Process Mining on ERP systems

BATU BUDIN, University of Twente, The Netherlands

Enterprises allocate investments with the anticipated Return On Investment (ROI) encompassing cost reduction, profit maximization, decision support, and fraud detection. In a competitive business environment, companies need to streamline and optimize their procurement processes. An ingredient for this kind of optimizations could be the use of information systems such as Enterprise Resource Planning (ERP) systems. Despite the prevalence of ERP systems in these organizations, many lack the proficiency to extract valuable insights from the processes. This process-centric nature aligns well with Process Mining techniques. This study is expected to contribute to the growing body of literature on the use of data-driven approaches in manufacturing organizations. To address this issue, this research aims to explore the correlation between ERP systems and process mining techniques. By conducting a systematic review of the literature, we will identify existing research on this topic and analyze the methods and results of previous stud-ies and imply the findings on data collected from a company. The goal of this research paper is to provide insights into how ERP systems and process mining techniques can be integrated to improve organizational eficiency and decision-making. With the acknowledged constraints of time and the data quality issues that will be investigated using limited techniques as emphasized in this research paper, we plan to utilize conformance-checking techniques to identify irregularities. Subsequently, we aim to assign significance to each anomaly and initiate discussions with the company for further resolution.

Additional Key Words and Phrases: ERP, Data Mining, Process Mining, bottleneck, ROI, process-centric, confermance-checking

1 Introduction

Data mining techniques have gained popularity [5] as a means of analyzing large datasets to extract valuable insights and identify bottlenecks in business processes. In the context of ERP systems, these techniques can be particularly useful due to the large amount of data generated by these systems. With the usage of process mining techniques, organizations can use process mining to discover their actual processes, offer insights, identify issues, and automatically initiate corrective actions [22]. Some process mining methodologies leverage supplementary data, including the entity (be it a person or a device) responsible for executing or initiating an activity, the timestamp (date and time) of the event, and accompanying data elements associated with the event (order quantity), to their advantage [20]. The utilization of limited manufacturing resources can be improved, the system throughput can be increased, and the overall cost of production can be reduced by accurately and effectively locating bottleneck spots[15].

ERP systems have emerged as one of the most extensively used business systems in recent years, changing the organizational infrastructures of corporations from functionally centered to procedure-driven [6]. To increase competitiveness, practically every company now considers ERP to be a "must have" technology [26]. The main objective of this paper is to leverage Budin Akarca's ERP Systems data for the purpose of detecting bottlenecks in their processes. Despite the existence of some research on manufacturing and big data in relation to process mining, there is a notable dearth of studies specifically addressing Process Mining for ERP systems, apart from the relevant work discussed in the third part. This indicates a significant gap in the current body of knowledge, emphasizing the necessity for further exploration and investigation in this domain.

In the upcoming sections, we will proceed with defining our research questions in alignment with the previously stated goal. The second section will specifically outline the main research question as well as the accompanying sub-questions. Following that, the third section will delve into the related work and existing literature pertaining to the topic under examination.

Additionally, the fourth part of the paper will elucidate the methods that will be employed in the research. This section will outline the specific approaches, techniques, and tools that will be utilized to address the research questions effectively.

Following that, the fifth section will delve into providing the necessary background information to comprehend process mining techniques and the analysis process. This section will offer a comprehensive overview of the key concepts, methodologies, and principles involved in process mining.

Lastly, the case study will be presented, detailing the practical choices made in order to detect bottlenecks in the company's data. This section will elucidate the specific steps taken, the data utilized, and the analytical techniques employed to identify and address the bottlenecks within the company's processes.

2 Research Question

To effectively analyze the data provided by the company, it is crucial to establish a foundation of background information regarding ERP systems and Process Mining techniques. This will enable a comprehensive understanding of the research context. In this study, we will address one main research question and three sub-questions to guide our analysis and exploration. These research questions will serve as a framework for our investigation and provide clarity in addressing the objectives of the study.

"How can process mining techniques be applied to analyze data from ERP systems in order to identify bottlenecks and extract valuable insights?"

To answer this question we have to answer the following research questions.

RQ1: What are the various types of data that contemporary ERP system's modules provide?

RQ2: What Process Mining techniques are applicable in the context of ERP systems?

RQ3: What specific data requirements are necessary for the successful implementation of the process mining techniques identified in RQ2?

Author's address: Batu Budin, b.budin@student.utwente.nl, University of Twente, P.O. Box 217, Enschede, The Netherlands, 7500AE.

3 Related Work

In this section, we explore the existing body of literature and research that is relevant to our study on applying process mining techniques to analyze data from ERP systems. Our aim is to gain insights from previous work and identify key findings and methodologies that have contributed to the field.

