Process Mining in Manufacturing: A Literature Review

Yüksel Yurtay
Sakarya University, Department of Computer Engineering; Sakarya, Turkey; yyurtay@sakarya.edu.tr;
Received 22 June 2022; Revised 30 September 2022; Accepted 17 November 2022; Published online 31.12.2022

Abstract

Process mining in manufacturing is a newly expanding field of research in the application of data mining and machine learning techniques and the focus of business processes. Although it is an exciting subject of the recent past and business processes, sufficient research has not been done. Decision support systems such as enterprise resource planning, customer relationship management, and management information systems store the most valuable resource data of process details and event logs. In the advanced information systems of tomorrow, the process management, analysis, and modelling functions of modern enterprises will take their place as a necessity. As a requirement, the fundamental purpose of process mining in production is to refine data from event logs, automatically create process models, compare models with event logs, and improve and make development continuous. Our study points to the definitions of studies published in the context of production with different titles and keywords. At the same time, it contributes to the access of researchers scanning in the main title of production and brings it to their attention. It is based on the literature review and primary stages of business process mining publications in the last decade with a production focus. An overview is discussed as a roadmap for future research with meaningful results.

Keywords: production process mining, business process mining, business process management, production, literature review

1. Introduction

Businesses consume a lot of resources to execute, analyze and manage business process models. Companies that want to reduce resource consumption tend to apply techniques and approaches that can structure a more efficient operation. Owning new solutions for business processes and operations with process mining instruments is an exciting business agenda. The primary purpose of process mining is to obtain process-centred information by analyzing event logs through business information systems. In the context of the process, it makes visible the actual situation, information and the secret, curable and problematic recipes of the operation of the business. With the process view being achieved, it is seen that business processes form the centre of the organization [3]. At this point, the meaningful association of the data owned by the enterprises includes the most appropriate functioning of the organizations in the context of the process. Suppose more than one organization contributes to the operation of the process. In that case, it requires consolidation of many business processes because organizations work interactively under the pressure of interdependence and intend to create shared value. Actions such as business process management, modelling and analysis are topics of increasing interest for organizational managers oriented towards a specific goal [4].

Today's organizations use process-aware information systems (PAIS) in automated modern business environments [14]. PAIS runs on structures where complex business processes are described as a “naturally distributed system” [2]. Information systems containing many open process models, such as supply chain management systems (SCM), financial systems (FS), and customer relationship management systems (CRM), can be given as examples. The processing and analysis of data obtained through business processes in information systems are manually performed [6,9,11]. With this workload in information systems, flexibility and definition have not received enough attention, as well as the developed open process concepts [1]. As the consistency or comparison of business process models emerged as a need, the first signs of process mining began to be seen. In the recent past, process mining was recognized as one of the significant innovations in business process management in 2012[7]. Then, key concepts, manifestos, roadmap and challenges were defined as a research area [5]. The process
mining research area, which uses data mining techniques to discover automatically and dynamically consistent process models and detect anomalies and their causes, has been announced [1]. Today, process mining is considered a broad discipline that combines the meaningful association, analysis, improvement, execution, and monitoring of the data that makes up the process under the umbrella of data science with information from information technology and management sciences [5]. As a new discipline, process mining focuses on the flexibility, dynamism, comparability and possible anomaly detection of dynamic models [34]. In particular, process mining studies continue to find a place for improvement, development, monitoring and detection of problems in production organizations. In addition, it is known to provide many contributions, such as determining optimal processes, improving product quality and production processes, and easy and fast performance analysis [35].

Its contributions stand out under five headings;
- It paves the way for model discovery,
- Detection of deviations in the process,
- Implementation of conformity check based on alignment,
- Ease of diagnosis in performance improvement,
- Increase process visibility and diagnosis with findings.

Process mining studies and research, which attract the necessary attention in production organizations, continue to attract attention in the scientific field. In our scans, it is seen that the titles given to the publications do not include the phrase "production" or "process mining in production". This situation makes it difficult for researchers and decision-makers in organizations to access research and applications carried out in the context of production. The article draws attention to the distribution in the main headings to facilitate access to the publications. Our motivation is to make the gains visible by drawing attention to the titles in the classification and entry of emerging publications. In addition, the answer is found under which main headings scientific publications on process mining are compiled in the context of production. It also helps to consolidate publications in the context of production by emphasizing the importance of including the title "production". Because facilitating access to publications to be made in the context of production, approaches and similarities used in publications will accelerate and guide various production processes waiting for solutions. At the same time, it will help process mining mature for sectors/areas that have not yet been researched and applied as a new discipline.

