

Article

# New Approach to the Analysis of Manufacturing Processes with the Support of Data Science

Martin Krajčovič <sup>1</sup>, Vsevolod Bastiuchenko <sup>1,\*</sup>, Beáta Furmannová <sup>1</sup>, Milan Botka <sup>2</sup> and Dávid Komačka <sup>1</sup>

<sup>1</sup> Department of Industrial Engineering, Faculty of Mechanical Engineering, University of Žilina, Univerzitná 8215/1, 010 26 Žilina, Slovakia; martin.krajcovic@fstroj.uniza.sk (M.K.); beata.furmannova@fstroj.uniza.sk (B.F.); david.komacka@fstroj.uniza.sk (D.K.)

<sup>2</sup> Asseco CEIT a.s., Univerzitná 8661/6A, 010 08 Žilina, Slovakia; milan.botka@asseco-ceit.com

\* Correspondence: vsevolod.bastiuchenko@fstroj.uniza.sk

**Abstract:** This article introduces process mining as an innovative approach to enterprise data analysis, offering a systematic method for extracting, analyzing, and visualizing digital traces within information systems. The technique establishes connections within data, forming intricate process maps that serve as a foundation for the comprehensive analysis, interpretation, and enhancement of internal business processes. The article presents a methodical procedure designed to analyze processes using process mining. This methodology was validated through a case study conducted in the Fluxicon Disco software (version 3.6.7) application environment. The primary objective of this study was to propose and practically validate a methodical procedure applied to industrial practice data. Focusing on the evaluation and optimization of manufacturing processes, the study explored the integration of a software tool to enhance efficiency. The article highlights key trends in the field, providing valuable insights into process flows and identifying areas for improvement. The results contribute to the growing body of knowledge in process mining, emphasizing its applicability in fostering a more efficient and competitive manufacturing environment. In the model example, we successfully achieved a reduction in the time required for production cycles by 15% and improved resource utilization by 20%. This resulted in an increased process efficiency and a potential reduction in the required number of workers by up to 10%. These outcomes offer promising evidence of the advantages of our method and its application in an industrial setting.

**Keywords:** process mining; process maps; data analysis; information systems; methodical procedure



**Citation:** Krajčovič, M.; Bastiuchenko, V.; Furmannová, B.; Botka, M.; Komačka, D. New Approach to the Analysis of Manufacturing Processes with the Support of Data Science. *Processes* **2024**, *12*, 449. <https://doi.org/10.3390/pr12030449>

Academic Editor: Diane Mynors

Received: 1 February 2024

Revised: 17 February 2024

Accepted: 20 February 2024

Published: 23 February 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

In today's environment, where the dynamics and competition in manufacturing are continually increasing, it is essential to apply advanced analytical methods before implementing any process changes. For instance, the simulation modeling of processes, conducted before their implementation, allows for a detailed prediction of the impact of proposed changes and the identification of potential waste and inefficient resource use, thereby ensuring their effective elimination and optimization. This approach, supported by data science and process mining technologies, forms the basis for strategic decision-making and enhancing the efficiency and competitiveness of manufacturing processes. The proper management and optimization of manufacturing and logistics processes can lead to increased productivity, cost reduction, and improvement in product quality. However, traditional analytical methods may be limited in their ability to process a large volume of data and handle the complexity of modern manufacturing systems. In this scenario, it is essential to use advanced data science technologies to gain a more detailed understanding of manufacturing processes and identify areas for improvement. One such technology is process mining, which brings an innovative approach to the analysis of manufacturing processes, supported by data science [1,2].

Existing process mining techniques provide a strong foundation for process analysis and optimization but face several challenges:

- Handling large volumes of data;
- Adapting to dynamic changes in processes;
- Supporting decision-making processes without extensive expertise.

These limitations restrict their effectiveness in identifying real improvement opportunities and reduce their adaptability to various industrial contexts, resulting in the inability to uncover critical areas for process improvement and optimization. Our methodology aims to address these challenges by integrating advanced data science techniques, which allow for deeper analyses of large data volumes, more flexible adaptation to changes, and a more intuitive user interface. This approach not only provides more accurate and comprehensive analyses but also simplifies the process of interpreting results, thereby increasing the added value for corporate decision-making processes.

In recent years, the role of process mining in analyzing and optimizing manufacturing processes has been pivotal. Despite its significance, there are still significant research gaps in the existing methodologies. Specifically, the current methods struggle to effectively tackle the dynamic nature of process changes and the complexity of data structures, thereby limiting their capacity to offer in-depth and actionable insights. This study aimed to overcome these challenges by proposing a new process mining technique that combines advanced data analysis and machine learning to offer a more thorough and flexible analysis of processes. The primary objectives of our technique include the following:

- Enhancing the ability to detect and react to process changes in real-time.
- Providing deeper insights into the causes and effects of process variations.
- Facilitating more precise and efficient process optimization.

For a more efficient analysis, therefore, it is recommended to use data science methods such as process mining. Process mining, as a new data analysis method, combines information sciences with analytical statistical technologies. This approach combines statistics, information sciences, and mathematical computations to determine the relationship between key factors such as utilization, timing parameters, and others. Moreover, it offers the possibility of a quick identification of the main root causes of waste throughout the process [3].

The current problem described in several scientific works is that various forms of waste can arise in industrial processes, reducing efficiency and productivity [4–6]. Waste can occur in various forms, including excessive use of resources, time, or energy, unnecessary steps or activities, and inefficient use of labor [7–9]. Among the main causes of waste are the following:

1. Insufficient planning and management: If processes are not properly planned and managed, it can lead to excessive use of resources and unnecessary tasks;
2. Inefficient use of resources: This can include using more materials or energy than necessary or employing labor that is not required for specific tasks;
3. Obsolete or inefficient technologies: Old or inefficient technologies can increase costs and reduce process efficiency;
4. Errors or failures: Errors in processes can lead to waste by requiring repairs or re-execution of tasks;
5. Lack of quality control: If the quality of processes and outputs is not properly monitored and controlled, it can lead to waste through poor performance or product defects.

Process mining can help identify and address these forms of waste in industrial processes [10]. Using the data analysis of processes, process mining can reveal areas where waste occurs, identify causes, and suggest solutions [11]. For example:

1. Identification of unnecessary or inefficient steps: Process mining can visualize the entire process and identify steps that do not add value;

2. Performance and quality analysis: Process mining can analyze the performance of individual process steps and the quality of outputs, which helps to identify areas for improvement;
3. Prediction and prevention of problems: Using advanced techniques such as machine learning, process mining can predict potential problems and enable their resolution before they occur;
4. Resource optimization: Process mining can help to optimize the use of resources by revealing where and when resources are needed and where they can be used more efficiently.

In this way, process mining can contribute to the improvement of efficiency and the reduction of waste in industrial processes. Using process mining technology, it is possible to uncover hidden waste, deficiencies, and opportunities for the optimization of manufacturing processes through a systematic and objective view of the data. This method enables the extraction, analysis, and visualization of digital traces that are created in information systems, providing valuable information to support decision-making processes.

#### Basic Definitions

With the increasing availability and lower costs of technologies, most manufacturing organizations can manage their processes using a certain information system (Enterprise Resource Planning, ERP, or Manufacturing Execution System, MES) [12–14]. Furthermore, in the era of Industry 4.0 and big data, modern manufacturing systems generate a large amount of data, which has the potential to be used as an information source [15]. Various data analysis techniques have been used to analyze processes specific to the manufacturing industry, such as production management and maintenance processes [16,17]. Business intelligence, knowledge discovery, and data mining are some of the tools that meet the need for automated data analysis. Process mining is a new research area that allows for the automatic discovery of business processes and a host of other process improvement techniques, such as performance analysis. It is defined by the IEEE Task Force on Process Mining as follows [18]: “The idea of process mining is the discovery, monitoring, and improvement of real processes (i.e., not assumed processes) by extracting knowledge from event logs that are readily available in today’s information systems”.

The objective is the automatic construction of a process model (abstraction) of an actual (unknown) process based on recorded events, which typically only cover a portion of reality or potential behavior, as shown in the following figure (Figure 1) [19].

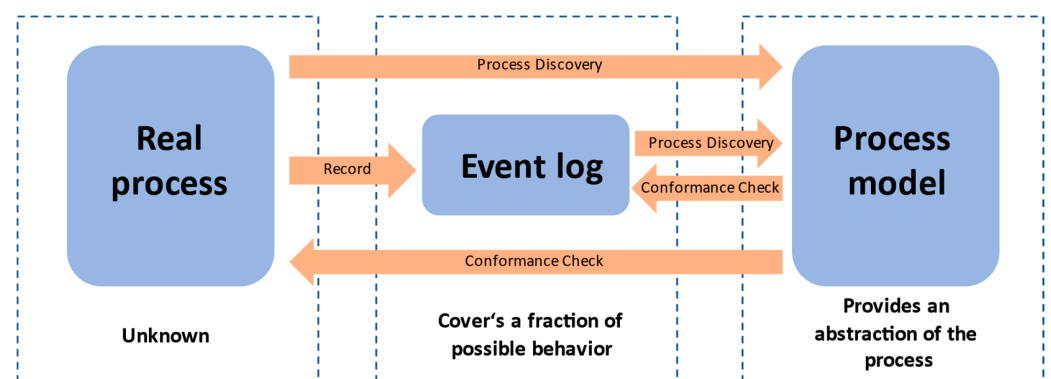


Figure 1. Process Mining Definition.

One definition of automated processes is those that are supported by applications that, after receiving an event log (input) that depicts the actual process, generate a representation of the process model using a variety of algorithms.

On the other hand, the process model refers to a visual and structured representation of processes, such as manufacturing processes, like entering production plans and managing forecast plans, and other processes, such as business transactions, administrative events,

etc. This structured depiction of a process is actually an abstraction and simplification of the real process, the behavior of its participants, the calculation of dynamic units, and so on [20].

The starting point for process mining is, therefore, the event log, i.e., a record of sequential transactions, events, or behaviors that have occurred in the process. These event logs can be structured and stored in a system, such as ERP or Workflow, or they can be dispersed and informally maintained in files or various machines that record events that partially explain the process.

This definition can be interpreted in two ways, as authors, developers, and creators of process mining also refer to the comparison of process models that are already structured for compliance or improvement purposes and perform performance analyses. But in reality, the function of creating process models is the main attribute of this research area, and all others are a natural consequence of the first; they are applications and outcomes of process discovery [21–23].

Given this, we consider that process mining techniques allow for the extraction of knowledge from event logs. For example, audit logs of a workflow management system, or “transaction logs”. An enterprise resource planning (ERP) system can be used to discover models describing processes, organizational roles, entire networks of recorded data, etc. In addition, process mining can be used to monitor process deviations, for example, by comparing observed events (reality) with predefined models or business rules.

Process mining can be considered a new area in business intelligence aimed at learning processes from recorded activities. Of course, this can only be carried out in the case of structured processes, i.e., behaviors that follow a consistent rule and sequence and can be assigned to the same case.

If it is the case of random, inexplicable data or anarchic process events, the process cannot be understood because there is no process, no cause and effect, and no relationship between an activity and its consequences.

The relationship between events must be rational, meaning there is a known or unknown logic behind it, which will be understood and structured through process modeling. Also, these events must be observable, meaning we should be able to record and monitor the execution of activities or behaviors.

As already explained, process mining is not limited to discovering control flow; the discovery function is just one of the three basic forms of process mining, with organizational, data, case, and temporal aspects also playing an important role in this discipline. This topic is elaborated on in the following chapter on process mining applications. In the next section, we present the background of process mining, but for now, we should understand that process mining is not just a specific type of data mining but can be considered the “missing link” between data-oriented data mining and traditional model-oriented areas of the study of business process management (BPM), as most data mining techniques are not process-oriented and BPM does not create process models from data and recorded events [24].