While there exists a wealth of data-driven Process Mining approaches documented in academic sources, a comprehensive and step-by-step explanation of Process Mining specifically tailored for ERP systems has been found lacking. However, noteworthy sources that contribute to the understanding of Process Mining in the context of ERP systems included [11, 14, 17, 18, 20, 22] The first most relevant publication found, the authors [18], use the Systems, Applications and Products Manufacturing Execution (SAP ME) in their example with the usage of process mining and value tracing. After analyzing the related work from the given authors, it offers examples for descriptive models such as conformance checking (checking two different models), streaming (monitoring real-time events), performance analysis (bottleneck analysis), and decision support systems[18]. Also gives an example of process mining algorithms such as Alpha Miner with Petri Net Notation, Inductive Miner, and Fuzzy Miner [18], the description of these miners will be explained later in the research.

The second article [17] applies process mining algorithms to a dataset comprising event logs from Oracle Financials-based enterprise resource planning (ERP) procurement processes from an organization in Pakistan. Their objectives include reducing the time between purchase generation and approval, identifying unnamed activities, and allocating the budget eficiently. The study introduces metrics that allow for the identification of both low-level and high-level deviations between the event log and the process model. This research aims to analyze ERP procurement log data to find an eficient process model that reflects the organization's actual flow. It also identifies correlated sub-processes, optimal time frames, and anomalies for improvement. The findings of the authors are the detection of five types of issues with quality in the logs through the use of different process mining tools from the ProM process mining tool.

4 Methodology

In this section, we will delve into the methodologies employed to answer the research questions and address the main research objective of this thesis, which is to explore how process mining techniques can be applied to analyze data from ERP systems in order to identify bottlenecks and extract valuable insights. To achieve this, we will use CRISP-DM [19] methodology which is constructed with 6 steps namely; business understanding, data understanding, data preparation, modeling, evaluation, and conclusion. The CRoss Industry Standard Process for Data Mining [19] technique plays a crucial role in Process Mining by guaranteeing quality and producing an industry-neutral methodology. Despite being used for two decades, the publication still adheres to De-Facto data mining standards [19]. To begin, our initial step involves investigating the diverse types of data offered by contemporary ERP system modules. Subsequently, we will explore process mining techniques that are applicable within the context of ERP systems. Furthermore, we will identify the specific data requirements necessary for the successful implementation of these techniques. Once these preparatory information's are given, we will proceed with the application of the CRISP-DM methodology, utilizing the company's data for analysis and further refinement.

The first research question, it aims to uncover the diverse types of data that are made available by modern ERP system modules. By examining these data types, we can gain a comprehensive understanding of the information landscape within ERP systems and lay the foundation for subsequent analyses. This exploration will enable us to ascertain the richness and granularity of the data, which will prove crucial in effectively applying process mining techniques.

Moving forward, second research question focuses on identifying process mining techniques that are applicable in the context of ERP systems. Process mining provides a valuable framework for analyzing and improving business processes by leveraging event data recorded in information systems [22]. By studying the current literature and industry practices, we will identify process mining techniques that are well-suited for analyzing the data within ERP systems. This exploration will enable us to determine the most relevant and effective techniques for our analysis.

To successfully implement the process mining techniques identified, seeks to ascertain the specific data requirements necessary for their application. This research question will guide us in identifying the essential data elements, attributes, and context necessary for accurately performing process mining analyses within an ERP system environment. Understanding these data requirements will contribute to the successful implementation of the chosen process mining techniques and enhance the quality and reliability of our subsequent findings.

Ultimately, the usage of these methodologies will allow us to address our main research question, which seeks to determine how process mining techniques can be applied to analyze data from ERP systems. By using the insights gained from the exploration of data types, the identification of applicable process mining techniques, and the understanding of specific data requirements, we will be able to conduct an in-depth analysis of Budin Akarca's [2] ERP data by following Crisp-DM [19] model using Prom-Lite 1.3.

In the following sections, we will discuss each research question in detail, outlining the research methodologies employed and the rationale behind their selection. Furthermore, we will provide a comprehensive overview of the data collection process, the analytical techniques used in recent related papers, while adhering the framework of the CRISP-DM methodology.

5 Results

This section aims to furnish the necessary background information to facilitate a comprehensive understanding of the analysis performed and the diagrams generated by Process Mining Tools (ProM) while addressing the research questions. By providing this contextual information, we intend to offer the reader a solid foundation for interpreting and grasping the subsequent analytical processes and visual representations presented in this study.

5.1 Modules and the Data provided by ERP Systems

It is stated that [4] process mining's first and most costly stage is the extraction, transformation, and loading of event logs from information systems. The traces that processes leave behind while they run are recorded in event logs[16]. This requires comprehending the fundamentals of the data, which are now found in ERP systems. Our goal is to identify the modules that are suited for producing accurate insights after studying the types of data generated by these systems. The report will answer this, however this section will detail the data for current ERP systems. One of the approaches will be to look at major businesses who have implemented ERP systems and the modules they focused on. And later understanding the data extracted from these Information Systems.

From the articles [3, 7, 10], it is mentioned that these Information Systems break down the departments in these enterprises in order to increase eficiency, decision support and reduction of the costs [7]. ERP is a software program that combines these departments and creates functional modules within big databases. Authors [3, 7, 9, 10] did concatenate these functional units into; Manufacturing, Sales, Human Resource, Finance, Supply Chain, Warehousing, Customer Relation, and Salary Management. These modules differ for each enterprise and some may need additional separate modules to support their businesses [9].

