INDUSTRY INTERNSHIP

SUMMARY REPORT

"Celonis Academy-Process Mining Internship"

BACHELOR OF TECHNOLOGY

in

COMPUTER SCIENCE AND ENGINEERING

CORE -Branch

Submitted by

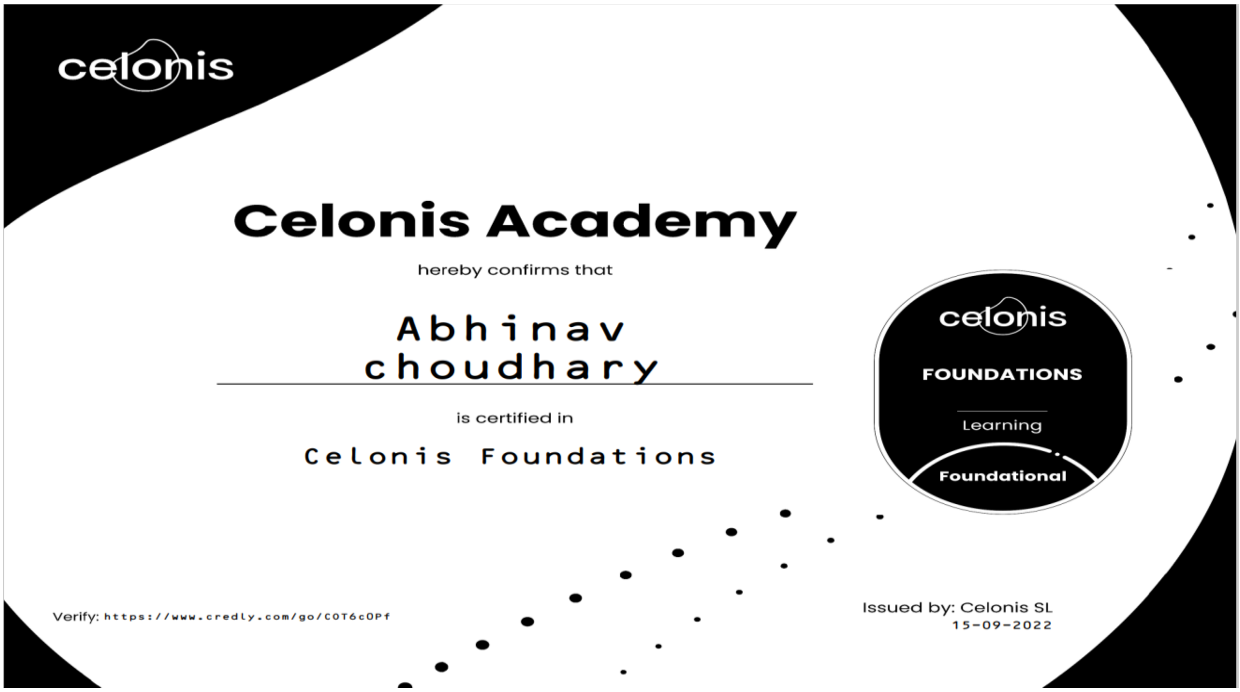
Abhinav Kumar Choudhary(21SCSE1011615)

Submitted to:

Dr. ARVIND KUMAR

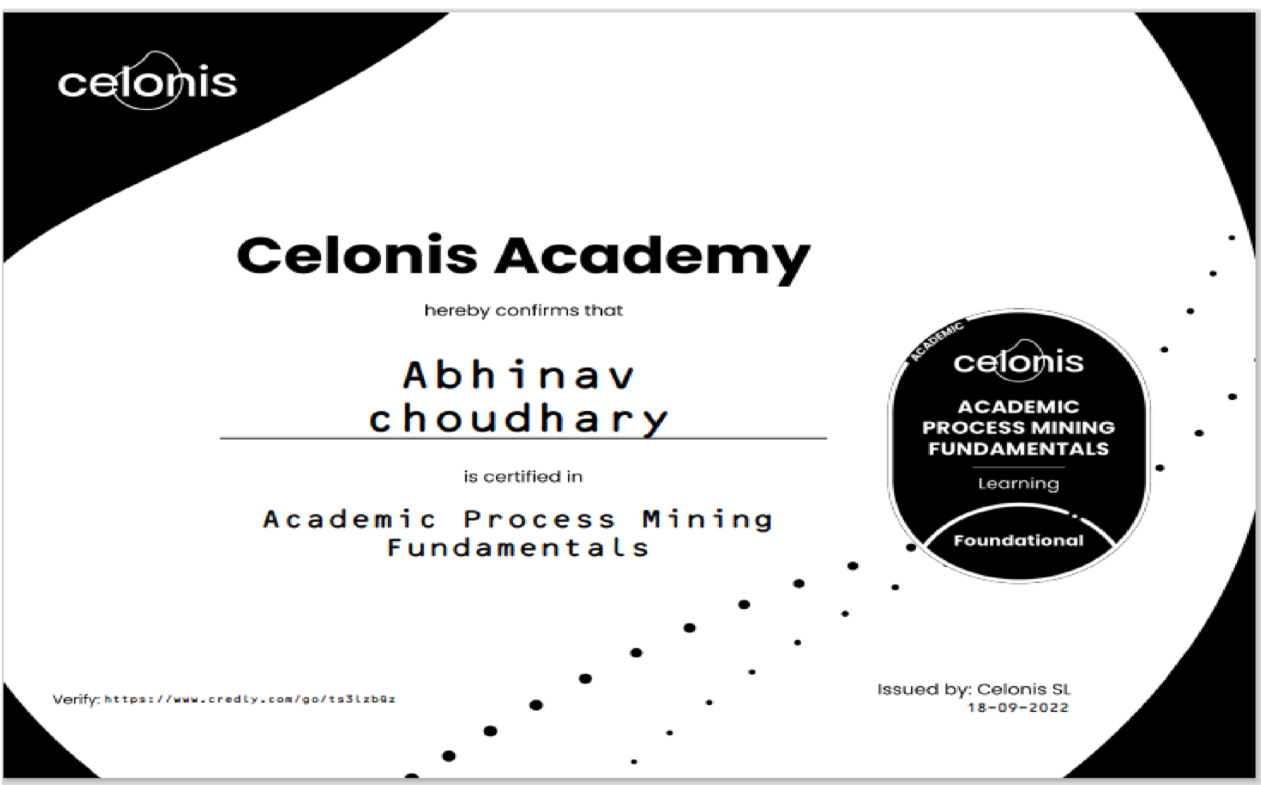
SCHOOL OF COMPUTING SCIENCE AND ENGINEERING
GREATER NOIDA, UTTAR PRADESH
Winter 2022 – 2023