The following diagram (Figure 2) [25] illustrates the possible relationship between three different subjects: on the left is BPM based on a model, on the right is DM based on data, and at the center is process mining based on data and processes.

It is important to mention that process mining is not limited to offline analysis; although information gathering is performed from historical event data, real-time records can be used, and the results can and should be applied to ongoing cases. For example, the completion time of a partially processed customer order can be predicted using a detected and running process model. In the rest of this article, we explore and go into detail of process mining, where and how it is applied, its types and perspectives, its goals and problems, and existing tools, owing to which this approach in a business or commercial organization is real and tangible.

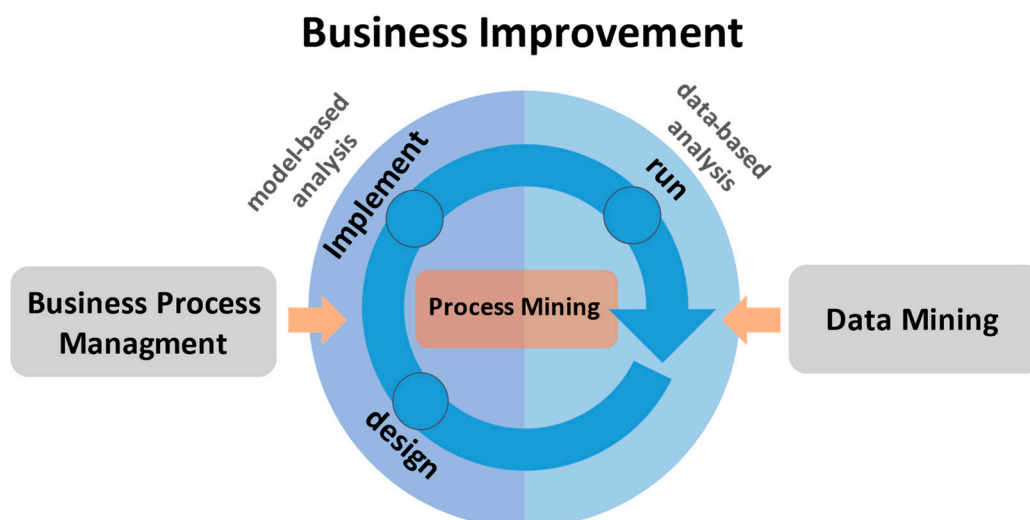


Figure 2. PM link between BPM and DM.

The basic requirements for a process mining tool should be aligned with the goals of implementation. This means that the solution should help the company [26] achieve the following:

1. Gain an overview of how the business process is defined, understand how it works, and know what its limitations and problems are;
2. Support discussion and brainstorming about the business process. It should create a common starting situation that is acceptable and concrete to start a discussion;
3. Document and educate, i.e., increase transparency and knowledge sharing; inform all relevant sources about their tasks, responsibilities, and the impact of their activities;
4. Audit and verify the process; review and search for errors and mistakes in design, systems, or even individual activities;
5. Analyze the performance of processes and enforce improvements. By understanding the process and also being able to determine the key weak points, the solution should allow for identifying key factors of change that will improve the level of services;
6. Animate and replay models, which allow users to explore the process and bring new insights and feedback about the process;
7. Design the process for system or process specification, meaning the model can serve as a binding contract or requirements list for system setup on the one hand and as a starting point for process change on the other.

## 2. Materials and Methods

This article not only highlights key trends in the field but also emphasizes the significance of understanding the algorithmic method of processing and the foundational mathematical concepts underpinning the described method. Providing valuable insights into process flows and identifying areas for improvement, the results contribute to the growing body of knowledge in process mining. This research underscores the broader applicability of process mining and its role in fostering a more efficient and competitive manufacturing environment.

### 2.1. Algorithms of Process Mining

The  $\alpha$ -miner, or  $\alpha$ -algorithm, is a foundational algorithm used in process mining aimed at reconstructing causality from a set of event sequences. It was first introduced by van der Aalst, Weijters, and Maruster [18]. The goal of the  $\alpha$ -miner is to transform an event log into a workflow net based on the relationships between various activities in the event log. An event log is a multiset of traces, and a trace is a sequence of activity names. Since its

inception, several extensions or modifications of the  $\alpha$ -miner have been introduced, which are discussed further here.

The  $\alpha$ -miner was the very first proposed algorithm for process discovery and provides a good overview of the objective of process discovery and how various activities within a process are carried out. The  $\alpha$ -miner was also the basis for the development of many other process mining techniques, such as the heuristic miner, and genetic mining was developed based on the idea upon which the  $\alpha$ -miner is built.

An event log is the basic requirement for using any algorithm to discover process characteristics. An event log consists of a unique case identifier, the name of the activity describing the action occurring in the process, and a timestamp. An event log can be represented as a multiset of activities. In the following example, for simplification, a letter of the alphabet is used to represent an activity. Consider the example of an event log presented in the following table (Table 1) [27]. The numbers “1, 2, 3” represent specific IDs of a case, for example, a business or manufacturing order. The letters “A, B, C, D, E” also represent a sequence of specific activities that occur within the ID, such as receiving an order, processing, shipping, etc. For example, under ID “1”, there were three different activities “A, B, C” that occurred at different times and in a specific sequence.

**Table 1.** Sample event log.

Case ID	Activity	Timestamp
1	A	5 December 2022 10:12:18.000
1	B	5 December 2022 11:30:18.000
1	C	5 December 2022 13:22:18.000
2	A	9 December 2022 12:11:18.000
2	C	9 December 2022 14:40:18.000
2	A	9 December 2022 16:16:18.000
2	D	9 December 2022 17:05:18.000
3	A	11 December 2022 08:02:18.000
3	E	11 December 2022 09:43:18.000
3	D	11 December 2022 11:37:18.000

An event log is a multiset of traces, and a trace is a sequence of activities. Such an event log can be represented by the following notation:

$$L1 = [\langle A,B,C,D \rangle, \langle A,C,B,D \rangle, \langle A,E,D \rangle] \quad (1)$$

Each event log can be decomposed into a multiset of traces, and such traces can be further used to parse out the relations between various activities in the process. According to the rules of the  $\alpha$ -miner, activities belonging to different cases can have 4 types of relations between them (Table 1) [27].

The  $\alpha$ -miner begins by converting the event log into direct succession, causality, parallel, and choice relationships and uses them to create a Petri net describing the process model (Figure 3). Initially, the algorithm constructs a footprint matrix. Using the footprint matrix and the pattern displayed above, a process model can be constructed. Based on the four relationships described previously, a matrix based on traces is first created. Using the matrix (Table 2) [28], based on traces, places are discovered. Two sets of duties are assigned to each place in order to keep the number of places minimal. (See Table 3).

- $Y_W$  is the set of all pairs (A, B) of maximal sets of tasks such that neither  $A \times A$  and  $B \times B$  contain any members of  $\succ$  and  $A \times B$  is a subset of  $\rightarrow$ ;
- $P_W$  contains one place  $p(A,B)$  for every member of  $Y_W$ , plus the input place  $i_W$  and the output place  $o_W$ .

The flow relation  $F_W$  is the union of the following:

$$\left\{ \left( a, p_{(A,B)} \right) \mid (A,B) \in Y_W \wedge a \in A \right\} \quad (2)$$



$$\left\{ \left( p_{(A,B)}, b \right) \mid (A,B) \in Y_W \wedge b \in A \right\} \tag{3}$$

$$\{ (i_w, t) \mid t \in T_1 \} \tag{4}$$

$$\{ (t, i_o) \mid t \in T_o \} \tag{5}$$

The result is as follows:

- A Petri net structure  $\alpha(W) = (P_W, T_W, F_W)$ ;
- One input place  $i_W$  and one output place  $o_W$ ;
- Every transition of  $T_W$  is on a  $F_W$ -path from  $i_W$  to  $o_W$ ; it is indeed a workflow net.

For the example given above, the following Petri net would be the resultant of the application of Alpha miner (Figure 4) [29]:

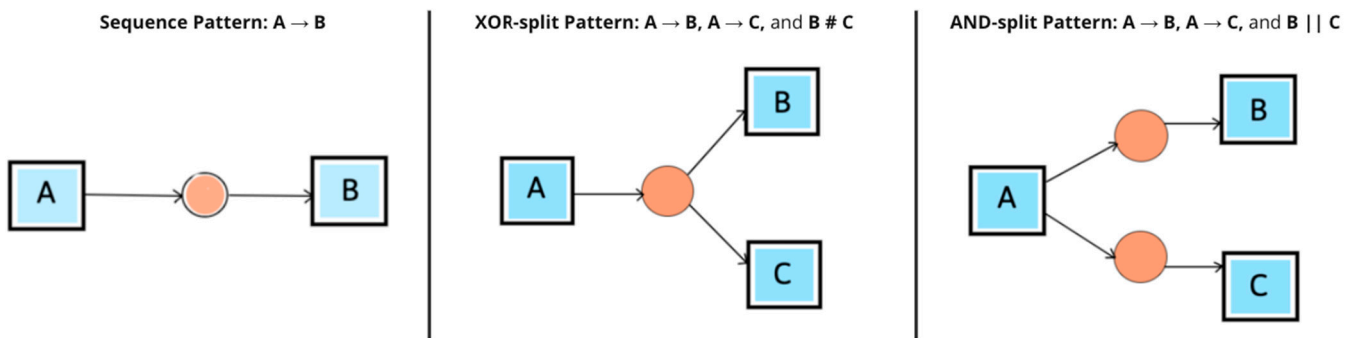


Figure 3. Patterns of alpha-algorithm.

Table 2. Description of symbols and their functions.

Name	Symbol	Function	Example
Direct Succession	>	$x > y$ if and only if some relation $x$ is directly followed by $y$ .	$A > E, A > C$
Causality	$\rightarrow$	$x \rightarrow y$ if $x > y$ and not $y > x$ .	$A \rightarrow E$
Parallel	$  $	$x    y$ if $x > y$ and $y > x$ .	$B    C$
Choice	#	$x \# y$ if not $(x > y)$ and not $(y > x)$ .	$A \# D$

Table 3. Footprint matrix for log L1.

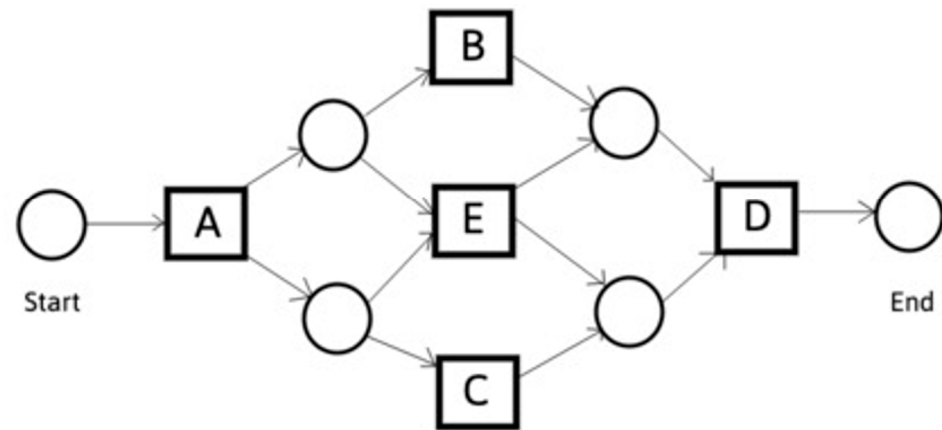
	A	B	C	D	E
A	#	$\rightarrow$	$\rightarrow$	#	$\rightarrow$
B	$\leftarrow$	#	$  $	$\rightarrow$	#
C	$\leftarrow$	$  $	#	$\rightarrow$	#
D	#	$\leftarrow$	$\leftarrow$	#	$\leftarrow$
E	$\leftarrow$	#	#	$\rightarrow$	#

Limitations [29]:

- Implicit places: The  $\alpha$ -miner cannot distinguish between implicit and required places and, therefore, may result in additional unnecessary places in the discovered Petri net.
- Loops: The  $\alpha$ -miner is unable to discover loops of lengths 1 and 2 in the process model.
- Representational bias: The  $\alpha$ -miner can only discover a Petri net, which introduces representational bias, such as the requirement for unique visible labels for each transition.