Gaining understanding of the data type alternatives offered by the current information systems is the aim of the next paragraph. The presence of event logs is a fundamental facilitator of process mining because study of run-time behavior is only possible if events are recorded [24]. Thankfully, all different kinds of information systems provide these logs such as SAP system, which is the most popular ERP system available [4]. These systems provide highly detailed information on the tasks that have been completed which is the essential in process mining [24]. CSV, XES and OCEL are types that are produced event data's by these systems [22].

The third research question aims to conduct a comprehensive investigation into the data storage types of CSV, XES, and OCEL.

5.2 Process Mining Techniques in the context of ERP Systems

Event data is pervasive and will continue to increase exponentially in all industries, economies, organizations, and even at homes [20]. Event log is the start point of process mining [20] and it is evident that ERP systems possess the necessary capabilities to generate such data. Process mining is related to data mining [20]. Whereas classical data mining techniques are mostly data-centric, process mining is process-centric [20]. Process mining approaches offer insights, spot bottlenecks and deviations, foresee and identify performance and compliance issues, and help the automation or elimination of repetitive work by combining event data and process models [22]. This part of the study aims to investigate and evaluate the applicability of process mining methods to ERP systems. Process modeling and techniques to create these models, conformance checking, performance analysis, and comparative and predictive process mining will be the four main areas of the investigation.

5.2.1 Process Modeling

Before digging deeper into the literature to learn more about what process modeling is, let's first define the terminology to create meaning. A straightforward but understandable alternative is to describe a process as a series of actions carried out in a particular order to accomplish a particular goal [16]. Modeling implies describing, and the objective here is to describe processes through the creation of models (diagrams) that are visually appealing rather than in CSV or XES format.

The techniques that produce these visuals are made from the raw data. Directly-Follows Graphs (DFGs) and UML diagrams to BPMN (Business Process Model and Notation) and Petri nets are the available methods Dear Aalst [22] demonstrates for visualizing these process models. Making the analysis of the next subsections requires the use of process modeling. Also to understand my analysis for better, these information's are required.

5.2.2 Process Model Formalism's

Numerous methods for examining and simulating business processes are included in the topic of process mining. Directly-Follows Graphs (DFGs), Petri Nets, and BPMN are three important topics in process mining that are briefly covered in this part. These notations provide a visual representation of how many-step processes are structured. Together, these topics and examples offer insights into different aspects of process analysis and formalism in the context of process mining.

5.2.2.1 Directly-Follows Graphs

When importing an event log, the majority of process mining tools immediately display a Directly-Follows Graph (DFG) [22] for showing initial behaviour of the events. It has special start and end nodes, represented by namely; a triangle as a start node, and a square representing the end node. Unlike other model notations, DFG's represent all the possible combination that could happen in process, from the event data[22]. The author [22] mentioned a problem occurring from this instance, when two activities occur in the event, DFGs automatically creates loop even they never happened together which creates ambiguity while reading the diagram.

5.2.2.2 Petri Nets

One of the most popular process modeling notations in process mining is the Petri Net [16]. Petri nets are ideal for simulating processes because of their formal semantics and mathematical background [16]. It is possible to convert Petri Net Notations to BPMN and other types of model notations [21]. One of the other feature the author[21] stated is that Petri Nets are executable and allowing many analysis techniques to analyze processes. It relies on logical expressions that establish connections between events by utilizing the sequence derived from the timestamps associated with those events. Also the next two representation techniques are based on Petri-Nets [22]. One example visualization could be Figure 2 which represents a Petri-Net notation generated by Alpha Miner.

5.2.2.3 Others

- Process Tree: In the sense of modeling, process trees are not used widely but some algorithms require [22]. A process tree

is a tree-like structure that has only one root node. The leaf nodes correspond to activities and connected by four type of operators. Namely; exclusive choice (out of many activities, any goes to the end state), sequential composition (multiple activities follow a sequence to reach the end state), parallel composition (multiple activities happens at the same time), redo loop (creates loop through activities). One example process tree visualization can be Figure 10 generated by inductive miner.
- BPMN: Author [21] states that Business Process Model Notation is widely used and one of the most used language to process models. In the practical part, tools only support a small subset of the BPMN standard which needs conversion to Petri-Nets and similar notations [22]. Figure 9 can be an example diagram.

5.2.3 Process Discovery Algorithms

5.2.3.1 Alpha Algorithm

According to the author [20], most of the classical process discovery approaches have problem dealing with concurrency. Alpha algorithm takes concurrency as its start point. The Algorithm's goal is to create causality relations from a collection of event sequences. The algorithm consumes event logs and produces Petri Nets with particular patterns betweeen events [17, 20]. An example Petri-Net notation produced by Alpha Miner, Figure 2.

5.2.3.2 Heuristic Miner

The algorithm starts by creating a dependency graph that accounts for activity frequencies and instances where one activity is followed by another [20]. Dependencies are either included or excluded from the graph based on predetermined thresholds. This dependency graph reveals the basic connections between activities. The intricate split and join behaviors of nodes are then revealed giving a more thorough understanding of the process flow [20]. An example produced Heuristic View can be Figure 5.