This article provides a general framework for process mining in the production context by focusing on the fundamental concepts of the process mining research field. Process mining types, tools, techniques and event logs are the prominent topics of the article. This research study has three main contributions;
1- The topics and content are compiled in the context of production process mining.
2- It presents the processing of event logs and the grouping of tools in production process mining.
3- The interest or analysis of the organizations engaged in production within the process mining framework is shared.

Process mining application points and literature reviews in production are shared. The second part of the study explains the basic concepts and current studies of process mining. The third section describes the form of the literature study, while the fourth section presents the literature research results and discussion. The fifth section includes the results evaluation and future productions.

2. Business process mining: an overview

Organizations' workflows have changed significantly with the innovations in computers and communication. Changing business processes brought complexity and increased the need for information systems. Visualization is critical to improving the understandability of business processes. Key actions of business process mining predict performance, provide insights, discuss responsibilities and analyze compliance. At the same time, it stands out as a new research area defined and promoted through process mining. Therefore, it most closely defines business process mining manifests of process
The primary function of process mining is to view the actual process and discover and improve it with the help of the analysis of event logs obtained from existing information systems. In particular, a business process is a cross-section of any case or process instance. The activity is the executing transaction part of the business process segment. An event is a well-defined step in a process in a particular situation. The event is the first assumption that refers to the activity or task in the process example. The second assumption is that the events in question are sequential. Event logs contain information about incidents and activities and store information about employees, devices, timestamps, and contributors in the data environment. Business process mining aims to bridge the gap between business process management (BPM), and workflow management (WFM) approaches on the one hand and data mining (DM), Machine Learning (ML) and business intelligence (BI) on the other. Process mining focuses on end-to-end process models, while DM, ML and BI concentrate on data. The results can detect/diagnose inefficiencies, performance, bottlenecks, risks and deviations. All disciplines affecting these points will, directly and indirectly, benefit from the process mining results. Institutions that improve and explore their processes will increase competitiveness while obtaining performance and productivity contributions. The following sections briefly share the stages, types, perspectives, tools, techniques, and main problems of process mining in achieving competitive advantage.

Stages of process mining:
- Extracting process models from event logs or performing automatic process discovery,
- Comparison of model and event logs, detection of unusual executions and compliance control,
- With the help of new models, techniques and technologies, steps can be taken for improvement and forward development.

In Figure 1, the context of process mining is shared with its significant concepts. Data such as transactions (messages, work orders, etc.) and event logs of business processes in daily life are stored. Data is filtered, cleaned and extracted. The software performs the desired process mining (discovery, compliance and development) steps.

Knowledge systems are the working plane of mining and automation Technologies. Process mining activities include data extraction, filtering and cleaning from information systems. Because, considering that data can be shared with various sources, it is necessary to correct existing data anomalies. Therefore, a set of manifests of event data is an implementation directive. In Figure 2, the principles to be considered in practice are below.
Figure 2 provides a high-level view of the information system to the success achieved. The process mining flow is defined as the upper level, from acquiring event data by inference from information systems to success [14]. Flow can be rearranged for unusual process flows. The guiding principles [15] to be considered in the figure are listed below.

1. The quality of the process mining result is related mainly to the data and its quality; it should be acted on in this context.
2. Extracting meaningful event data with concrete questions is a requirement.
3. Basic control flow (concurrency, selection, etc.) structures should be supported.
4. It should be related to events and model elements.
5. Models derived from event data provide insights into reality.
6. The continuity of process mining should be ensured.

2.1. Process mining types

It analyzes three types of event logs under process mining, discovery, compliance, and development [10].

• Discovery requires creating an event log and generating a model without prior knowledge. Typically, the model explored is a process model, similar to a Petri net, BPMN (Business Process Model and Notation), or UML (Unified Modeling Language) activity diagram. In addition, the discovered model can also describe other perspectives, such as social networking [10].