INTERNSHIP CERTIFICATE



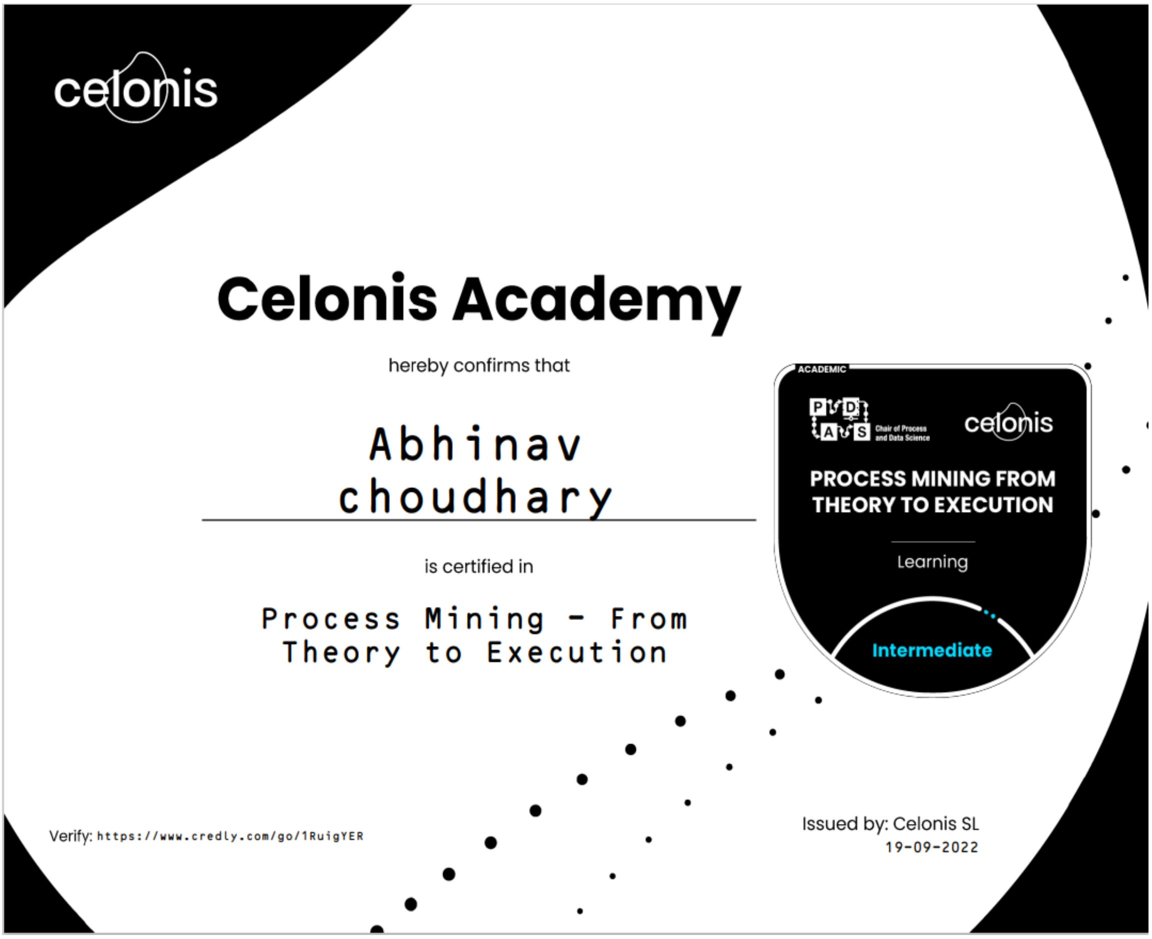
Celonis_Foundations_Badge20221122-30-1ersg96.pdf

VERIFY ME : https://www.credly.com/go/COT6c0Pf



VERIFY ME : https://www.credly.com/go/ts3lzbQz

[**Academic_Process_Mining_Fundamentals_Badge20221122-35-18xqp33.pdf**]



VERIFY ME : https://www.credly.com/go/1RuigYER

[**Process_Mining___From_Theory_to_Execution_Badge20221122-30-dvc9w0.pdf**]

CERTIFICATE

I hereby certify that the work which is being presented in the Internship project report entitled “Celonis Academy Internship <<Process Mining >> “in partial fulfillment for the requirements for the award of the degree of Bachelor of Technology in the School of Computing Science and Engineering of Galgotias University , Greater Noida, is an authentic record of my own work carried out in the industry.

To the best of my knowledge, the matter embodied in the project report has not been submitted to any other University/Institute for the award of any Degree.

Abhinav Kumar Choudhary

(21SCSE1011615)

This is to certify that the above statement made by the candidate is correct and true to the best of my knowledge.

Signature of Internship Coordinator

Dr.N.Partheeban
Professor & IIIC
School of Computing Science & Engineering
Galgotias University
Greater Noida.

Signature of Dean (SCSE)

Dr. MUNISH SABHARWAL
Professor & Dean
School of Computing Science & Engineering
Galgotias University
Greater Noida.