5.2.3.3 Fuzzy Miner

One of the most recent process detection methods is the fuzzy miner. It simplifies the model interactively, turning spaghetti-like models into ones that are more manageable and making it suited for mining [17]. It has benefits [20] to certain mining algorithms, such as handling noise and incompleteness.

5.2.3.4 Inductive Miner

Inductive Miner's architecture resembles a series of tasks for analysis and visualization [13]. In contrast to other mining techniques, a user can alter any parameter at any time to promote exploration. The authors[13] support that, most academic tools lack features like animation and seamless zooming, making it dificult to support the repetitive nature of process exploration. One example could be Figure 1.

5.2.4 Conformance Checking

Event data collection for processes within ERP systems is not arbitrary. It can be based on predefined models. Process visualization predates [16] the advent of process mining and contributes to credibility within a company. One of the examples could be, the availability of one of the ISO (International Organization for Standardization) [12] standards allows companies to purchase verification services that assess the quality of their processes for their custwomers.

These predefined process models allows conformance checking based on the generated process models from event logs [16, 22]. These generation could be made by tools. Some example process mining tools [24] namely, ProM, Celonis, Disco, and PM4Py.

The goal is to find the positions in the diagrams which related (predefined model) and observed (event log) disagree [22]. For instance, the predefined model's footprint can demonstrate that x comes after y [20] rather than the other way around. If a relationship between y to x is observed, this will cause ambiguity to be discovered by conformance checking in the system. These spots that has undesired behaviours could be reasoned by frauds and ineficiencies in the processes [22]. The authors [16, 20, 22] mentions there are two popular methods used in order to check both diagrams conformance. Namely, token-based replay and alignments. To understand the causes of the deviations, the findings of the model comparison should be discussed with a supervised person. Who, what, when, and where are the key topics of discussion in these variations [17].

5.3 Formalizing the ERP data for Process Mining Techniques

Author [22] mentions that Information Systems are eligible to produce such data that supports representing processes. These systems typically contain event data [22], which is the essential component for process mining techniques. However, it is not done for the application of process mining techniques by only extracting databases. Correct relationships across databases are necessary to develop a format that can be used for process mining. The event logs are created in databases or information system interfaces using these formats, which should be built to the XES and OCEL specifications which will be described in the next paragraphs.

5.3.1 Event Log

Events are the building blocks of event logs, and each event, in the authors' [4, 22] opinion, must include a least of three components: event, activity, and timestamp. Each event has an associated action, and the timestamp establishes the order of the events that creates processes.

5.3.2 OCEL

The authors[4] suggests to construct the extracted data from SAP (ERP system) as Object Centered Event Logs (OCEL) in their example. Unlike traditional intermediate storage units to collect events (such as CSV), in OCEL, an event could handle multiple object associated with it. To understand the basics of OCEL, it is necessary to deconstruct each constituents. According to the OCEL definition; An Object is specific kind of item that the business is related to, the data should explain the features of these objects and creates meaningful connections between database tables [8]. An Event is associated to certain objects and has a relational identifier such as bar-code number, an action as event type, a timestamp to determine the sequences of the events [8]. Finally, Logs are made from sets of events and objects, creates a sequence that creates processes [8]. These logs can be utilized in conjunction with Process Mining methods [4]. However this format still needs to converted to a XES format for further analysis which process mining tools require.

5.3.3 XES (eXtensible markup language Encoded Source)

Previous version [25] for XES called MXLM, published in 2003, the first standardize event data storage unit. Because of the limitations, new version they published become XES. Also XES became the oficial IEEE standard for storing event data in 2016 [22].

The format they constructed contains of logs. Each log contains traces, that contain events. Additionally, event classifiers that give each event a name can be supplied in the log element. Events can now be compared to one another based on their associated identities [23, 25].

By employing these conversion techniques, it becomes feasible to analyze the data using the process mining tools described in the previous section.

5.4 ERP Systems and Process Mining

Modern ERP system modules offer a variety of data kinds that are essential for eficient business operations [3, 7, 10]. These comprise financial data, sales and customer information and contact details, inventory and supply chain data, procurement, and logistics data, and human resources information. Accoring to the author [22], information systems are eligible to produce event data, which is essential for process mining. There are numerous strategies that can be employed when process mining techniques are used in the context of ERP systems [7, 22]. Process discovery, conformance verification, and performance analysis are particular process mining approaches that are applicable to ERP systems which stated in previous chapters. Based on the literature review mentioned in the previous paragraphs, it can serve as a valuable guideline for approaching Budin Akarca's ERP data. By following the Crisp-DM methodology, this project could be effectively planned, executed, and evaluated in each state, ensuring that the results are reliable and the objectives are met. The next paragraph will ensure the step-by-step notation on how to extract valuable insights from data generated from ERP systems using Process Mining Techniques.