• Compliance is an audit technique that uses event logs and the model as input. Explain the similarities and differences between the output, model and event logs. The fit check can be used to show the conformity or nonconformity of reality to the model, as in recorded event logs [10]. Compliance control is used mainly to detect, explain and analyze deviations in the process [16]. Two different types of metrics are used. It is the accuracy observed in the behaviour of the process model and the degree of openness it represents [17].

• Build is a model development technique that uses an event log and model as input. The resulting output defines the enhanced and extended model. It uses to process information describing event log records
to develop and extend the existing process model [10]. The process model can be changed, corrected and demonstrated at this stage.

2.2. Process mining perspectives

Process analysis of institutions is carried out using process execution data. Process data also stores information about its execution. The process perspective reveals the details of how a process is executed. Thus, the process perspective is based on the data and activities of an event log [10]. From the process perspective, the activity sequence is the study's focus. The work also referred to as the control-flow perspective, is to characterize all possible paths with Petri net or other process notations [10]. Control flow stakeholders, people, roles, and the relational status of departments are defined from the organizational perspective. The primary purpose is to observe the event logs and look for answers to many questions such as processing time estimates, bottlenecks, service levels, and resource usage [10].

2.3. Process mining software tools and techniques

Process mining software analyzes hidden options, anomalies, and business processes using event logs. Process mining software tools primarily include process discovery, design, analysis and recovery functions. Provides feedback and recommendations for process design, improvement and development. Process analytics examines event log data to detect potential problems. Process recovery or revitalization takes the process one step further for dynamic improvement.

Examples of process mining software include ProM (Open Source, TU/e), ARIS PPM (Software AG), Icris (Icris), Process Mining, BusinessOptix, Disco (Fluxicon), and LANA Process Mining (Lana Labs).

Preferred software should provide the necessary features to implement priority use cases and processes. It should be possible to process commonly used event logs in selecting software in which the user and place of use stand out. Otherwise, it will increase the user's workload in obtaining insights.

Algorithms used in the general framework can be classified into three main categories.

1. Deterministic algorithms: Since the algorithm's output, which accepts all variable data as input, is constant, it produces repeatable models [3].

2. Heuristic algorithms are solution approaches based on insights when the algorithm cannot conclude [3].

3. Genetic algorithms: Better solution models are sought by combining features and adding random variations in operations carried out with an arbitrary starting point. Although the approach is powerful and complex, it can yield many models [18].

2.4. Process mining problems

In addition to many techniques applied in process mining, it also contains difficulties [19],[20]. In practice, the most crucial challenge of process mining is the preparation and processing of data. There are difficulties with event data such as non-process-oriented, incomplete, scattered, noisy, and mismatched timestamps. Concept shift concerning processes is another challenge [21]. Concept shift means that the process may change during the analysis. Data may vary due to changing business conditions. Choosing the appropriate data range based on changing business conditions can simplify the solution. Therefore, event data also changes due to changes in process models or information systems. For example, the content mismatch between old and new data will create a concept shift. Four different conceptual shifts are introduced as sudden, gradual, repetitive and incremental [22]. The different levels of detail of event logs is another challenge that should not be ignored [44].

Since process mining is an emerging discipline, various problems and difficulties arise during its application in business processes [10]. As process mining challenges and possible solutions is a topic of interest, the process mining manifest [10] may be needed for further reading.
Data is one of the most valuable assets of the organization's business processes. The data waiting to be evaluated added value in proportion to its impact on the entire functioning of the organization by refining potentially hidden patterns. For example, it allows them to correctly understand the processes operating during mining operations, check compliance and improve them. For organizations, the primary purpose of process mining is to obtain action-oriented information by refining event logs obtained from existing information systems. On the other hand, with the increasing amount of data, there is a need for techniques and methods to process the growing data. Case studies exist to refine the event logs of organizations built on distributed systems [24],[25]. It should be remembered that mining process models through event logs of different structured information systems can be challenging.

3. Research design and methodology
The research aims to make the current situation visible regarding process mining in production. It is to compile information about recent studies of software tools, types and perspectives with a focus on process mining within the scope of production. The Scopus and Google Scholar databases, including data sources, books, scientific journals and conference papers, were searched for the purpose. In Figure 3, the stages of the literature review process are given with the evaluation parameters.