Currently, in addition to the  $\alpha$ -algorithm, there are several different types of process discovery algorithms. Each algorithm is used to extract various process models using an event log. Abstraction-based algorithms generate models by arranging event relationships

in the event log, while on the other hand, the heuristic miner generates models where events are arranged based on the frequency of occurrences. Events that occur less than a certain threshold number are ignored, so heuristic miners work better with event logs containing noisy data. Search-based algorithms (like the Genetic Algorithm Miner (GA)) attempt to mimic the process of evolution. The key factors of the process discovery technique are the balance between fitness, precision, generalization, and simplicity [30].



**Figure 4.** Petri-like result of the Alpha miner application.

The most commonly used algorithm for process discovery is the  $\alpha$ -algorithm, which is an abstraction-based algorithm. The  $\alpha$ -algorithm creates a dependency graph based on the order of events in the event log. It does not take into account the frequency of observations, so the noise present in the log causes a high discrepancy from the expected behavior. To avoid this, one can use the fuzzy miner, which is a heuristic approach that generates a model according to the frequency of traces. The fuzzy model is difficult to translate into other process modelling languages, but its representation is simple for understanding the behavior of the system [10].

## 2.2. Methodology for Analyzing Manufacturing Processes Using Process Mining

In this chapter, a proprietary methodological approach to the analysis of manufacturing processes using process mining tools is described. The proposed procedure is based on a schematic flowchart (Figure 5), which contains specific steps and decisions. These steps help to properly process the input data, systematically analyze the production processes, and identify areas for improvement.

Defining KPIs:

For proper analysis, it is necessary to define key performance indicators (KPIs) to be monitored. These indicators should align with the organization's objectives, for example: work-in-progress, lead times (orders, production, delivery), order lifecycle, among others (rework, capacities).

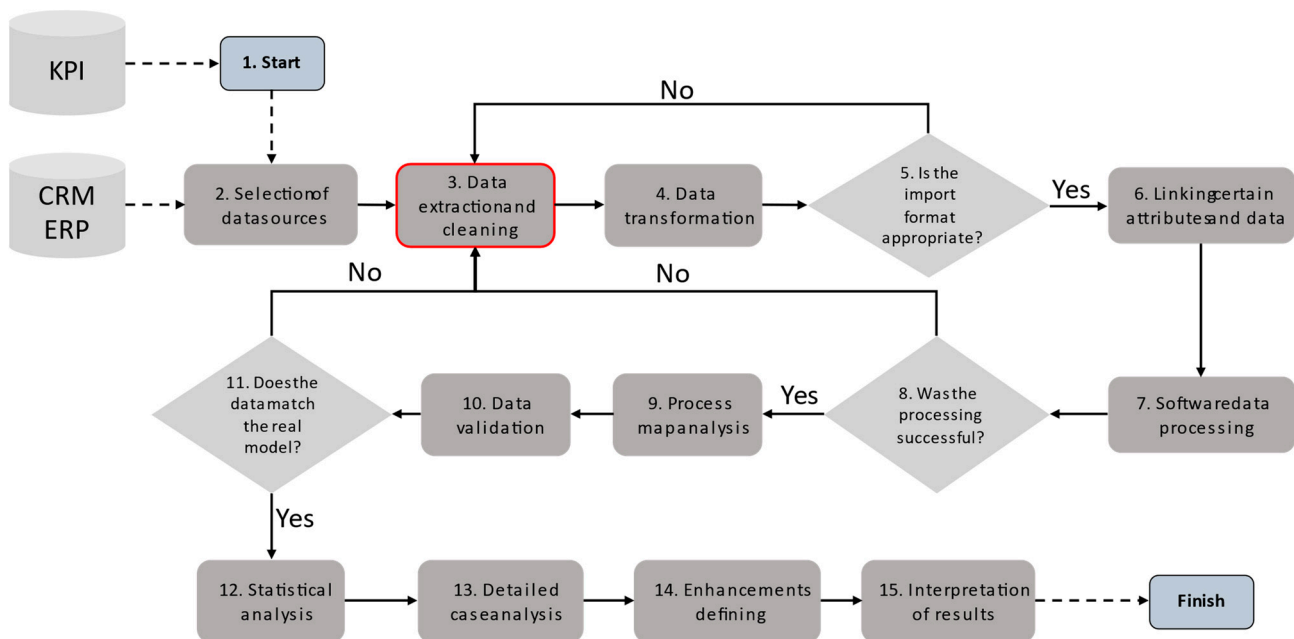
### 1. Beginning:

At the start of the analysis process, the goal and questions to focus on are defined. These questions should aim to understand manufacturing processes, uncover deficiencies, and seek opportunities for improvement. Goals and questions are directly related to key performance indicators:

- A map of the sequence of the overall manufacturing process and individual orders;
- Time utilization of all or selected critical equipment;
- Rework analysis (quality indicator);
- Lead times;
- Number of work-in-progress manufacturing orders;
- Effective times;



- Order lifecycle.



**Figure 5.** Procedure for the analysis of manufacturing processes using process mining tools.

## 2. Selection of data sources:

Data sources to be used in the analysis of manufacturing processes are identified. These sources can include logs from information systems, system records, sensor data, and data from identifiers.

- **Process logs:** These data are generated automatically during the execution of manufacturing processes. They may contain records of individual steps, times, actions performed by the operator, order statuses, and other relevant information. Process logs are often stored in the form of log files or records in information systems.
- **System Records:** These data come from various information systems, such as Manufacturing Execution Systems (MESs), Enterprise Resource Planning (ERP) systems, or production-tracking systems. These records contain information about orders, times, quantities, operators, and other relevant data [31].
- **Sensor data:** In some cases, manufacturing processes are monitored using sensors that collect data on temperature, pressure, speed, vibrations, and other physical parameters. These sensor data can be integrated into process mining analysis to gain additional insights into processes.
- **Identifier data:** In some cases, identifiers such as barcodes or RFID tags are used to track the movement of materials, products, or workers within manufacturing processes. These data can provide information on material flow, timing, and resource movement in the process.

Some suitable formats for data import:

- **Event Log:** CSV (Comma-Separated Values), XES (eXtensible Event Stream);
- **Network Log (Petri net logs):** PNML (Petri Net Markup Language), BPMN (Business Process Model and Notation);
- **Relational Databases:** SQL (Structured Query Language)—for the process of extracting from the database and importing data into the tool;
- **Digital Document Format Files:** XLS, XLSX (XML Spreadsheet), CSV (Comma-Separated Values).

### 3. Data extraction and cleaning:

Necessary data are extracted from the sources, and a data cleaning process is performed, which includes the removal of errors, duplicates, and inconsistencies. This ensures that the data are reliably prepared for analysis. The following are the basic steps for data extraction and cleaning:

- **Identification of relevant data:** In the first step, it is necessary to identify data sources and determine which of them are relevant for the analysis of manufacturing processes. This includes identifying relevant logs, database tables, or other sources that contain necessary information about processes.
- **Data extraction:** After identifying relevant data, the extraction of these data from the source systems is performed. Extraction can be performed using automated scripts or data extraction tools that allow for the necessary information to be obtained from structured or unstructured sources.
- **Data cleaning:** Various issues, such as missing or invalid values, inconsistent formats, or duplicates, can occur during the extraction process. Data cleaning involves removing these impurities and ensuring that the data are accurate, complete, and consistent. Various techniques are used, such as removing duplicates, supplementing missing values, correcting erroneous values, or normalizing data.
- **Normalization of data structures:** When analyzing different source data, data may be stored in various formats or structures. For a successful analysis and comparison, it is important to normalize these data structures into a uniform format. This step involves aligning attributes, creating a uniform table structure, or grouping data from different sources into a single file.

### 4. Data transformation:

The data are transformed into an appropriate format and structure for further analysis. This may involve normalization, aggregation, and merging data from various sources to create a coherent picture of manufacturing processes. The following are the basic steps for data transformation:

- **Attribute Transformation:** This step involves transforming individual attributes (columns) of data. This can include converting data into a different format, such as changing time to a different time zone, converting text to uppercase, or creating new attributes by combining existing ones.
- **Aggregation and Grouping:** In some cases, it may be necessary to aggregate and group data to obtain a summary or high-level information. This step allows for the reduction of data complexity and provides an overview of the overall course of processes.
- **Time Normalization:** If data come from various sources or systems, it may be necessary to normalize time data to make them comparable. This includes aligning timestamps, creating a uniform timeline, and adjusting time intervals.
- **Identifier Normalization:** If data contain various forms of identifiers, such as different names for the same process activities, they need to be normalized. This ensures a uniform identifier for the same objects and facilitates later analysis and process mapping.

### 5. Decision block: "Appropriate format for import":

At this point, a decision is made as to whether the transformed data are in the appropriate format for import into the tool. If the format is correct, proceed to the next step. If not, return to step 3 and repeat the data extraction and cleaning process.

### 6. Linking specific attributes and data:

In this phase, certain attributes are linked with data in the input data. This creates connections between various aspects of manufacturing processes and provides more comprehensive and detailed information. The resulting analysis directly depends on the assigned attributes.

When linking attributes with data, at least three main elements must be identified. It is possible to gain a process view of the data by defining at least these three elements [32]:

- Case ID: This determines the scope of the process. The process ID is a specific case of the process. The significance of a case in a particular situation depends overall on the process and input data.
- Activity: This determines the steps on the process maps. For example, a process in production may consist of steps such as “sanding”, “cutting”, “pressing”, “levelling”, “bending”, “storing”, “dispatching”, or naming specific machines (performed by various people, e.g., workers and masters). Some of these steps may occur more than once in a case, and not all may always be performed.
- Timestamp: If a sequential log file is not available, timestamps determine the sequence of activities in the process. The third important prerequisite for process mining is to have at least one column with a timestamp indicating when the individual activities took place. This is important not only for analyzing the timing behavior of the process but also for establishing the sequence of activities in the event log.

#### 7. Data processing using software:

Data processing using software represents mathematical and algorithmic computations. It involves transforming input data into a process map and statistical summarization, and dividing by main ID attributes into separate processes for a complete and structured overview and analysis of processed data.

Within the case study, the Process Mining tool—Fluxicon Disco is used for data processing and analysis. This software enables the extraction, visualization, and analysis of process data and the identification of patterns, trends, and anomalies in processes.

If there are unreadable data or technical errors, the software alerts the user and indicates the specific problem. Such an approach facilitates the identification of import errors and provides the possibility to adjust the data for correct processing. Inaccuracies can be identified using the following indicators:

- Data inconsistency error (missing, incorrect format, unexpected data volume, incorrect attribute identification, incorrect timestamps, unreadable text format, time axis gaps).
- The process map does not correspond to the actual state of the processes. Visually, it is possible to identify the sequence of steps in the process map and check lead times.
- Statistical evaluations do not match actual values.
- Case analysis incorrectly identifies and divides individual cases based on the assigned ID attribute.