6 Case Study

In the forthcoming sections, a comprehensive description of the company [2] that is under investigation will be provided, including an overview of its operations and key aspects. Additionally, the nature of the available data will be examined, emphasizing its relevance and significance in relation to the project at hand. The data extraction process will employ methods and techniques acquired from previous chapters, ensuring a systematic and eficient approach. By utilizing these established techniques, valuable insights will be extracted and patterns within the data will be uncovered, thereby facilitating a deeper understanding of the company's production processes. The focus of the analysis will be on reviewing the produc -tion data for a specific product during a three-month period, namely January to March 2023. The last part of this section will present a comprehensive analysis of the data, offering valuable insights and recommendations aimed at enhancing the company's production processes.

6.1 Business Understanding

To gain a sense of the resources that are needed and accessible, the business environment should be evaluated. One of the most crucial aspects of this phase is choosing the data mining objective. According to the author [19] is important to first describe the types of process mining techniques (conformance checking, Performance Analysis, etc.) in this section. The company under study, Budin Akarca, is actively engaged in the production of printing inks and since 2007 has successfully implemented Net-Sis ERP, a widely utilized ERP solution provided by Logo company [1]. According to the CEO, the modules within the ERP system have undergone continuous development and have progressively gained enhanced functionality over time. Through extensive consultations with the accountant and ERP support personnel, as well as a thorough analysis of the database, it became evident that the ERP system is built upon various modules such as Orders, Production, Receipts, Stocks, Payments, and Human Resources, among others. These specific modules have been the focal point of the Data Preparation phase. Further examination of the company's data, combined with relevant literature, revealed the close association between production processes and Process Mining. Initially, the project aimed to develop a predictive model to address stock-related issues. However, due to the limited time-frame of this research, the focus has been narrowed down to analyzing the production line for a specific ink type, namely single color (LAK-1011). According to the accountant, the company's most produced and sold product can be identified. In order to check for bottlenecks in the system, Conformance Checking will be employed. The creation of a model and a thorough understanding of the data are crucial for this analysis. The company's adherence to ISO standards, as mentioned earlier, is of great importance in the context of conformance checking. The ISO standards have facilitated the modeling of all the available processes within the company, ranging from employee on boarding to production. This serves as a crucial aspect for conformance checking, as the comparable model will be based on these established ISO processes.

6.2 Data Understanding

Key tasks in this phase involve collecting data from various sources [19], analyzing and condensing it, and evaluating the quality of the data. In this research work, the event data were extracted from SSMS (SQL Server Management Studio) that the ERP system build onto. Hence, for this study, the organization's production cycle was considered as the primary input. The actual data in the form of an Excel spreadsheet was extracted using SQL queries, encompassing 66 events associated with 1 specific activity (production of LAK-1011) across 632 cases. The data spanned from 1 January 2023 to 23 March 2023. This event was taken from the stock control table (TBLSTHAR), which also contains sales-related events.

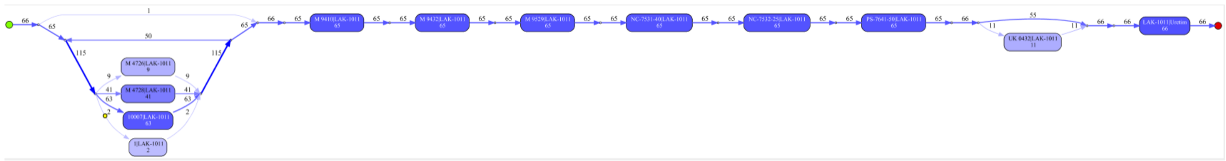

Fig. 1. Inductive view on Production Data by Inductive Miner 5.2.3.4

6.3 Data Preparation

The process of data selection involves establishing specific criteria for inclusion and exclusion [19]. In this step, the data needs to undergo a transformation to a suitable format for the process mining tool, namely Prom Lite 1.3. This tool specifically accepts data in CSV format, which requires converting it to the XES format through the process of transformation. During the transformation process, it is essential to select the relevant attributes associated with each event and case. This includes identifying the necessary information such as the event identifier (e.g., barcode number), case (e.g., type of action) and the timestamp. These selected attributes play a crucial role in generating the model from the event data.

Within the company's dataset, each column represents a specific attribute containing valuable information. The chosen attributes include: the stock name of the product associated with the event, a unique identification number serving as a code to link instances, and a description that indicates either the name of the product produced (utilizing the stock name attribute) or the name of the event itself, such as the company name that purchased the product or "URETIM" denoting production. Additionally, the timestamp attribute plays a crucial role in establishing the sequence of events, which is essential for analyzing processes effectively.

After being done with the identification of attributes, to address issues associated with inadequate data quality, it is essential to undertake data cleaning procedures[19].In the specific case of the company, it was crucial to differentiate and filter the data because the dataset contained irrelevant event logs related to sales of the color LAK-1011. Python and its libraries (std and pandas) were employed to process and filter the extensive event logs, eliminating irrelevant entries and ensuring a more concise and focused dataset for subsequent analysis. The ProM framework facilitates the conversion of raw data format into XES format, as the descriptions discussed earlier. The identification number is chosen as the case column since it serves as the relevant identifier in the context. Respectively the chosen event attributes are the "stock name" (LAK-1011) and the "description" (URETIM). The selected event attributes are inherently interconnected as they contribute to the creation of the production process with the help of the timestamp.