It is seen that the process mining studies carried out in the focus of production have started to attract new attention. So much so that the studies are compiled under business process mining in scientific publication environments. Under the title of business process mining, it is categorically collected in the content of production, management and accounting. While the categories do not give clear boundaries for new work, they define a facilitating framework for research. With the increase in interest and studies, the classes will become apparent as frames defined under a separate title. The research reports the trend and development of production-oriented studies from two different source collection points.

It resulted in 2786 articles initially obtained from the SCOPUS resource. Research criteria start with process mining and narrow down to business process mining. They are reduced from 206 business process mining articles to 27 production content articles. When the production and process mining titles on the same source are searched together in keywords, 56 papers are compiled. Out of the 56 article
intersections, ten articles are obtained as a process mining study in production. Among the articles scanned in our research, no article with the title "production process mining" was found.

Secondly, the number of articles obtained from the Google SCHOLAR resource is 95. The number of essays compiled in the scanning process narrowed by business process mining is 7. Scanned papers are gathered over the title. An article is obtained by searching the keywords "production" and "process mining". It is noteworthy that the related articles, titles, keywords and content are increasing daily. There was no study on a topic with the range of process mining in production and the expression of production process mining.

In the literature review, the abstracts of the related articles were analyzed and scanned as in Table 1 and written according to their titles. In Table 1, examples of the articles that are the research subject are given.

Table 1: Sample publications summarizing the work done

<table>
<thead>
<tr>
<th>Year</th>
<th>Title</th>
<th>Authors</th>
<th>Purpose / Type of source</th>
<th>Keywords:</th>
<th>Consequitual / Theoretical framework</th>
<th>Major themes</th>
</tr>
</thead>
<tbody>
<tr>
<td>2020</td>
<td>Process mining-based anomaly detection of additive manufacturing process activities using a game theory modeling approach [27].</td>
<td>S.Saraeian etc.</td>
<td>Anomaly detection in production / Research</td>
<td>*Event-based anomaly detection *Additive manufacturing *Business process management system *Process mining technique *Game theory modeling *Distributed production system</td>
<td>Intrusion detection through the integration of process mining and game theory techniques.</td>
<td>Process improvement development in production</td>
</tr>
<tr>
<td>2017</td>
<td>ERP Post Implementation Review with Process Mining: A Case of Procurement Process [29].</td>
<td>M.Er etc.</td>
<td>Workflow process editing and development / Research</td>
<td>*Post Implementation review *Process Mining *Procurement *SAP Materials Management</td>
<td>Purchasing process definition/editing and development</td>
<td>Improvement in the purchasing process</td>
</tr>
</tbody>
</table>

Although the studies in Table 1 are process mining studies that focus on production systems, they are mostly keyed as methods or analyses used as categories.

4. Research results and discussion

It is seen that the articles studied in terms of the use of process mining are primarily and mostly control-flow, and secondly, case/time and organizational perspective. It is determined that there are studies for analyzing the expectation process that is prioritized in the field of exciting process mining [3]. Especially in process analysis, decision-makers who want to improve flow and control are representatives of different production sectors. The case/time perspective is studied for the needs of the process, such as analysis of processing, waiting, output times, discovering bottlenecks, anomalies and case frequencies. When evaluated from a higher level, studies have been carried out in the organization's title to support...
strategic decisions and obtain an organizational perspective. While many papers apply more than one perspective, no significant model combines perspectives [3].

The article types reviewed in our searches are aligned in order of discovery, development, and relevance. It is seen that researchers care more about process improvement and solutions to problems. Emerging and prominent improvement options are flow and temporal requirements. Mining studies for compliance checking are less common in scanned articles and other sources [3]. Motivation is a need to identify compliance and deviations and make recommendations. They compare the model that emerged/discovered in the work of decision-makers with the reference model, with software tools such as "ProM" and "Disco", respectively. The digitization and evaluation of the appropriate metrics take place through the reports produced by the software tools. Considering the application areas, are listed as "Healthcare", "IT", "Finance", "Manufacturing", and "Education"[3]. The ranking is predictable, considering the rapidly changing digital world's expectations put pressure on industries. Especially in the health sector, it is known that the development, improvement or personalization of the variety of services given to the patient [31], [32], [33], [34], [35], [36], [37], [38], [39], [40] has come to the fore in the recent past. The contribution of the IT industry to other fields is an undeniable fact. In this respect, the software process, positive/negative user activities and service management process continue to attract attention [41],[42],[43]. While the finance sector has attracted interest in every period, analysis [23],[45] has been focused on deposit, customer and loan service points in the recent past. At the same time, studies on security [46] continue to attract attention.