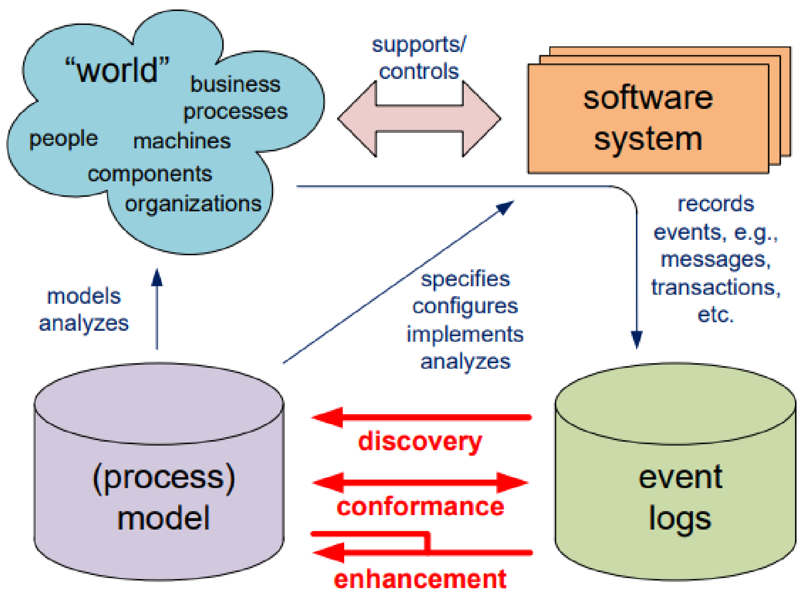
TABLE OF CONTENTS

ABSTRACT

Process models are considered to be the cornerstone of Business Process Management since they enable the documentation, analysis and improvement of business processes. Traditionally, these process models are subjective and depend strongly on perceptions. Process mining remedies the subjectivity issue by utilizing trails left by operational processes (event logs). This novel research field tightly couples event data and process models. Thus, it offers a comprehensive set of tools capable of providing fact-based insights into actual end-to-end processes. The different approaches to process mining and the plethora of tools it offers, make it applicable to a large spectrum of domains. However, blindly applying process mining to real-life, complex and highly flexible processes, proves to be inefficient. In fact, it is necessary to adapt the adopted approach to the particularities of each case study, and to frame its application by the understanding the problem at hand. This paper illustrates, through studies documented in scientific literature, that the applicability and efficiency of process mining in real-life scenarios relies on adapting the approach to their inherent challenges: Some processes might be too complex (e.g. healthcare processes), or too flexible in the absence of a system that enforces a normative model (e.g. loan applications, interaction with radiology equipment...). Event logs might present some challenges as well: Some might be extremely large, heterogenous, related to different processes or suffer from missing data. Plus, in some scenarios, we might need to account for some other aspects that traditional process mining doesn't consider (e.g. the financial aspect of processes).The papers reviewed, try to adapt to the challenges faced by defining a clear set of questions and combining process mining techniques with other analysis approaches. The papers are related to two application domains. The first is healthcare since medical procedures are too complex and sometimes involve different intertwined processes. The second is banking, finance and audit, since it is highly flexible, and in some cases, it requires the integration of the financial dimension.

INTRODUCTION

Process mining aims to discover, monitor and improve real processes by extracting knowledge from event logs readily available in today's information systems. Over the last decade there has been a spectacular growth of event data and process mining techniques have matured significantly. As a result, management trends related to process improvement and compliance can now benefit from process mining. Starting point for process mining is an event log. Each event in such a log refers to an activity (i.e., a well-defined step in some process) and is related to a particular case (i.e., a process instance). The events belonging to a case are ordered and can be seen as one "run" of the process. Event logs may store additional information about events.
In fact, whenever possible, process mining techniques use extra information such as the resource (i.e., person or device) executing or initiating the activity, the timestamp of the event, or data elements recorded with the event (e.g., the size of an order).



The three basic types of process mining: (a) discovery, (b) conformance, and (c) enhancement.

