in on the facts, in order to obtain more details such as who performed a transaction, what time, etc. This can help to get more transparency on how the organization really works.

Compliance: As already explained, process mining facilitates the identification of nonconformities vis-à-vis the desired or defined process.

Root-causes & bottlenecks: Process Mining allows (more) easy to discover root causes, bottlenecks, etc. .; which obviously facilitates the improvement practices (continuous) processes. Predictions & simulations: thanks to the knowledge of the facts, it is possible to predict the future behavior of a process. Event Logs also allow for even more reliable process simulations than those purely based on theoretical assumptions. This blog explains in more detail how Process Mining contributes to better process simulations.

9. Challenges of process mining:

Cost of data preparation: Even when Event Logs are available, their extraction or manipulation (for example when combining data from different systems) can be a chore.

Data quality: Depending on the source, it may be difficult to obtain Event Logs of acceptable quality

Usability of the tool: any process mining software is not user-friendly. This depends of course from one tool to another.

Limits of the tools:

some software are rather limited as regards their functionalities or algorithms.

Mastery of algorithms or software:

the good control of the software of Process Mining requires a minimum of experience and competence in the matter.

Access to data:

It is not always possible to obtain valid data. Especially when dealing with 'older' computer systems that do not produce Event Logs.

10. Analyzing an Example Log :

After providing an overview of process mining we use the event log shown in table 1.1, and the table 1.2 represent more compact representation of log shown in tab 1.2 with algorithm such as (a = *register request*, b = *examine thoroughly*, c = *examine casually*, d = *check ticket*, e = *decide*, f = *reinitiate request*, g = *pay compensation*, and h = *reject request*)

Case id	Event id	time	EVENT	EX
Case17	356234	2017-4-02 12:40:11	registration	
Case17	356235	2017-5-02 01:40:35	pay compensation	
Case17	356346	2017-12-02 13:40:20	Accepte request	
Case17	356765	2017-11-02 12:40:16	check ticket	
Case17	356679	2017-1-02 16:40:18	decide	
Case17	356796	2017-3-02 15:40:19	pay compensation	

Case17	356863	2017-2-02 18:40:11	examine causally	
Case18	356234	2017-2-02 19:40:50	registration	
Case18	356456	2017-3-02 8:40:10	decide	
Case18	356567	2017-4-02 10:40:30	pay compensation	
Case18	356345	2017-8-02 9:40:17	reject request	
Case18	356345	2017-7-02 :40:19	Examine	
Case18	356345	2017-6-02 11:40:45	check ticket	
Case18	356345	2017-5-02 11:40:55	pay compensation	
Case18	35634	2017-4-02 11:40:59	examine causally	
Case18	35664	2017-9-02 11:40:14	decide	
Case18	35678	2017-11-02 11:40:44	reject request	
Case18	35697	2017-2-02 11:40:36		
Case19	35668	2017-10-02 1:40:24	registration	
Case19	356546	2017-11-02 5:40:45	examine	
Case19	356657	2017-9-02 13:40:33	decide	
Case19	356562	2017-12-02 14:40:36	pay compensation	
Case19	356568	2017-11-02 16:40:23	check	
Case20	356634	2017-07-02 15:40:43	registration	
Case20	356768	2017-02-02 17:40:22	check ticket	
Case20	356123	2017-03-02 18:40:27	examine causally	
Case20	356456	2017-05-02 9:40:28	reject request	
Case20	356672	2017-04-02 12:40:14	decide	
Case21	356796	2017-6-02 13:40:05	registration	
Case21	356678	2017-05-02 14:40:57	pay compensation	
Case21	356789	2017-04-02 9:40:14	examine causally	
Case21	356798	2017-02-02 10:40:11	Accepert request	
Case21	356798	2017-01-02 8:40:29	compensation	
Case21	356780	2017-08-02 11:40:45	decide	
Case22	356787	2017-09-02 13:40:1	registration	
Case22	356708	2017-10-02 12:40:11	check ticket	
Case22	356709	2017-11-02 14:40:11	decide	
Case22	356798	2017-12-02 02:40:11	registration	
Case22	356698	2017-11-02 05:40:11	pay compensation	
Case22	356456	2017-01-02 08:40:11	examine causally	
Case22	356547	2017-11-02 9:40:11	decide	