Even when the "Scopus" publication source of process mining is scanned for the last ten years, it is seen that the increasing interest continues (Chart 1). When the graph is examined, it is seen that institutions continue to apply for process mining instruments in a short time, and positive feedback is received. With the active use of software tools developed for solutions that can meet the needs of decision-makers, the application areas have differentiated. The development of the tools and the positive results in the application areas have highlighted the interest in process mining. Research results on "Scopus" and "Google Scholar" contain similar results. Below, the research results are shared via the "Scopus" publication source.

![Chart 1 Number of publications with "process mining" in the title](image)

Graph 2 of the application areas of interest in the same source and date range is shared. When the summaries of the studies appearing in Graph 2 are examined, it is seen that they are defined under different titles instead of the production title. Remarkably, "Computer Science" is continuing the development of algorithms and software tools in this new field of process mining. So much so that the modelling and development of mining techniques in all industries with processes led to publications in the areas of "Engineering", "Mathematics", and "Decision Sciences". Process mining in production is gathered under the title "Business, Management and Accounting".
Process mining solutions for production are shared under the "Business, Management and Accounting" title and the "Business" sub-title. In the "Business" heading, there are solutions for the production process and all solutions for business processes. From this point of view, the variety and number of methods in the production title will pave the way for examining process mining in production under a separate heading.

The number of studies obtained with the keywords "KEY (manufacture OR production) AND KEY (business AND process AND mining)" in the scanned articles is 56. The distribution of article subject areas is given in graph 3.

As can be seen in Graph 1, it is seen that the number of studies on "process mining" in the scanned article titles continues to increase rapidly. When the article contents are examined in Graph 2, scientific publications related to mining research are in the first three places. This output is geared towards developing tools and scientific approaches in the maturing field of "process mining". It determines the sectoral titles and boundaries of the studies following the first three lines. Remarkably, the studies defined by "Decision Sciences" are at the centre of attention of all sectors. So much so that the most effective support used in planning and management issues is sought under "decision sciences". Starting from the fourth row in Graph 2, the number of publications related to sectoral fields is shared. "Business, Management and Accounting" is one of the common sectoral areas. This title covers processes such as business, production, management and cost. As the variety and number of publications in the scope increase, new titles will emerge.
"Business, Management and Accounting" corresponds to 17.8% of the articles. Likewise, research on the content of production processes in 10 articles is collected under "Business". Although studies in the context of production are categorically under this title, the number of titles will increase with the increase in acquisition and use. Chart 4 shows the number of publications by keywords that support this situation. It is seen that the number of publications increases according to the keywords that support this situation. When queried with the term “KEY (manufacture OR production) AND KEY (business AND process AND mining)”, the number of accessible publications in the last four years is shared in graph 4.

![Chart 4 Number of publications by keywords](image)

Graph 5 shows the citation numbers of process mining publications in the context of production for the last ten years. Process mining tends to decline after the 2019 pandemic as it corresponds to the application in the field.

![Char 5: The number of citations of publications with the title "Process Mining" between 2012-2022.](image)

Looking at the last four years, research studies defined under "Business" in the context of production continue to increase. Sharing the data of production processes by decision-makers and announcing their applications will attract attention to production processes.

Therefore, after the readings, the possible benefits of use can be listed as follows, focusing on application in production processes today.
1. Measurement and optimization of machine process performance
2. Improve and better understand operations
3. Identifying automation opportunities
4. Reducing operating costs  
5. Contribution to production planning  
6. Support for resource management

Considering the listed benefits, awareness and interest in the production processes of organizations will increase. The interest turned into practice will open the door to businesses' competitive advantage in global markets.