Event logs can be used to conduct three types of process mining as shown in Fig. The first type of process mining is discovery. A discovery technique takes an event log and produces a model without using any a-prior information. Process discovery is the most prominent process mining technique. For many organizations it is surprising to see that existing techniques are indeed able to discover real processes merely based on example behaviors stored in event logs. The second type of process mining is conformance. Here, an existing process model is compared with an event log of the same process. Conformance checking can be used to check if reality, as recorded in the log, conforms to the model and vice versa. The third type of process mining is enhancement. Here, the idea is to extend or improve an existing process model thereby using information about the actual process recorded in some event log. Whereas conformance checking measures the alignment between model and reality, this third type of process mining aims at changing or extending the a-prior model. For instance, by using timestamps in the event log one can extend the model to show bottlenecks, service levels, and throughput times.

Unlike traditional Business Process Management (BPM) techniques that use handmade models process mining is based on facts. Based on observed behavior recorded in event logs, intelligent techniques are used to extract knowledge. Therefore, we claim that process mining enables evidence-based BPM. Unlike existing analysis approaches, process mining is process-centric (and not data-centric), truly intelligent (learning from historic data), and fact-based (based on event data rather than opinions). Process mining is related to data mining. Whereas classical data mining techniques are mostly data-centric, process mining is process-centric. Mainstream business process modeling techniques use notations such as the Business Process Modeling Notation (BPMN), UML activity diagrams, Event-driven Process chains (EPC), and various types of Petri nets. These notations can be used model process processes with concurrency, choice, iteration, etc.

“ TECHNICAL DESCRIPTION ”

Digital event data is everywhere – in every sector, in every economy, in every organization, and in every home – and will continue to grow exponentially. The omnipresence of such data allows for new forms of process analysis, i.e., based on observed facts rather than hand-made models. Starting point for process mining is an event log. Process Mining supports this transition by delivering the necessary visibility over actual process activity, and thus highlighting problems, bottlenecks, and opportunities. To the extent that the digital enterprise runs on optimized and agile processes, Process Mining is a decisive tool to highlight opportunities for improvement and problem-resolution. The drill down capabilities of state-of-the-art Process Mining solutions uncover troublespots and provide proof of inefficiency, which supports improvement initiatives. However, Process Mining is not a panacea. Exceptions will always remain and will require specialists. In addition, while Process Mining covers the end-to-end process, it does not extend downstream or upstream, and so it is important that management gains transparency over the bigger picture to drive additional improvements. Process Mining is a highly effective tool, but it is not in itself a transformation driver, nor is this tool necessarily about automation.

SYSTEM DESIGN

#	trace
455	acdeh
191	abdeg
177	adceh
144	abdeh
111	acdeg
82	adceg
56	adbeh
47	acdefdbeh
38	adbeg
33	acdefbdeh
14	acdefbdeg
11	acdefdbeg
9	adcefcdeh
8	adcefdbeh
5	adcefbdeg
3	acdefbdefdbeg
2	adcefdbeg
2	adcefbdefbdeg
1	adcefdbefbdeh
1	adbefbdefdbeg
1	adcefdbefcdefdbeg
1391	