#### 8. Decision block: “Was processing successful?”

After data processing using the software, it is necessary to verify whether the process was successful. If so, proceed to the next step. If not, return to step 3 and repeat the data extraction, cleaning, and transformation process.

#### 9. Process map analysis:

The analysis of the created process map is the first outcome of the data analysis. The flow of processes is examined, critical areas are identified, and potential problems and opportunities for improvement in manufacturing processes are uncovered.

The software provides various options for analyzing the process map, which allow for a detailed examination of the data and the revelation of important information about manufacturing processes. The main steps in analyzing the process map include the following:

- Time Filtering: The software allows the user to define the time span in which to analyze events. This allows for the analysis of a specific period, such as a month, week, or day, and it also allows for the observation of changes and trends in the processes within that time frame.
- Source Filtering: The user can select specific roles (individuals, departments, or machineries) and limit the analysis to their activities in the process. This can help explore

how individual resources contribute to the process flow and identify any issues associated with their participation.

- Step Filtering: The user has the option to select specific steps in the process and narrow down the analysis to these steps. This can examine a specific part of the process and obtain detailed information on how it is performed and how it differs from the expected model.

#### 10. Data validation:

The correctness and reliability of the data used in the analysis are validated. Data validation represents a comprehensive process of verifying the accuracy and reliability of the obtained results of the analysis of manufacturing processes. For this purpose, the following specific criteria and validation methods were introduced:

- Comparison with historical data: This represents the validation of the accuracy of the simulation results by comparing them with available historical data from a real operational environment. This comparison allows for the identification of any discrepancies and ensures that the model adequately reflects actual processes.
- Quantitative analysis of deviations: A quantitative analysis of deviations is performed to numerically express the differences between the results and the actual data. This method helps us determine the accuracy and reliability rates of the model.
- Statistical verification: We use statistical tests, such as the *t*-test or ANOVA, to verify whether the differences between the results and historical data are statistically significant. This allows us to assess whether the model's results can be considered trustworthy.
- Expert evaluation: The model's results are provided for evaluation by a group of experts on the subject, who provide a qualitative assessment of the realism and relevance of the results.
- Iterative optimization: Based on the results of the validation, iterative modifications of the model are performed until the set criteria for accuracy and reliability are met.

This detailed and multi-level approach to validation ensures that the conclusions and recommendations are based on verified and reliable data, thereby increasing the value and applicability of the analysis outputs.

#### 11. Decision block: "Do the data match the real model?"

The obtained data are compared with the real model of manufacturing processes. If they match, proceed to the next step. If not, return to step 3 and repeat the data extraction, cleaning, transformation, and processing.

#### 12. Statistical analysis:

A statistical analysis is conducted on the obtained data to uncover patterns, trends, and dependencies in the processes. Various methods and techniques, such as time series, regression analysis, variance evaluation, and others, are used.

In the statistics section, the following information areas can be evaluated [33]:

- a. Overview information: key overview data, global statistics, the number of events and cases in the dataset, the number of different activities, the median and average case duration, and the time frame included in the log.
- b. Performance charts: Several pre-generated charts visualizing performance metrics are displayed. Charts can be exported.
- c. Detailed information: At the bottom of the screen, detailed statistical information is displayed in tabular form. Tables can be exported as CSV files for further processing in other tools.
- d. Global statistics: Overview statistics about the entire dataset, individual cases, and variants.
- e. Activity statistics: Statistics on individual process steps in the dataset are displayed.
- f. Resource statistics: Statistics about people or organizational units in the dataset are displayed.

### 13. Detailed case analysis:

Focus on a detailed analysis of specific cases or events that may affect manufacturing processes. Deviations and causes of problems are examined, and ways to address them are sought.

Additionally, looking at individual cases with their history and all their attributes can provide additional context (such as a comment field), which sometimes explains why something happened. For example, if deviations from the described process or violations of an important manufacturing rule are identified, it is possible to obtain a list of these cases and discuss them with the people involved or provide them with additional training.

The case display consists of the following areas:

- Complete log: Displays a list of all cases in the data set.
- Individual variants: It is possible to select an individual variant, which is a specific sequence of activities. The case list will then show only the cases that follow the same sequence of activities. Sorting is carried out by the frequency of variants.
- List of cases: A list of case identifiers for either the complete log or the selected variant.
- Individual cases: It is possible to view further details about the currently selected case.
- Searching: For specific case IDs or attribute values.
- Filtering: Filtering controls for logs of the current data set are accessible for each examined item. Filters are, indeed, important for examining specific aspects of the process and, of course, for focusing the analysis.
- Copying, removing, and exporting data files: Data files can be copied, removed, and exported directly from the current analysis view.

### 14. Defining Improvements:

Based on the analysis, opportunities for improving manufacturing processes are identified. Specific steps, measures, and improvements are proposed that could lead to a more efficient and higher-quality manufacturing process.

Breakdown of the main steps in defining improvements for manufacturing processes:

- Identification of improvement opportunities: Based on the analysis of available data and the results of previous steps, areas of manufacturing processes where improvements can be achieved are identified. This may include identifying activities that are inefficient or lead to errors and problems.
- Formulation of specific steps and measures: Based on identified areas for improvement, specific steps, measures, and improvements that could be implemented are proposed. These measures should be clearly defined and should focus on improving the efficiency, quality, and performance of manufacturing processes.
- Use of key improvement indicators: Key performance indicators (KPIs) are the main indicators for improvement. These indicators may include factors such as time costs, quality, productivity, resource utilization, time cycles, and other relevant metrics. Based on these indicators, priority areas for improvement are identified, and goals and measurable parameters for improvement are determined.
- Identification of secondary problematic situations and discrepancies: In addition to the main areas for improvement, secondary problematic situations and discrepancies may be identified. These can provide additional opportunities for improving manufacturing processes and can be included in the overall improvement plan.

### 15. Interpretation of results:

The obtained results of the analysis and the expected outcomes of improvements are interpreted in the context of the set objectives:

- Analysis of obtained results: Based on the processing and evaluation of data and information from previous steps, an analysis of the obtained results is conducted. Patterns, trends, relationships, and dependencies identified during the analysis process are analyzed.

- Context of set objectives: The analysis results are interpreted in the context of the set objectives and requirements of manufacturing processes. It is examined whether the set goals were achieved and how the results relate to these goals.
- Provision of conclusions: Based on the interpretation of results, conclusions are drawn, and main findings and discoveries are formulated. Relevant information and facts from the collected analytical results bolster these conclusions.
- Recommendations and guidance for future decision-making: Based on the conclusions and knowledge, recommendations and guidelines for future decisions regarding manufacturing processes are developed. These recommendations focus on improving the efficiency, productivity, quality, and performance of processes.
- Actions concerning manufacturing processes: Based on the interpretation of results and recommendations, specific actions and measures aimed at implementing the proposed improvements in manufacturing processes are planned and implemented. These actions may include changes in procedural processes, technology enhancements, employee training, and other relevant initiatives.

### 3. Results

The proprietary methodological approach described in Section 2 was applied and verified within a case study using real data from a manufacturing enterprise.

The enterprise that this case study focuses on is one of the European leaders in the segment of large, welded steel structures. Orders are processed both in series and custom-made, and all important dynamic and static data are recorded in a specific ERP system. These data were collected from this system and served as the basis for our analytical work. Specifically, they involve production-related data, which includes information about machines, equipment, handling, preparation, machining, and more.

The aim of this case study was to implement and verify the proposed methodological procedure for implementing process mining on the actual production data of this enterprise and analyze them based on the defined key performance indicators. The software Fluxicon Disco was used for this purpose, with this specialized application serving to analyze processes using the process mining method. In this way, an attempt was made to gain a deep insight into the manufacturing processes of this enterprise and to identify potential areas for improvement and optimization.

In the figure (Figure 6), the extracted raw data from the ERP system are displayed. These data represent records of the work operations of the organization's production section. They contain detailed information over a certain period about each activity within the production hall:

- Timestamp;
- Article number;
- Worker ID;
- Machine ID;
- Work order;
- Operation numbers of the technological process;
- Identifiers of batches and transactions;
- Number of completed pieces;
- Operation names;
- Planned deadlines for each activity;
- Additional information and other similar details.

The raw data were subsequently transformed into a clear, unified format, and essential attributes were identified for analysis (Figure 7):

- Timestamp (time attribute);
- Article number (main case attribute);
- Production order (secondary case attribute);
- Worker ID (human resource attribute);
- Machine ID (production resource attribute);



- Numbers of technological procedure operations (attribute of the sequence of technological operations).

Start Date Time	End Date Time	Worker	Article number	Machine	Transac Time	Worker Order	ID	Operation	Production crew
2022/10/02 08:10:00.000	2022/10/02 08:10:00.000	62762	B433453	BC 2L	2110	VLV	3278104	LABOR	4066770
2022/10/02 17:36:00.000	2022/10/02 17:36:00.000	62762	A143503	BC 2L	2110	VLV	999301	LABOR	1218326
2022/10/02 17:50:00.000	2022/10/02 17:50:00.000	62762	B433453	BC 2L	2110	VLV	999201	LABOR	1218326
2022/10/02 18:05:00.000	2022/10/02 18:05:00.000	62762	A143503	BC 2L	2110	VLV	999201	LABOR	1218326
2022/10/02 18:20:00.000	2022/10/02 18:20:00.000	62762	B433453	BC 2L	2110	VLV	999201	LABOR	1218326
2022/10/02 21:29:00.000	2022/10/02 21:29:00.000	62762	A143503	BC 2L	2110	VLV	2613104	LABOR	4066770
2022/10/03 03:44:00.000	2022/10/03 03:44:00.000	62762	B433453	BC 2L	2110	VLV	999201	LABOR	1218326
2022/10/03 15:38:00.000	2022/10/03 15:38:00.000	62762	C402200	BC 2L	2110	VLV	3278105	LABOR	7155104
2022/10/03 15:52:00.000	2022/10/03 15:52:00.000	62762	C402200	BC 2L	2110	VLV	999201	LABOR	1218326
2022/10/03 19:30:00.000	2022/10/03 19:30:00.000	62762	C402200	BC 2L	2110	VLV	999301	LABOR	1218326
2022/10/03 23:31:00.000	2022/10/03 23:31:00.000	62762	A143503	BC 2L	2110	VLV	2613104	LABOR	4066770
2022/10/03 01:11:00.000	2022/10/03 01:11:00.000	113805	B433453	R Cast	2120	VLVB	942101	LABOR	1218326
2022/10/03 05:32:00.000	2022/10/03 05:32:00.000	63304	A143503	BC 2R	2110	VLV	941201	LABOR	1218325
2022/10/03 07:41:00.000	2022/10/03 07:41:00.000	63304	A143503	BC 2R	2110	VLV	384301	LABOR	1218325
2022/10/04 07:02:00.000	2022/10/04 07:02:00.000	63304	D376200	PL 28	2305	AC-W	999201	LABOR	4144272
2022/10/04 12:53:00.000	2022/10/04 12:53:00.000	63304	D376200	PL 28	2305	AC-W	999201	LABOR	4144272
2022/10/04 15:28:00.000	2022/10/04 15:28:00.000	63304	D376200	PL 28	2305	AC-W	999301	LABOR	4144272
2022/10/04 16:11:00.000	2022/10/04 16:11:00.000	63304	D376200	PL 28	2305	AC-W	2613104	LABOR	4144272
2022/10/04 17:51:00.000	2022/10/04 17:51:00.000	63304	D376200	PL 28	2305	AC-W	3278105	LABOR	4144272
2022/10/05 04:03:00.000	2022/10/05 04:03:00.000	63304	C402200	PL 28	2305	AC-W	942101	LABOR	7155104
2022/10/05 08:17:00.000	2022/10/05 08:17:00.000	63304	C402200	PL 28	2305	AC-W	1622501	LABOR	7155104
2022/10/07 15:02:00.000	2022/10/07 15:02:00.000	63304	D376200	R Cast	2320	VLVB	942101	LABOR	4144272

Figure 6. Extracted data from the company’s ERP system.