Knowing what you have in your system is a necessary and sufficient condition for change, transformation and competitive advantage. All resources are brought together with the help of intuitive dashboards and investigative interfaces, optimizing the collected data by analyzing the differences in your systems. Ensuring the synchronization of resources with customer expectations and needs with daily business operations is a challenging task. At this point, traceable process metrics, synchronization, and catching and correcting instant performance deviations are the result of process intelligence. Also, process mining provides a quick solution to understanding, sharing and improving the current situation. On the other hand, the correlation and meaningful association of the data supports the decision points. Decisions based on data and accurate process information empower organizations against risks. Therefore, a process-oriented shift enables organizations in all sectors to compare and position their current situation as they should in the future. Research conducted in this framework and solutions in the industries will increase the acceleration of the transformation process. In this context, analysis and solutions in the sectors will contribute to the transformation process. Working in a new discipline and using new research methods has lost the sensitivity of the title of "production". The term "production", especially in the title and keywords, will facilitate access to publications. At the same time, it was evaluated that the publications being easily accessible or found would contribute positively to the search and transformation process of the sectors.

5. Conclusion

Adequate information on performance, optimization, improvement in production processes, detection of downtime/faults and a better understanding of the process are among the indicators for the purpose. How to analyze actionable operational data to achieve the goal. How to provide end-to-end visibility into your system that hosts your crucial process metrics and key performance indicators (KPIs). How to analyze baseline models automatically. How to get the real reasons and compare the variables in the poorly performing processes. These and many more questions are among the questions that the manufacturing sectors urgently need to find answers to. One of the easy ways to reach answers and solutions is to access scientific publications. Our study draws attention to the definitions in titles and keywords in accessing publications. The article draws a general picture in which research and studies carried out within the scope of production are compiled.

In the article, it is given under which headings the publications are distributed in the production field to increase the gains in mining works and production processes. While this distribution includes new methods and subheadings, the term "production" is not included in the top heading. Our study shows the place, status and result of production processes in process mining research areas quantitatively. At the same time, emphasis is placed on using the term "production" in developing mining studies in the context of access. The importance and conceptual framework of process mining are shared in the first stage. The second stage discusses the scope and development of business process mining. To answer our question, it is seen that the publications are classified under identical/similar titles in two different scientific databases. Both database sources show that the publications are classified under "Business, Management and Accounting" instead of "Production". The results obtained share limited examples of scientific publications. Apart from production, there is a need for research in the known titles of different sectors. Research needs more databases and methods for a general description. The results obtained in our study were scanned, evaluated and limited based on abstract, title and keywords. While the publications specific to the scope of process mining in production are compiled under the title of "business", they also include different business process publications. The results of changing the scan
sequence in Charts 3 and 4 show that process mining efforts are increasing in business and production. The works that emerge with a more accurate title definition or keywords will be more visible in the main title of the paper. Making the definitions correct in the classification of the studies, which is the article's main subject, will open the way for easier access to the researchers. As a result, it is evaluated that the remarkable process mining studies will continue to grow in production as the quality and access to mining techniques and scientific approaches increase.

In future studies, it is noteworthy that mining studies and research of the production processes of enterprises with big data infrastructures, whose benefits we clearly show, are a niche area.

Three critical benefits, in particular, attract the attention of relevant employees.
1. Critical workflows and operational processes: User error is minimized by improving continuity and business process performance. Continuity in improvement is ensured by process mining.
2. Adaptive technology: Warning and action are taken before potential bottlenecks occur. The automation potential can be determined, e.g. batch processing.
3. Scalable: It can manage millions of events, tasks, multiple processes and workflows. A digital twin can be created at an appropriate scale.

The automation era in production has brought industries to the brink of extensive data analysis. For this reason, the gaps between organizations’ physical and digital structures make it difficult for developing institutions to mine. However, developments continue rapidly in creating dynamic manufacturing businesses capable of objective decision-making based on interconnected facts. Researches and solutions on process mining in production provide practical benefits in the digital twin of organizations, intelligent process management and mining studies.

It will take place among the decision support tools quickly if comprehensive and high-quality process mining research is carried out in production. Interagency mining research is another interesting topic. In addition, process mining over the Internet and refining process and event data in real-time are other research subjects.

References