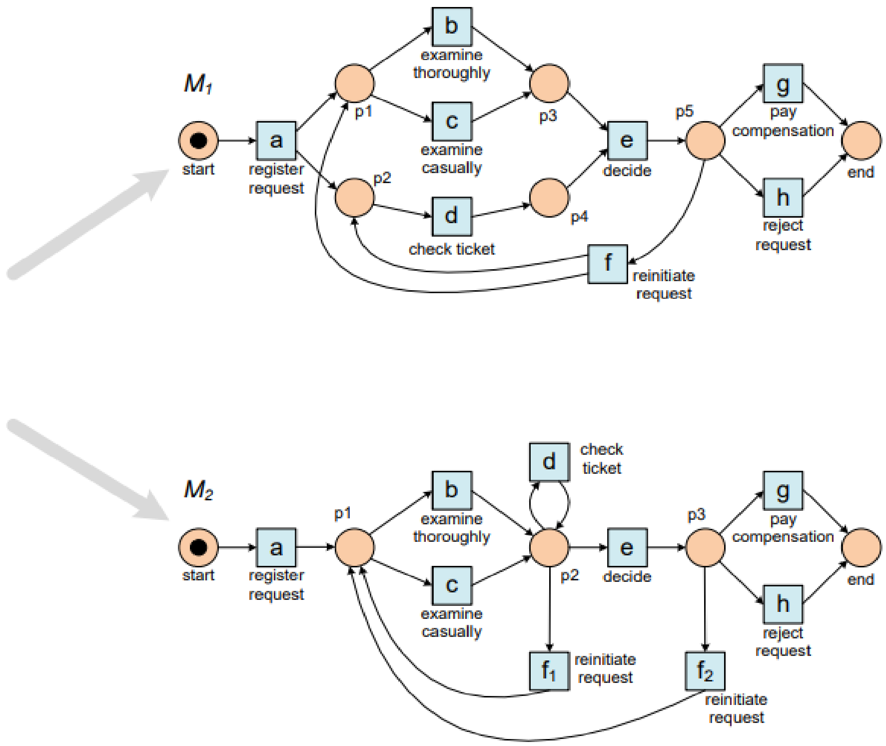


Fig. 2. One event log and two potential process models (M_1 and M_2) aiming to describe the observed behavior

“SYSTEM IMPLEMENTATION ”

As with any tool, there are best practices that support improved results. The first is to start small and then to scale up. It makes sense to collect first experiences and learn how the tool works and what exactly it can deliver before tackling a big complex process. Another aspect is managing expectations. If you start with high expectations, you run into the risk of disappointment. To stress this again, start small and learn how it works. Many organizations are leveraging Centers of Expertise that develop specialist skills in areas such as Process Mining, data analytics, or automation, and by this approach, they can push these strategies out across the organization effectively. Another advice is to consult an experienced implementation partner to ensure a best practice approach. Tapping into a partner who knows the tool inside out and who has experienced other implementations can prove invaluable in terms of avoiding costly mistakes and ensuring Process Mining is deployed effectively.

Challenges:

Process mining is an important tool for modern organizations that need to manage non-trivial operational processes. On the one hand, there is an incredible growth of event data. On the other hand, processes and information need to be aligned perfectly in order to meet requirements related to compliance, efficiency, and customer service. Despite the applicability of process mining there are still important challenges that need to be addressed; these illustrate that process mining is an emerging discipline

Some of the Most Important Process mining Challenges Identified:

C1	**Finding, Merging, and Cleaning Event Data** When extracting event data suitable for process mining several challenges need to be addressed: data may be *distributed* over a variety of sources, event data may be *incomplete*, an event log may contain *outliers*, logs may contain events at *different level of granularity*, etc.
C2	**Dealing with Complex Event Logs Having Diverse Characteristics** Event logs may have very different characteristics. Some event logs may be extremely large making them difficult to handle whereas other event logs are so small that not enough data is available to make reliable conclusions.
C3	**Creating Representative Benchmarks** Good benchmarks consisting of example data sets and representative quality criteria are needed to compare and improve the various tools and algorithms.
C4	**Dealing with Concept Drift** The process may be changing while being analyzed. Understanding such concept drifts is of prime importance for the management of processes.
C5	**Improving the Representational Bias Used for Process Discovery** A more careful and refined selection of the representational bias is needed to ensure high-quality process mining results.
C6	**Balancing Between Quality Criteria such as Fitness, Simplicity, Precision, and Generalization** There are four competing quality dimensions: (a) fitness, (b) simplicity, (c) precision, and (d) generalization. The challenge is to find models that score good in all four dimensions.
C7	**Cross-Organizational Mining** There are various use cases where event logs of multiple organizations are available for analysis. Some organizations work together to handle process instances (e.g., supply chain partners) or organizations are executing essentially the same process while sharing experiences, knowledge, or a common infrastructure. However, traditional process mining techniques typically consider one event log in one organization.
C8	**Providing Operational Support** Process mining is not restricted to off-line analysis and can also be used for online operational support. Three operational support activities can be identified: *detect*, *predict*, and *recommend*.
C9	**Combining Process Mining With Other Types of Analysis** The challenge is to combine automated process mining techniques with other analysis approaches (optimization techniques, data mining, simulation, visual analytics, etc.) to extract more insights from event data.
C10	**Improving Usability for Non-Experts** The challenge is to hide the sophisticated process mining algorithms behind user-friendly interfaces that automatically set parameters and suggest suitable types of analysis.
C11	**Improving Understandability for Non-Experts** The user may have problems understanding the output or is tempted to infer incorrect conclusions. To avoid such problems, the results should be presented using a suitable representation and the trustworthiness of the results should always be clearly indicated.