Start Date Time	End Date Time	Art. Number	Machine	Work order	ID	Operation
2022/10/02 14:31:00.000	2022/10/02 15:23:00.000	A143503	3278105	1218326	5996710	5
2022/10/02 17:36:00.000	2022/10/02 17:38:00.000	A143503	999301	1218326	5996710	10
2022/10/02 17:50:00.000	2022/10/02 18:03:00.000	A143503	999201	1218326	5996710	15
2022/10/02 18:05:00.000	2022/10/02 18:20:00.000	A143503	999201	1218326	5996710	20
2022/10/02 21:29:00.000	2022/10/02 21:52:00.000	A143503	2613104	1218326	5996710	25
2022/10/03 23:31:00.000	2022/10/03 09:22:00.000	A143503	942101	1218326	5996710	30
2022/10/03 01:11:00.000	2022/10/03 01:11:00.000	A143503	942201	1218325	5996709	5
2022/10/03 05:32:00.000	2022/10/03 13:46:00.000	A143503	941201	1218325	5996709	10
2022/10/03 07:41:00.000	2022/10/03 07:42:00.000	A143503	384301	1218325	5996709	15
2022/10/13 07:31:00.000	2022/10/13 07:43:00.000	A143503	3278105	1218327	5996711	5
2022/10/13 23:12:00.000	2022/10/13 23:38:00.000	A143503	999201	1218327	5996711	10
2022/10/15 05:39:00.000	2022/10/15 06:07:00.000	A143503	942101	1218327	5996711	15
2022/10/15 10:59:00.000	2022/10/15 12:45:00.000	A143503	1385401	1218327	5996711	20
2022/10/18 03:09:00.000	2022/10/18 03:20:00.000	A143503	3278105	7202342	6237996	5
2022/10/18 07:21:00.000	2022/10/18 17:48:00.000	A143503	999201	7202342	6237996	10

Figure 7. Result of transformed data (part of the data table).

The transformed data were imported into the Fluxicon Disco software environment. The process involves identifying and assigning specific attributes presented by the software to the data being analyzed. The method of algorithmic processing and the resulting analysis depend on the assigned attributes. The selected attributes for analysis were as follows:

- Timestamp (time attribute): for correct time sequencing and duration calculations;
- Article number (main case attribute): for the correct division of each production order;
- Work order (secondary case attribute): for the complex division of each order into individual work orders;
- Machine ID (production resource attribute): a visual representation of the process map for analyzing work efficiency and inefficiencies, logistical times, and production resources;
- Numbers of operations of the technological procedure (attribute of the sequence of technological operations)—for understanding the technological sequence of operations within each order.

The imported data are then processed using mathematical algorithms available in the software solution (Figure 8). The initial result is two types of process maps: a frequency map

and a performance map. The first represents an organized display of ongoing production processes, focusing on the frequency of orders. The second process map has the same structure but represents the lead times within a certain period of production—efficient and inefficient logistical times. One of the benefits of these maps is their clarity and visualization of the flow of the processes being examined.

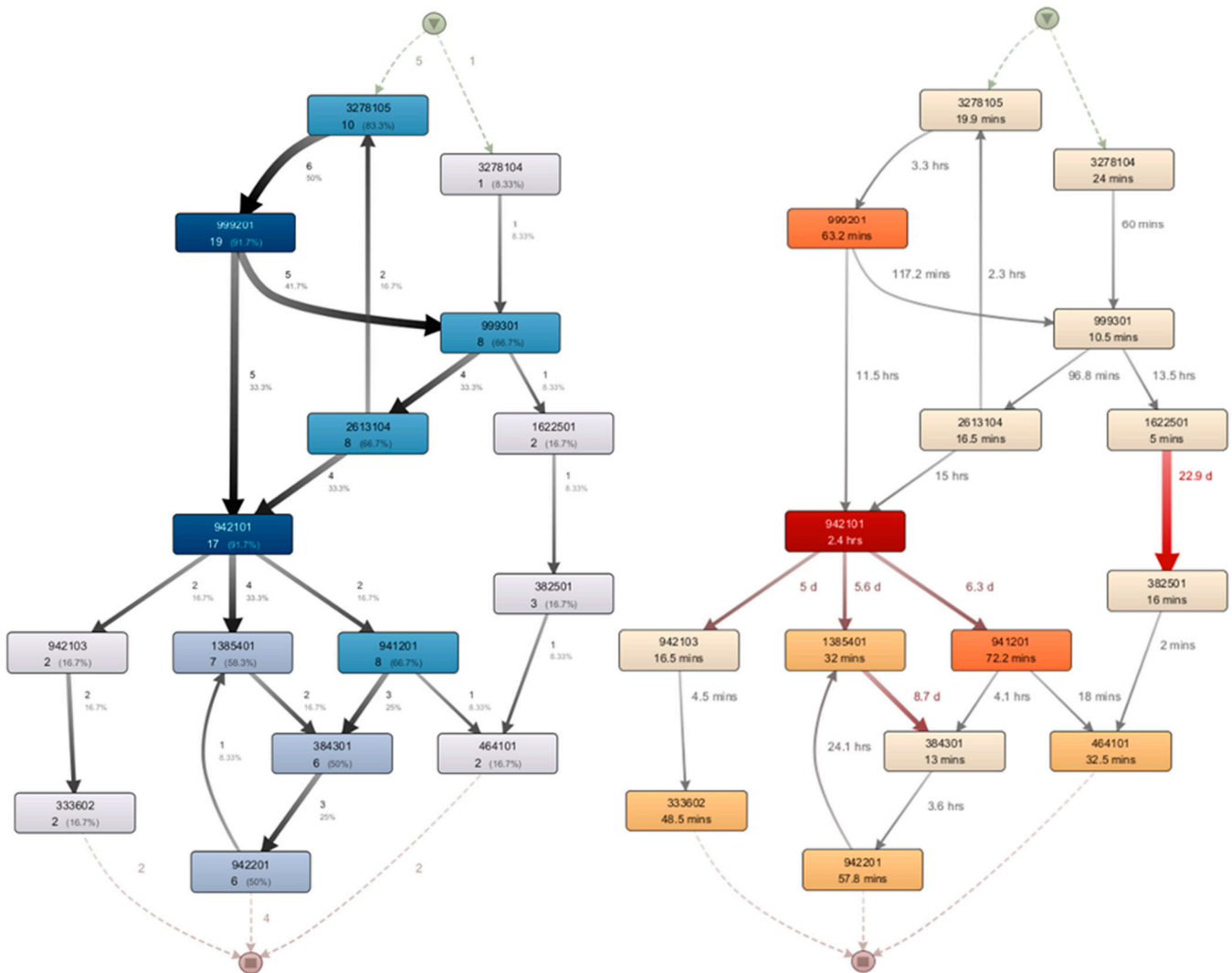
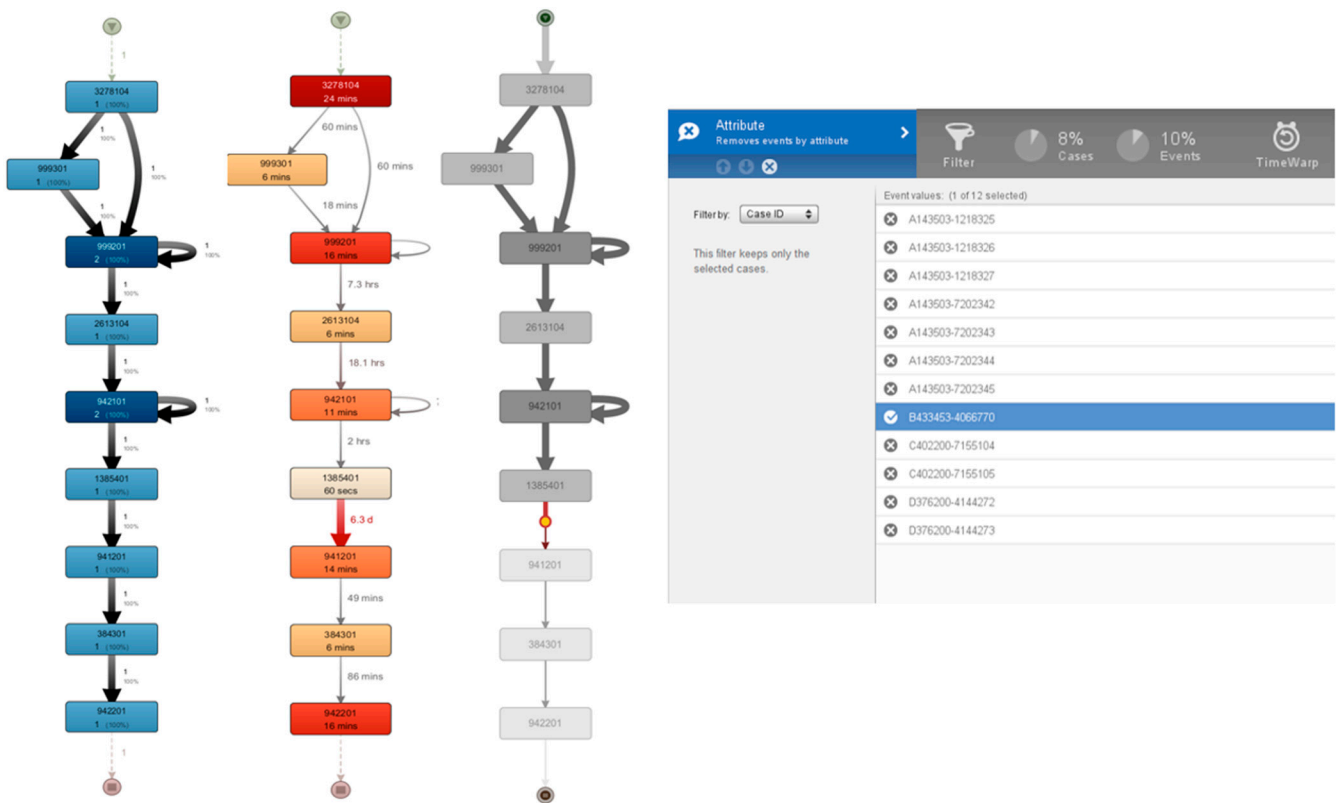


Figure 8. Processed data using mathematical algorithms in software—process map.

The processed process map can be further analyzed using various filters. Through the visually created process map with the characteristics of paths and activities, it is possible to define the time utilization of all or selected critical equipment and identify rework, lead times, and the frequency of value movement. Using filters allows for the visual display of individual orders or their groups, defining a time range (from–to, specific days), searching for information according to needed attributes (source, day, work order, etc.), and determining relationships between activities.

The illustration (Figure 9) shows an example of one of the filtering methods. Using the “Attribute” filter, a specific order with the number “B433453-4066770” is selected. After applying this filter, the shape of the process map changes based on the selected filter value. Based on the principles described above, the order map can be analyzed in terms of frequency, time performance, and animation of the process flow.