6.4 Modeling

The data modeling step encompasses the development of the test case, the test model, and the modeling approach by process mining techniques. During this phase, various data mining methods can be utilized. In this case, test model is provided by the company and edited for the clarification of the language (Turkish). The future analysis will be based on the test model and the recipe of the color represented by Figure 12 and 11.

6.4.1 Control Flow and Conformance Analysis

As indicated in the appendix section, a range of widely used process mining tools are available for visualizing event logs to facilitate a clear understanding. It is evident that each method employs a distinct approach to analyze the event data. By examining Figure 1 and 2, it becomes apparent that the inductive miner method provides a clearer understanding of the sequence and control flow of events. This is attributed to its ability to generate occurrence counts, thereby facilitating the detection of anomalies.

To enhance the understandability of visualizations, filtering techniques are applied to the attributes used , particularly in the case of the Alpha Miner. By referring to Figure 6 and 2, it becomes more apparent how the choices are made in the production of the color LAK-1011 because of filtering on attributes. However, it is worth noting that the occurrences are not clearly visible in these graphs, which raises the question of whether the flow choices could be influenced by data quality issues. It is possible that other process mining tools could provide a more comprehensive view of the occurrences and shed further light on the analysis.

As depicted in the recipe in Figure 11, the production line for creating the color LAK-1011 follows a linear process, where semi-products are added one by one. However, when examining the graphs provided in the Appendix, it is evident that tree structures occur, leading to anomalies in terms of the utilization of different semi-products to create the same color. These discussions regarding flow choices were conducted with individuals from the Research and Development department, as well as the Accountant, as mentioned in the acknowledgments section of this research.

Inductive mining techniques, particularly when applied to process mining, provide valuable insights into flow choices by visualizing each trace by date and showcasing their occurrences in a sequential manner. From the Figures 1 and 3, it indicates that these choices are not related to a data quality problem. This approach enables the identification and filtering of anomalies, allowing us to understand when and how they occurred, thus facilitating the reasoning process. The outcomes of these discussions and analyses will be presented in the Evaluation section of this study, shedding light on the implications and significance of the identified anomalies.

6.5 Evaluation

It appears that there are anomalies in the process flow structure based on the test models. The anomalies involve the starting points

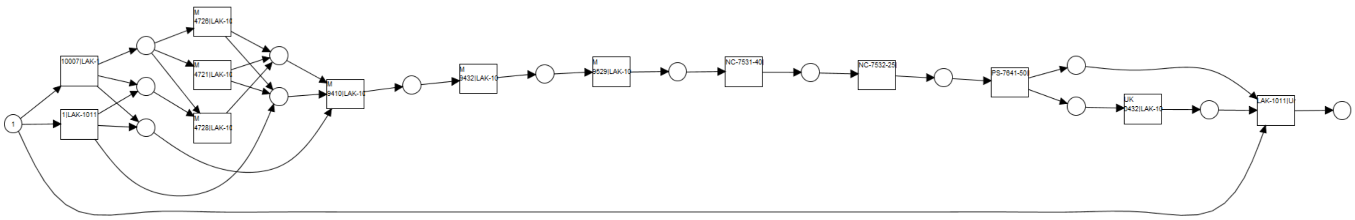

Fig. 2. Petri-Net produced by Alpha Miner on Production Data. Link to Petri-Net's 5.2.2.2

of the process, either 1/LAK-1011 or 10007/LAK-1011, and the subsequent addition of recipe instances, namely UK0423, M4726, M4721, and M4728/LAK-1011. According to the recipe, resources 10007 and M4721 should be used in the first 2 steps.

The discussion mentioned with the related people from the company, the results were not creating a problem from the companies side, which they reason the anomalies as "Sometimes while creation of a product, some semi-products could be not in the stock or not available at some moments. There are alternatives that could used for the same product resulting the same. In this case for M4721; UK0423, M4726 and M4728 can be used as the substitute for LAK-1011, and "10007" used instead of "1" which they both represent metal barrels for storing the ink.".

If we consider Figure 2, we can observe that at the bottom of the diagram, there is a direct flow to the LAK-1011/URETIM end result. Upon analyzing the data using the inductive miner and examining the occurrence of that event, it became apparent that it was the first event created from the data itself. Further investigation revealed that it was related to the inventory count, which is typically conducted after the new year.

From this evaluation, the anomalies found have significantly hindered the identification of bottlenecks, but in the start of your thesis, the objectives were to create a comprehensive understanding of the entire progress, including Order (customers giving orders via mail or phone call), Production (the process under examination), and Sales (which was excluded from this particular analysis).

During the data collection phase, it is worth noting that the company specifically designed an attribute related to the desired delivery date of the product. This attribute was considered essential as it can help identify potential bottlenecks in the process and understand the impact of delivery time on overall process performance. However this field were never filled and along with that, the order and the production of the customs, the data were filled when the order has produced, which creates ambiguity in the data to follow processes. This allowed for the exclusion of the examination of the company's sales process flow, thus enabling a focused analysis solely on the Production process for the purposes of the thesis.