“ CONCLUSION AND FUTURE WORK ”

This paper introduced process mining as a new technology enabling evidence-based process analysis. We introduced the three basic types of process mining (discovery, conformance , and enhancement) using a small example and used some larger examples to illustrate the applicability in real-life settings. Nevertheless, there are still many open scientific challenges and most end-user organizations are not yet aware of the potential of process mining. Shared Services are constantly seeking out new solutions to identify problem areas, drive performance improvement, and leverage data analytics to learn more about the activities they run. Process Mining serves all three objectives by identifying where an ‘as is’ process is not aligned with its defined model, highlighting obvious inefficiencies – including where automation is a best fit solution. Mining event data also provides a full and real time picture of the activities running at any given time. For enterprises targeting digitization - and the transparency and agility that go with it - Process Mining

Solutions present indispensable tools. The data derived from Process Mining – data that was traditionally hidden – serves to evaluate process efficiency. The benefit of Process Mining is that it identifies weaknesses, inefficiencies and gaps that are not visible to the human eye, because they are difficult or impossible to analyze with the tools traditionally at our disposal. Process Mining offers a brand-new opportunity to analyze data and identify areas for improvement, replacing the ‘guessing game’ of the past with a bona fide value stream map offering full transparency and frictionless processes.

“ REFERENCES ”

1. Teece DJ (2010) Business Models, Business Strategy and Innovation. Long Range Planning 43(2-3): 172–194
2. 2. Tiwari A, Turner CJ, Majeed B (2008) A review of business process mining: State-of-the-art and future trends. Business Process Management Journal 14(1): 5–22
3. 3. van der Aalst WMP, Weijters AJMM (2004) Process mining: a research agenda. Process / Workflow Mining 53(3): 231–244
4. van der Aalst WMP (2011) Process mining: Discovery, conformance and enhancement of business processes. Springer, Berlin, Heidelberg, New York 15
5. van der Aalst WMP (2014) Process Mining in the Large: A Tutorial. In: Zimányi E (ed) Business Intelligence, vol 172. Springer International Publishing, Cham, pp 33–76
6. Indulska M, Green P, Recker J et al. (2009) Business process modeling: Perceived benefits. A.H.F. Laender et al. (Eds.): ER 2009, LNCS 5829: 458–471
7. van der Aalst WMP (2013) "Mine your own business": using process mining to turn big data into real value. Proceedings of the 21st European Conference on Information Systems
8. Loebbecke C, Picot A (2015) Reflections on societal and business model transformation arising from digitization and big data analytics: A research agenda. The Journal of Strategic Information Systems 24(3): 149–157
9. van der Aalst WMP, Reijers HA, Weijters AJMM et al. (2007) Business process mining: An industrial application. Information Systems 32(5): 713–730
10. Wurm B, Schmiedel T, Mendling J et al. (2018 (forthcoming)) Development of a Measurement Scale for Business Process Standardization. ECIS 2018 Research-In-Progress Papers 11. Botta-Genoulaz V, Millet PA
11. WWW.Google.com
12. WWW.celonis.com
13. WWW.Wikipedia.com
14. https.//www.researchgate.net