**Figure 9.** Example of filtering a specific order from the entire process.

Statistical analysis is also available, which, through mathematical computation methods, displays a criterion evaluation:

1. Events over time: The timeline on the horizontal axis of the log displays the total time frame recorded in the log. The metric of events over time shows the activity in the process and the number of activities carried out. The chart allows for viewing individual data points and uncovering seasonal and weekly patterns in the process, such as peaks on workdays and lower activity during weekends or increased activity during a certain season.
2. Activity statistics: This section contains performance metrics about activities in the process. Depending on the import configuration, an activity may consist of multiple columns. The cumulative count and Pareto diagram are analytical tools used for identifying and prioritizing key processes in the organization.

In this case, the activity statistics represent a list of machinery and equipment that process orders and, based on which, work orders are created. From these statistics, it is possible to define, for example, the following:

- The number of event occurrences at each machine facility.
- The level of process stability.
- The percentage share of a specific machine facility out of all available facilities. This is also an indicator of the utilization of production resources.
- A detailed characterization of the lead times of machinery.

Statistical evaluation within the software provides the possibility of a broad overview of processed data. Therefore, in this case, the sequence of steps is as follows:

1. Analysis of available attributes and gaining an overview of the current state: The case study focuses on orders and machinery equipment;
2. Identification of critical orders and equipment: In this case, orders with the longest lead times and the most utilized equipment are searched for (Figure 10):



- Order number “A143503-1218326” includes 10 activities, with a total duration of 29 days and 5 h;
- Order number “C402200-7155104” includes 11 activities, with a total duration of 25 days and 18 h;
- Machineries numbered “999201” and “942101” are the most active and, together, make up more than 1/3 (36%) of the share of the entire manufacturing process compared to other machineries.

Case ID	Events	Variant	Started	Finished	Duration
A143503-1218326	10	Variant 1	02.10.2022 14:31:00	31.10.2022 19:00:00	29 days, 5 hours
B433453-4066770	11	Variant 2	02.10.2022 08:10:00	10.10.2022 06:36:00	7 days, 22 hours
C402200-7155104	11	Variant 3	03.10.2022 03:44:00	28.10.2022 21:59:00	25 days, 18 hours
A143503-1218325	3	Variant 4	03.10.2022 01:11:00	03.10.2022 13:46:00	12 hours, 35 mins
D376200-4144272	8	Variant 5	04.10.2022 07:02:00	10.10.2022 03:14:00	5 days, 20 hours
A143503-1218327	7	Variant 6	13.10.2022 07:31:00	26.10.2022 19:57:00	13 days, 12 hours
A143503-7202342	10	Variant 7	18.10.2022 03:09:00	31.10.2022 20:25:00	13 days, 18 hours
C402200-7155105	13	Variant 8	19.10.2022 09:15:00	28.10.2022 23:06:00	9 days, 13 hours
D376200-4144273	10	Variant 9	22.10.2022 09:12:00	02.11.2022 12:28:00	11 days, 4 hours
A143503-7202343	7	Variant 10	28.10.2022 02:04:00	11.11.2022 10:25:00	14 days, 9 hours
A143503-7202344	4	Variant 11	04.11.2022 19:26:00	06.11.2022 17:38:00	1 day, 22 hours
A143503-7202345	7	Variant 12	08.11.2022 23:41:00	09.11.2022 23:54:00	1 day, 13 mins

Activity	▲ Frequency	Relative frequency	Median duration	Mean duration	Duration range
999201	19	18.81 %	18 mins	1 hour, 3 mins	10 hours, 27 mins
942101	17	16.83 %	11 mins	2 hours, 25 mins	14 hours, 9 mins
3278105	10	9.9 %	16 mins	19 mins, 54 secs	49 mins
999301	8	7.92 %	6 mins	10 mins, 30 secs	26 mins
2613104	8	7.92 %	10 mins, 30 secs	16 mins, 30 secs	49 mins
941201	8	7.92 %	14 mins, 30 secs	1 hour, 12 mins	8 hours, 14 mins
1385401	7	6.93 %	28 mins	32 mins	1 hour, 45 mins
384301	6	5.94 %	9 mins, 30 secs	13 mins	27 mins
942201	6	5.94 %	17 mins	57 mins, 50 secs	4 hours, 23 mins
382501	3	2.97 %	10 mins	16 mins	20 mins
1622501	2	1.98 %	5 mins	5 mins	10 mins
464101	2	1.98 %	32 mins, 30 secs	32 mins, 30 secs	45 mins
942103	2	1.98 %	16 mins, 30 secs	16 mins, 30 secs	19 mins
333602	2	1.98 %	48 mins, 30 secs	48 mins, 30 secs	1 hour, 31 mins
3278104	1	0.99 %	24 mins	24 mins	0 millis

Figure 10. Statistical results of orders and machinery equipment.

### 3.1. Use of Filters to Reveal Selected Processes and Equipment

The software reprocesses the information based on selected criteria using filters. In line with the previous step, the selection was narrowed down to the following:

- Order numbers “A143503-1218326” and “C402200-7155104”;
- Machinery numbers “999201” and “942101”.

### 3.2. Analysis of the Process Map

After processing the new data limited by filters, the software redraws the process map, which is displayed in the figure (Figure 11). The process map points to the following:

- Long handling times. This could be caused by various types of waste: transport, inventory, waiting, overproduction, errors, and more.
- Long processing and retooling times within machinery equipment. In this case, the machines “942101” and “999201” were identified as the most utilized. Waste can come from a variety of sources, including transportation, inventory, waiting, overproduction, oversights, and more.
- Long rework time within machinery “942101”: 31.5 days (Figure 12). Such a problem can be caused by poorly performed operations, reduction in order prioritization, extension of the order lifecycle, waste within other workplaces, lack of material, and human error.

For a better analysis, it is possible to use internal company information from the planning department to find out what the main time difference is between the planned and actual statuses. Subsequently, it is advisable to have records of the activities of the machinery and equipment to find out why some cycle times do not match the planned times.

Detailed analysis of a selected case: Case evaluation within the software provides the opportunity to dissect a specific case in more detail. In the case study, it is necessary to

analyze selected attributes: orders and machines. Figure 13 shows the progression of order A143503-1218326 with more detailed characteristics.

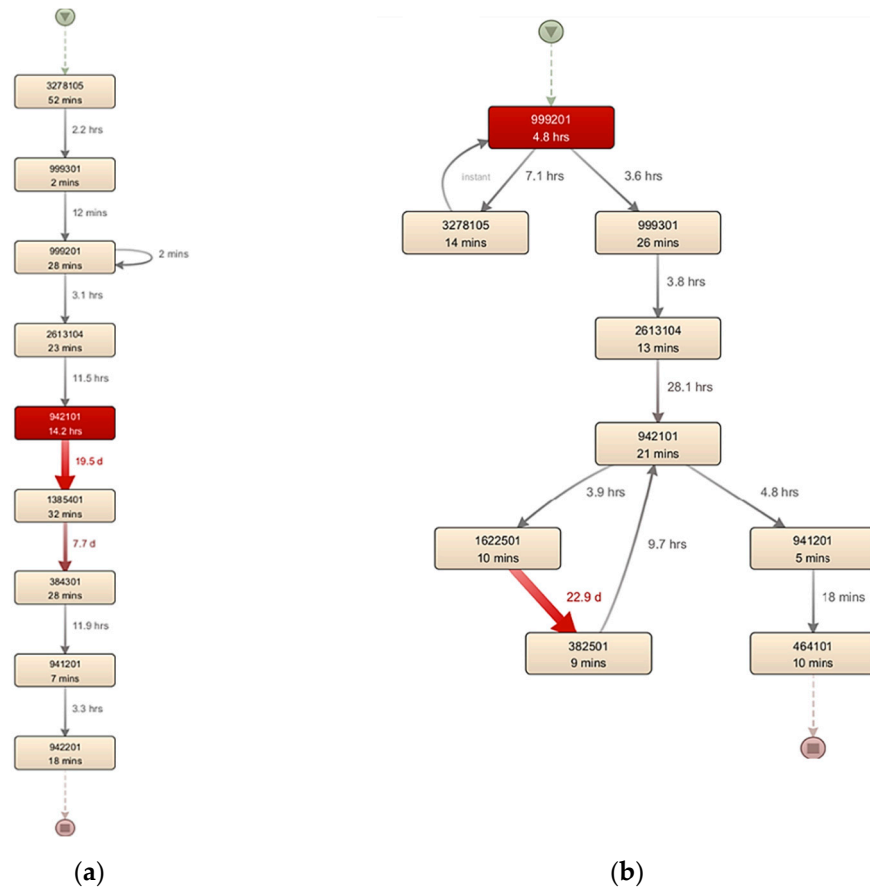


Figure 11. Process maps of selected orders of (a) “A143503-1218326” and (b) C402200-7155104.

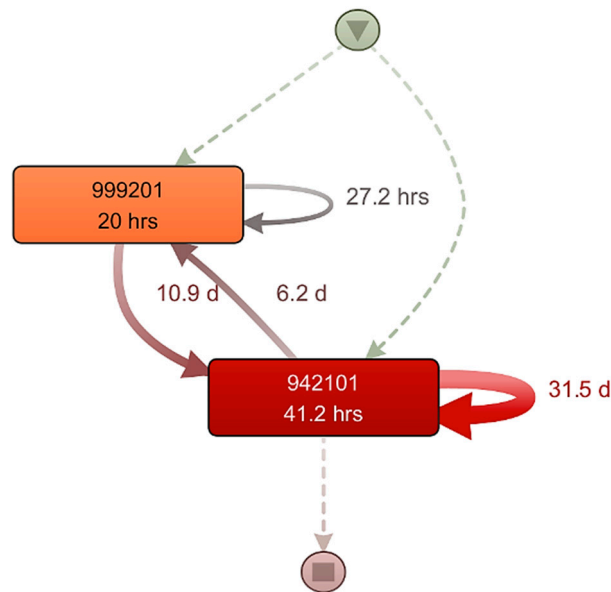


Figure 12. Process map of selected critical machinery equipment—“999201 and 942101”.

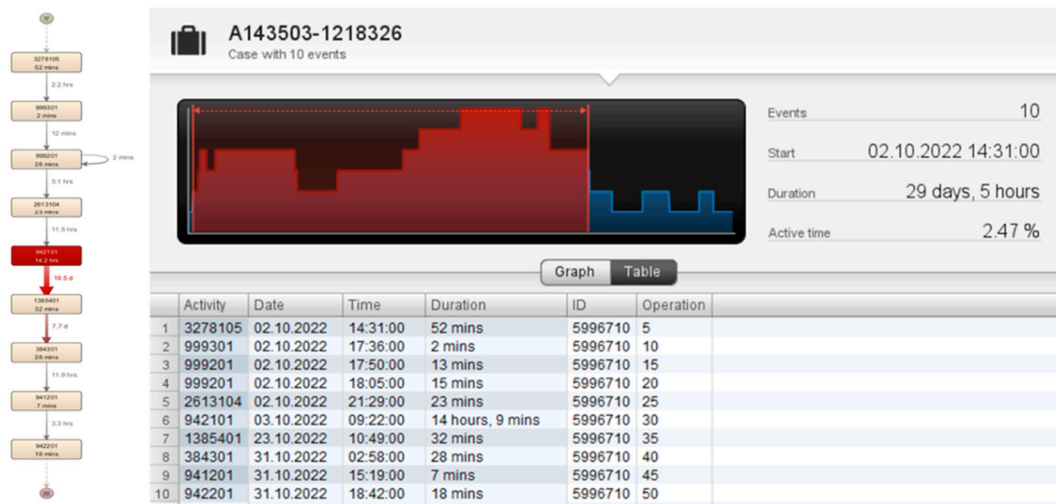


Figure 13. Detailed analysis of order event “A143503-1218326”.