7 Conclusion

All things considered, the objective of this research was to leverage ERP systems and Process Mining techniques to derive benefits. The study followed a systematic approach, starting with an examination of existing resources and formulating relevant sub-questions to establish a background for addressing the problem. The methodologies section then addressed these sub-questions and the methods needed to answer each of the question, culminating in a practical assessment using the company's data and employing Process Mining techniques. Which are namely; investigation on ERP modules and data provided from them, process mining techniques in the content of ERP system data, data formulation for Process Mining tools.

Through a systematic literature review, strong correlations between ERP systems and Process Mining were identified [20, 22], laying a solid foundation for further research in this domain. By utilizing aligned methodologies and tools provided by mentors, the research successfully uncovered process ambiguities within the company's production data, particularly in relation to a specific color. The results of the analysis were conducted using ProM 1.3, a popular process mining tool. Through the utilization of this tool, the generated graphs from the event data allowed for the detection of bottlenecks within the company's processes. These identified bottlenecks were then discussed and clarified by engaging with the appropriate individuals within the company.

Looking ahead, it would be beneficial to evaluate the ERP system at Budin Akarca based on the observations and comments highlighted in the evaluation section, possibly with the assistance of a third-party business provider specializing in process mining. Modifying the modules within the ERP systems and introducing compulsory attributes could enable the discovery of previously unknown bottlenecks, thereby further enhancing the overall eficiency of the processes.

Acknowledgments

I would like to express my heartfelt gratitude to Mr. Usman Ahmad and Ms. Faiza Bukhsh, my esteemed professors, for their invaluable guidance and support throughout my research project journey. Their dedication and expertise have played a pivotal role in shaping my knowledge and findings in this research paper.

Furthermore, I would like to extend my appreciation to Mr. Suayip and Mr. Abdullah from Budin Akarca A.S. [2]. Their assistance with the SQL queries, data collection, and explanations has been instrumental in the success of my project. Their expertise and willingness to help have been truly invaluable.