Case22	356697	2017-12-02 12:40:11	reject request	
--------	--------	---------------------	----------------	-------

table 1.1 a fragment of some event log

Case id	Trace
1	<a,b,d,e>
2	<b,e,f,g>
3	<h,g,e,f>
4	<j,k,l,m,f,r,d>
5	<a,f,g,h,j>
6	<A,b,c,d,f,g,h,e,>
.....	

Table1. 2 a more compact representation of log shown in table 1 .

Process mining algorithms for process discovery can transform the information shown in Table 1.2 into process models (petri net model), using the α -algorithm that can discovers the fig 1,4 , It is easy to check that all six traces in Table 2 are possible in the model.Let us replay the trace of the first case—a,b,d,e,h—to show that the trace “fits” the model. In the initial marking shown in Fig 1,4 , a is indeed enabled because of the token in start. After firing a places c1 and c2 are marked,, both places contain a token. b is enabled at this marking and its execution results in the marking with tokens in c2 and c3. Now we have executed a,b and the sequence d,e,h remains. The next event d is indeed enabled and its execution results in the marking enabling e (tokens in places c3 and c4). Firing e results in the marking with one token in c5. This marking enables the final event h in the trace. After executing h, the case ends in the desired final marking with just a token in place end. Similarly, it can be checked that the other five traces shown in Table 1.2are also possible in the model and that all of these traces result in the marking withjust a token in place end.The Petri net shown in Fig. 1,4 also allows for traces not present in Table 2.2.For example, the traces a,d,c,e,f,b,d,e,g and a,c,d,e,f,c,d,e,f,c,d,e,f,c,d,e,f, b,d,e,g are also possible[1]

Note : The α -algorithm is just one of many possible process discovery algorithms.

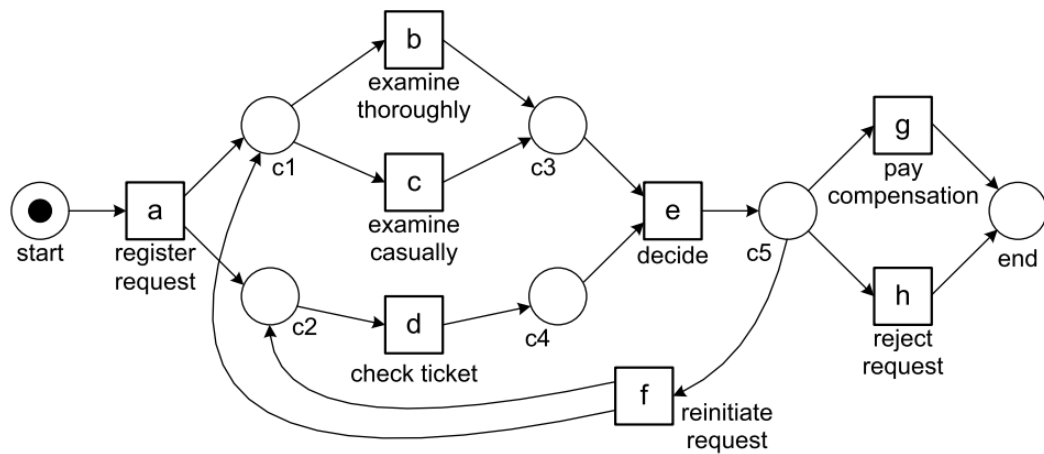


fig I,4 example for process model [2]

11. conclusion :

In conclusion Process mining is a technique that facilitates the analysis of business processes as they have been done on the basis of event logs, which can be extracted from computer systems. It is somewhat similar to Data Mining, but the focus is primarily on acquiring process knowledge. in chapter two we will present the basic concepts of process mining .