### 3.3. Outcomes of the Analysis

The results of the analyses were then processed and interpreted through the examination of the following KPIs:

- A map of the sequence of the overall manufacturing process and individual orders;
- Time utilization of all or selected critical equipment;
- Rework analysis (quality indicator);
- Lead times;
- Number of work-in-progress manufacturing orders;
- Effective times;
- Lifecycle of the order.

Additionally, column charts processed based on data exported from the Fluxicon Disco tool for a period of 2 months are illustrated (Figure 14).

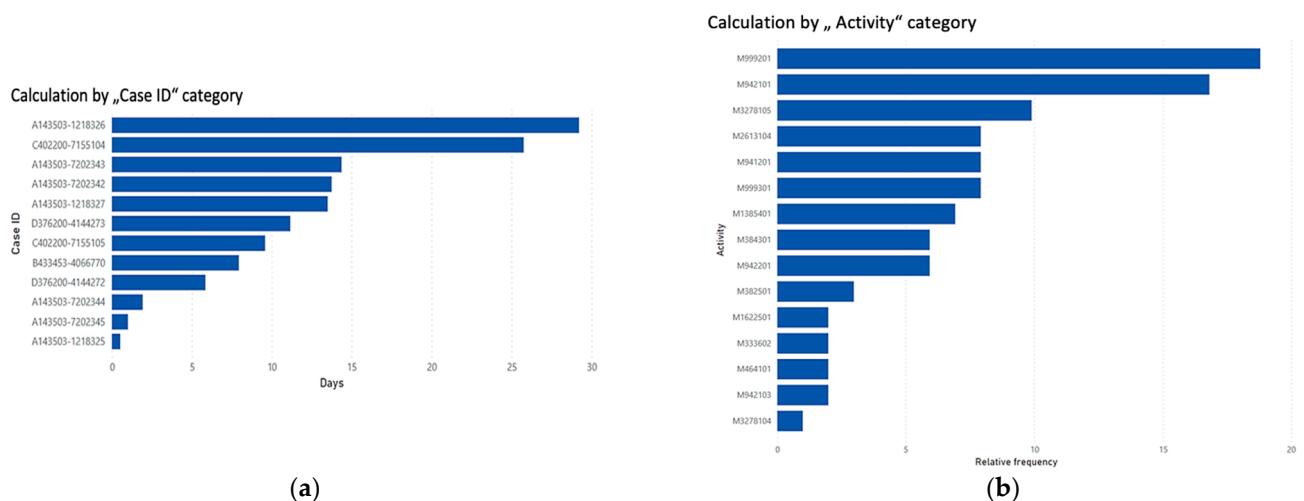


Figure 14. Statistical graphs of the (a) duration of order in days and (b) machinery utilization percentage share.

During the analysis, the following problematic orders and machinery equipment were identified:

1. Order “A143503” with work order “1218326”, operation of technological procedure “30”, and machinery equipment “942101”. Identified problem:
  - The processing time of the order itself is 29 days and 5 h;



- The machining time of processing machinery equipment is 14.2 h;
  - Logistical time for material delivery from machine 942101 to machine 1385401: 19.5 days;
  - Rework time on machinery equipment (999201): 17 min.
2. Order “C402200” with work order “7155104”, operation of technological procedure “5”, and machinery equipment “999201”. Identified problem:
- The processing time of the order itself is 29 days and 5 h;
  - The machining time of processing machinery equipment is 14.2 h;
  - Logistical time for material delivery from machine 942101 to machine 1385401: 19.5 days;
  - Rework time on machinery equipment (999201): 17 min.

By reducing the main identified problems (reducing WIP and shortening lead time), there is a potential to increase the efficiency of ongoing processes by at least 30%.

One of the advantages of process mining analysis is the ability to segment machinery equipment based on exported data. Through segmentation, it is possible to divide equipment into groups based on the percentage similarity of performed operations or the frequency of occurrences across all orders. In this case, three groups of equipment were created, based on which it was possible to propose a cellular production arrangement with the aim of shortening lead time and, overall, increasing the efficiency of manufacturing and logistical processes (Figure 15).

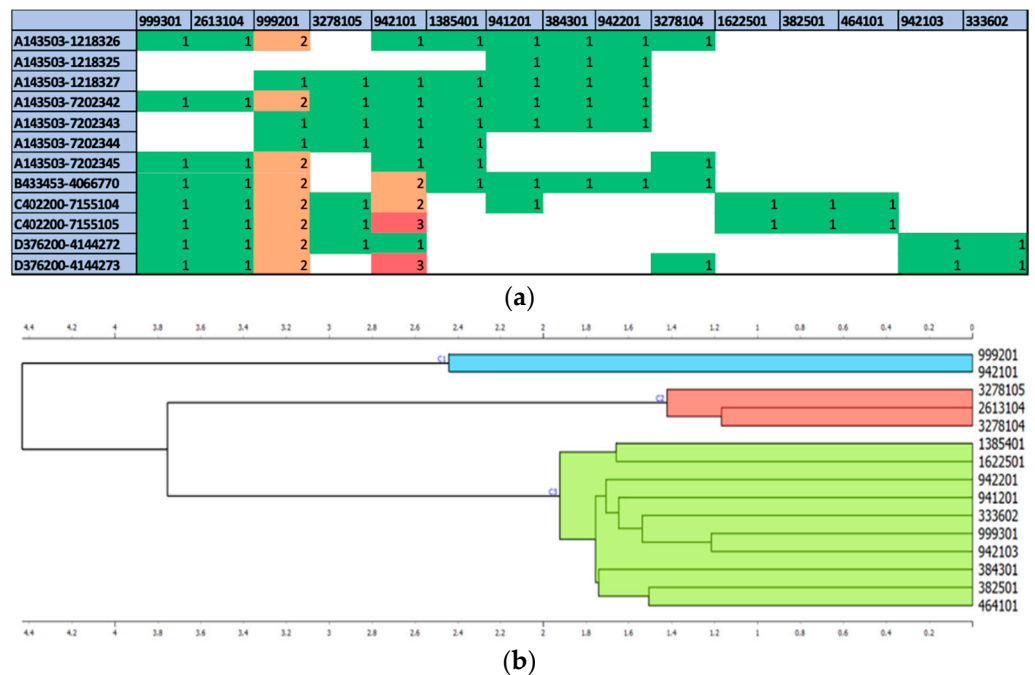


Figure 15. Statistical results of the main attributes: (a) “Order-Machine” relationship table; (b) machinery equipment segmentation.

#### 4. Discussion

The analysis of industrial production and its processes is the foundation for the effective management and optimization of industrial production. The complexity of these processes and the dynamic environment in which they operate often complicate their proper understanding and management. In this context, the process mining method appears to be a powerful tool that allows enterprises to better understand and improve their processes.

Before we move to the concluding remarks, it is important to consider the managerial perspective on our model and discuss the key identified advantages. The integration of the proposed process mining technique into the management of manufacturing processes

brings significant benefits for decision-making and strategic planning. The main findings include the following:

1. **Increased process transparency:** Our model provides a deep insight into operational processes, enabling managers to identify key bottlenecks and effectively monitor process performance in real-time.
2. **Resource optimization:** The analysis revealed opportunities for better utilization of resources and reduction of waste, leading to cost reductions and increased efficiency.
3. **Support for data-driven decision-making:** With more accurate predictions of performance and potential risks, managers can make more informed decisions, enhancing the adaptability and competitiveness of the organization [34].
4. **Improvement in quality and customer satisfaction:** The model allows for a better alignment of processes with customer needs and expectations, leading to a significant improvement in the quality of products and services.

Owing to process mining, it is possible to discover hidden patterns, dependencies, and trends in process data that might be missed by traditional analytical methods. This method of data science offers a new way to understand and optimize production processes, which ultimately leads to higher productivity, increased efficiency, and profitability for the enterprise. The article presents the authors' own methodology for the implementation of process mining. In the model example, we successfully achieved a reduction in the time required for production cycles by 15% and improved resource utilization by 20%. This resulted in increased process efficiency and a potential reduction in the required number of workers by up to 10%. These outcomes offer promising evidence of the advantages of our method and its application in an industrial setting.

The benefits of using this methodology are as follows:

1. **Quick analysis of massive statistical sets of all orders**

Using process mining tools allows for efficiently identifying trends and patterns. These tools automatically visualize data, facilitating the identification of problems and opportunities and speeding up processing and outcomes. Additionally, owing to a quick analysis, it is possible to immediately respond to deviations and improve process performance. Discovering hidden patterns and revealing relationships between activities assists in rational decision-making.

2. **Uncovering inefficiencies and opportunities for improvement**

Even well-designed production processes can often contain hidden waste or issues that can significantly affect performance and productivity. By analyzing production processes with the process mining method, organizations can uncover these hidden problems and identify areas where improvement is possible. These improvements can be in the form of reduced lead times, reduced costs, better quality outputs, or increased customer satisfaction.

3. **Understanding the true performance of processes**

Although many organizations have detailed models and documentation of their production processes, the actual execution of these processes can significantly differ from planned scenarios. By analyzing actual data from process logs with the process mining method, organizations can obtain a detailed picture of the true performance and utilization of their processes. This information can help organizations better understand their processes and identify KPIs that need improvement.

4. **Predicting and preventing problems**

Process mining allows organizations not only to analyze the past performance of their processes but also to predict future outcomes. Analytical tools for process mining can use historical data to create predictive models that can forecast future outcomes or identify potential issues before they occur. This way, organizations can prevent problems and optimize their processes for the future.

The process mining method is a very powerful tool for process analysis, but it also has its limitations and constraints that are important to consider:

1. Dependency on input data

The quality of process mining results depends on the quality of the input data. If the data are incomplete, inconsistent, or missing, it can affect the reliability and accuracy of the analysis. Process mining can also be limited by existing data records. If there are insufficient records of processes, the analysis may be inaccurate or impossible.

2. Time-consuming

Processing a large amount of process data can be time-consuming. This can be a problem for organizations with a huge volume of data. Likewise, the complexity of interpreting results exists, which can be complex. There is a need for the ability to analyze and understand process patterns. This also depends on the technical competence of the person conducting the analysis and, of course, the ability to work with analytical software.

3. Limited predictive capacity and analysis time period:

Process mining is primarily a method for analyzing past processes. It does not have strong predictive capabilities and may not accurately predict future events. Another significant point is that processes can change over time, so it is important to perform analysis within the correct time period. Old data may not fully reflect the current state of processes.

4. Alternative methods of analysis

There are also other methods of process analysis, such as Six Sigma, LEAN, or traditional data analysis methods. For some organizations, these methods may be more suitable or effective. Process mining is not always the best tool for every kind of process analysis. It is important to consider the specific use case and decide whether this method is the most appropriate.

Despite the many advantages of process mining, it is important to be aware of its limitations and to prepare thoroughly for its implementation. It is not a universal tool, and its success depends on many factors, including the quality of the data and the analyst's skills.

The discussion of these advantages highlights the significant potential of the proposed technique for improving management and operations in organizations. The implementation of this approach can serve as a foundation for the development of more sustainable, efficient, and profitable manufacturing processes.

## 5. Conclusions

The article presents a proprietary methodological approach to the analysis of manufacturing processes using process mining tools. The schematic development diagram, which was created as part of the study, includes specific steps and decisions that allow for a systematic analysis and improvement of manufacturing processes in an organization. This procedure provides valuable insights into process flows, identifies areas for improvement, and contributes to a more efficient and competitive manufacturing environment.

The results of the manufacturing process analysis using process mining are proof of its importance and contribution to production management. The study aimed to explore the possibilities of using this method in the analysis and optimization of manufacturing processes and improving efficiency. Based on the obtained results, several key findings can be identified.

The first finding is that process mining allows for the discovery of hidden patterns and variability in the analyzed manufacturing processes. It helps identify and visualize various activities and their relationships, supporting managers and analysts in better understanding the course of processes and identifying potential deficiencies or problems.

The second benefit is the ability to identify inefficient steps and the possibility of their optimization. By analyzing data, it is possible to identify activities with excessive duration,

repeated steps, or unnecessary delays, allowing for a focus on key areas for improvement and increasing the overall efficiency of production.

The third benefit is the increased transparency of manufacturing processes. Process mining allows for the detailed viewing and tracking of data and activity flows, which aids in better process management and monitoring and contributes to improved communication between different departments and stakeholders.

In conclusion, process mining provides significant advantages in the analysis and optimization of manufacturing processes. The achieved results and findings are in line with the original objectives and research questions of the study and contribute to the development of the professional field of manufacturing process analysis, providing practical insights for production management.

Further research in this area should aim to explore additional factors and areas of manufacturing processes, as well as to expand the application of process mining methods to other industrial sectors and organizations. Collaboration between academia and industry is key to the successful application of these methods and to achieving specific improvements in the management of manufacturing processes.

Overall, this study contributes to the professional field of manufacturing process analysis, and its findings have the potential to influence practical procedures and improve the management of manufacturing processes. It is important to continue future research and further explore the opportunities offered by process mining.

**Author Contributions:** Conceptualization, M.K. and V.B.; methodology, V.B.; software, V.B.; validation, V.B., M.B. and M.K.; formal analysis, B.F.; investigation, B.F.; resources, M.B.; data curation, D.K.; writing—original draft preparation, V.B.; writing—review and editing, D.K.; visualization, V.B.; supervision, M.K.; project administration, M.K.; funding acquisition, B.F. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Slovak Research and Development Agency, grant number APVV-19-0305.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author.

**Acknowledgments:** This work was supported by the Slovak Research and Development Agency under the contract No. APVV-19-0305.

**Conflicts of Interest:** Author Milan Botka was employed by the company Asseco CEIT a.s. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## References

1. Gregor, M.; Hodon, R.; Binasova, V.; Dulina, L.; Gašo, M. Design of Simulation-Emulation Logistics System. *MM Sci. J.* **2018**, *2018*, 2498–2502. [[CrossRef](#)]
2. Schuh, G.; Gützlaff, A.; Schmitz, S.; Kuhn, C.; Klapper, N. Process Mining as Alternative to Traditional Methods to Describe Process Performance in End-to-End Order Processing of Manufacturing Companies. *J. Adv. Manag. Sci.* **2022**, *10*, 1–8. [[CrossRef](#)]
3. Stefanovic, D.; Dakic, D.; Stevanov, B.; Lolic, T. Process Mining in Manufacturing: Goals, Techniques and Applications. In *Advances in Production Management Systems. The Path to Digital Transformation and Innovation of Production Management Systems*; IFIP Advances in Information and Communication Technology; Springer International Publishing: Cham, Switzerland, 2020; pp. 54–62. [[CrossRef](#)]
4. Luo, H.-J.; Wang, S.; Chen, C.; Tai, H.-E.; Chien, C.-F.; Chen, P.-N. Using Process Experienced Correlation Table to Improve the Accuracy and Reliability of Data Mining for Yield Improvement. In *2004 Semiconductor Manufacturing Technology Workshop Proceedings*; IEEE: Hsinchu, Taiwan, 2004; pp. 127–130. [[CrossRef](#)]
5. Susilawati, A.; Tasri, A.; Arief, D.S. Yohanes Wastes Analysis to Improve the Productivity and Sustainability in Manufacturing Industry. *J. Phys.Conf. Ser.* **2019**, *1351*, 012110. [[CrossRef](#)]
6. Subhashish, S.; Lance, H. Waste Elimination: The Common Denominator for Improving Operations. *Ind. Manag. Data Syst.* **1993**, *93*, 13–19. [[CrossRef](#)]
7. Dewi, S.; Utama, D.M.; Rohman, R.N. Minimize Waste on Production Process Using Lean Concept. *J. Phys.Conf. Ser.* **2021**, *1764*, 012201. [[CrossRef](#)]

8. Kučerová, M.; Mílkva, M.; Sablik, J.; Gejguš, M. Eliminating Waste in the Production Process Using Tools and Methods of Industrial Engineering. *PEA* **2015**, *9*, 30–34. [[CrossRef](#)]
9. Dreher, S.; Reimann, P.; Gröger, C. Application Fields and Research Gaps of Process Mining in Manufacturing Companies. *INFORMATIK* **2021**, 621–634. [[CrossRef](#)]
10. Abonyi, J.; Dorgo, G. Process Mining in Production Systems. In Proceedings of the 2019 23rd International Conference on Intelligent Engineering Systems (INES), Gödöllő, Hungary, 25–27 April 2019; IEEE: Gödöllő, Hungary, 2019; pp. 000267–000270. [[CrossRef](#)]
11. Chiu, T.; Jans, M. Process Mining of Event Logs: A Case Study Evaluating Internal Control Effectiveness. *Account. Horiz.* **2019**, *33*, 141–156. [[CrossRef](#)]
12. Gola, A. Design and Management of Manufacturing Systems. *Appl. Sci.* **2021**, *11*, 2216. [[CrossRef](#)]
13. Beric, D.; Havzi, S.; Lolic, T.; Simeunovic, N.; Stefanovic, D. Development of the MES Software and Integration with an Existing ERP Software in Industrial Enterprise. In Proceedings of the 2020 19th International Symposium INFOTEH-JAHORINA (INFOTEH), East Sarajevo, Bosnia and Herzegovina, 18–20 March 2020; IEEE: East Sarajevo, Bosnia and Herzegovina, 2020; pp. 1–6. [[CrossRef](#)]
14. Więcek, D.; Więcek, D.; Dulina, L. Materials Requirement Planning with the Use of Activity Based Costing. *Manag. Syst. Prod. Eng.* **2020**, *28*, 3–8. [[CrossRef](#)]
15. Crnjac, M.; Veža, I.; Banduka, N. From Concept to the Introduction of Industry 4.0. *Int. J. Ind. Eng. Manag.* **2017**, *8*, 21–30. [[CrossRef](#)]
16. Lopes, I.; Figueiredo, M. Criticality Evaluation to Support Maintenance Management of Manufacturing Systems. *Int. J. Ind. Eng. Manag.* **2020**, *11*, 3–18. [[CrossRef](#)]
17. Pekarčíková, M.; Trebuňa, P.; Markovič, J. Case Study of Modelling the Logistics Chain in Production. *Procedia Eng.* **2014**, *96*, 355–361. [[CrossRef](#)]
18. Van Der Aalst, W.; Weijters, T.; Maruster, L. Workflow Mining: Discovering Process Models from Event Logs. *IEEE Trans. Knowl. Data Eng.* **2004**, *16*, 1128–1142. [[CrossRef](#)]
19. Da Silva, L.F.N. Process Mining: Application to a Case Study. 2014. Available online: <https://core.ac.uk/download/pdf/302923848.pdf> (accessed on 21 November 2014).
20. Van Der Aalst, W.M.P.; Reijers, H.A.; Weijters, A.J.M.M.; Van Dongen, B.F.; Alves De Medeiros, A.K.; Song, M.; Verbeek, H.M.W. Business Process Mining: An Industrial Application. *Inf. Syst.* **2007**, *32*, 713–732. [[CrossRef](#)]
21. Pena, M.R.; Bayona-Ore, S. Process Mining and Automatic Process Discovery. In Proceedings of the 2018 7th International Conference on Software Process Improvement (CIMPS), Guadalajara, Mexico, 17–19 October 2018; IEEE: Guadalajara, Jalisco, Mexico, 2018; pp. 41–46. [[CrossRef](#)]
22. Ďurica, L.; Gregor, M.; Vavřík, V.; Marschall, M.; Grznár, P.; Mozol, Š. A Route Planner Using a Delegate Multi-Agent System for a Modular Manufacturing Line: Proof of Concept. *Appl. Sci.* **2019**, *9*, 4515. [[CrossRef](#)]
23. Van Der Aalst, W.M.P. *Process Mining: Discovery, Conformance and Enhancement of Business Processes*; Springer: Berlin/Heidelberg, Germany, 2011. [[CrossRef](#)]
24. Van Der Aalst, W. Spreadsheets for Business Process Management: Using Process Mining to Deal with “Events” Rather than “Numbers”? *BPMJ* **2018**, *24*, 105–127. [[CrossRef](#)]
25. Trebuna, P.; Straka, M.; Rosova, A.; Malindzakova, M. Petri nets as a tool for production streamlining in plastics processing. *Przem. Chem.* **2015**, *94*, 1605–1608.
26. Kregel, I.; Stemann, D.; Koch, J.; Coners, A. Process Mining for Six Sigma: Utilising Digital Traces. *Comput. Ind. Eng.* **2021**, *153*, 107083. [[CrossRef](#)]
27. Weerapong, S.; Porouhan, P.; Premchaiswadi, W. Process Mining Using Alpha-Algorithm as a Tool (A Case Study of Student Registration). In Proceedings of the 2012 Tenth International Conference on ICT and Knowledge Engineering, Bangkok, Thailand, 21–23 November 2012; IEEE: Bangkok, Thailand, 2012; pp. 213–220. [[CrossRef](#)]
28. Van Der Aalst, W.M.P.; Van Dongen, B.F. Discovering Petri Nets from Event Logs. In *Transactions on Petri Nets and Other Models of Concurrency VII*; Jensen, K., Van Der Aalst, W.M.P., Balbo, G., Koutny, M., Wolf, K., Eds.; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2013; Volume 7480, pp. 372–422. [[CrossRef](#)]
29. Van Der Aalst, W.M.P. A Practitioner’s Guide to Process Mining: Limitations of the Directly-Follows Graph. *Procedia Comput. Sci.* **2019**, *164*, 321–328. [[CrossRef](#)]
30. Buijs, J.C.A.M.; van Dongen, B.F.; van der Aalst, W.M.P. On the Role of Fitness, Precision, Generalization and Simplicity in Process Discovery. In *On the Move to Meaningful Internet Systems: OTM 2012*; Meersman, R., Panetto, H., Dillon, T., Rinderle-Ma, S., Dadam, P., Zhou, X., Pearson, S., Ferscha, A., Bergamaschi, S., Cruz, I.F., Eds.; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2012; pp. 305–322. [[CrossRef](#)]
31. Janík, S.; Szabó, P.; Mílkva, M.; Mareček-Kolibiský, M. Effective Data Utilization in the Context of Industry 4.0 Technology Integration. *Appl. Sci.* **2022**, *12*, 10517. [[CrossRef](#)]
32. Fluxicon. Technical Documentation of Disco Software—Import Data Sets. Disco User Guide. Available online: <https://fluxicon.com/book/read/statisticsview/> (accessed on 1 July 2022).

33. Fluxicon. Technical Documentation of Disco Software—Analyzing Statistics. Disco User Guide. Available online: <https://fluxicon.com/book/read/import> (accessed on 1 July 2022).
34. Grznár, P.; Burganová, N.; Mozol, Š.; Mozolová, L. A Comprehensive Digital Model Approach for Adaptive Manufacturing Systems. *Appl. Sci.* **2023**, *13*, 10706. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.