References

[1] [n. d.]. Logo Netsis 3 Entegre. Retrieved from https://www.logo.com.tr/en/product/logo-netsis-3-entegre. Accessed: June 19, 2023.
[2] 2022. Printing Inks and Masterbatch Manufacturer. https://www.budingroup.com/.
[3] Kohei Arai (Ed.). 2021. ERP Implementation in the Oil Sector of Middle-East: A Case Study in Sultanate of Oman and Saudi Arabia. Springer International Publishing, Cham.
[4] Alessandro Berti, Gyunam Park, Majid Rafiei, and Wil M. P. van der Aalst. 2021. An Event Data Extraction Approach from SAP ERP for Process Mining. CoRR abs/2110.03467 (2021). arXiv:2110.03467 https://arxiv.org/abs/2110.03467
[5] L. Bántay and J. Abonyi. 2023. Frequent Pattern Mining-based Log File Partition for Process Mining. Engineering Applications of Artificial Intelligence (2023). https://doi.org/10.1016/j.engappai.2023.103190
[6] Partha Pratim Deb, Diptendu Bhattacharya, Indranath Chatterjee, Abhijit Saha, Arunodaya Raj Mishra, and Sk Hasane Ahammad. 2022. A Decision-Making Model With Intuitionistic Fuzzy Information for Selection of Enterprise Resource Planning Systems. IEEE Transactions on Engineering Management (2022), 1–15. https://doi.org/10.1109/TEM.2022.3215608
[7] Mohamed A. A. Elmonem, Ehab S. Nasr, and Mohamed H. Geith. 2017. Benefits and Challenges of Cloud ERP Systems – A Systematic Literature Review. Future Computing and Informatics Journal 2, 3 (2017), 162–171. https://doi.org/10.1016/j.fcij.2016.11.001
[8] Anahita Farhang Ghahfarokhi, Gyunam Park, Alessandro Berti, and Wil M. P. van der Aalst. 2021. OCEL: A Standard for Object-Centric Event Logs. In New Trends in Database and Information Systems, Ladjel Bellatreche, Marlon Dumas, Panagiotis Karras, Raimundas Matulevičius, Ahmed Awad, Matthias Weidlich, Mirjana Ivanović, and Olaf Hartig (Eds.). Springer International Publishing, Cham, 169–175.
[9] M. Haddara and A. Constantini. 2018. View of Fused or Unfused? The Parable of ERP II. DOAJ. (November 8 2018). https://doi.org/10.12821/ijispm080303
[10] Moutaz Haddara and Ahmed Elragal. 2015. The Readiness of ERP Systems for the Factory of the Future. Procedia Computer Science 64 (2015), 721–728. https://doi.org/10.1016/j.procs.2015.08.598 Conference on ENTERprise Information Systems/International Conference on Project MANagement/Conference on Health and Social Care Information Systems and Technologies, CENTERIS/ProjMAN / HCist 2015 October 7-9, 2015.
[11] R. Ismail, Zalinda Othman, and Azuraliza Abu Bakar. 2009. Data mining in production planning and scheduling: A review. 2009 2nd Conference on Data Mining and Optimization, DMO 2009, 154–159. https://doi.org/10.1109/DMO.2009.5341895
[12] ISO/IEC. 2013. ISO/IEC 19510:2013 - Information technology —Object Management Group Business Process Model and Notation. https://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/06/26/62652.html
[13] Sander JJ Leemans, Dirk Fahland, and Wil MP Van Der Aalst. 2014. Process and Deviation Exploration with Inductive Visual Miner. BPM (demos) 1295, 8 (2014).
[14] Lin Li, Qing Chang, and Jun Ni. 2009. Data driven bottleneck detection of manufacturing systems. International Journal of Production Research 47, 18 (2009), 5019–5036. https://doi.org/10.1080/00207540701881860
[15] L. Li, Q. Chang, J. Ni, G. Xiao, and S. Biller. 2007. Bottleneck Detection of Manufacturing Systems Using Data-driven Method. IEEE Transactions on Automation Science and Engineering 4, 4 (2007), 428–438. https://doi.org/10.1109/TASE.2007.909501
[16] Jorge Munoz-Gama. 2016. Conformance Checking and its Challenges. Springer International Publishing, Cham, 11–18. https://doi.org/10.1007/978-3-319-49451-7_2
[17] A. B. Naveed, Z. Mahmood, M. U. Sana, I. de la T. Díez, J. C. Galán, S. Brie, and I. Ashraf. 2023. Behavioral and Performance Analysis of a Real-Time Case Study Event Log: A Process Mining Approach. Scopus (2023). https://doi.org/10.1016/j.procs.2021.02.170
[18] Julia Rudnitckaia, Hari Santhosh Venkatachalam, Roland Essmann, Tomáš Hruška, and Armando Walter Colombo. 2022. Screening Process Mining and Value Stream Techniques on Industrial Manufacturing Processes: Process Modelling and Bottleneck Analysis. IEEE Access 10 (2022), 24203–24214. https://doi.org/10.1109/ACCESS.2022.3152211
[19] C. Schröer, F. Kruse, and J. M. Gómez. 2021. A Systematic Literature Review on Applying CRISP-DM Process Model. Procedia Computer Science 181 (2021), 677–686. https://doi.org/10.1016/j.procs.2021.02.170
[20] Wil van der Aalst. 2012. Process Mining: Overview and Opportunities. ACM Trans. Manage. Inf. Syst. 3, 2, Article 7 (jul 2012), 17 pages. https://doi.org/10.1145/2229156.2229157
[21] Wil M. P. van der Aalst. 2011. Process Modeling and Analysis. Springer Berlin Heidelberg, Berlin, Heidelberg, 29–57. https://doi.org/10.1007/978-3-642-19345-3_2
[22] W. M. P. van der Aalst. 2022. Process Mining Handbook. Springer.
[23] W. M. P. van der Aalst, B. F. van Dongen, C. W. Günther, A. Rozinat, H. M. W. Verbeek, and A. J. M. M. Weijters. 2009. ProM: The Process Mining Toolkit. In Proceedings of the BPM 2009 Demonstration Track (CEUR Workshop Proceedings, Vol. 489). CEUR-WS.org, 1–4.
[24] H. M. W. Verbeek, Joos C. A. M. Buijs, Boudewijn F. van Dongen, and Wil M. P. van der Aalst. 2011. XES, XESame, and ProM 6. In Information Systems Evolution, Pnina Soffer and Erik Proper (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 60–75.
[25] Moe Thandar Wynn, Julian Lebherz, Wil M. P. van der Aalst, Rafael Accorsi, Claudio Di Ciccio, Lakmali Jayarathna, and H. M. W. Verbeek. 2022. Rethinking the Input for Process Mining: Insights from the XES Survey and Workshop. In Process Mining Workshops, Jorge Munoz-Gama and Xixi Lu (Eds.). Springer International Publishing, Cham, 3–16.
[26] HsiuJu Rebecca Yen and Chwen Sheu. 2004. Aligning ERP implementation with competitive priorities of manufacturing firms: An exploratory study. International Journal of Production Economics 92, 3 (2004), 207–220. https://doi.org/10.1016/j.ijpe.2003.08.014

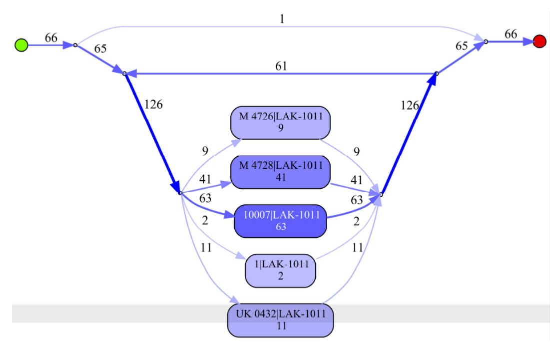

Fig. 3. Filtered model on attributes that creates ambiguity, result produced by Inductive Miner.

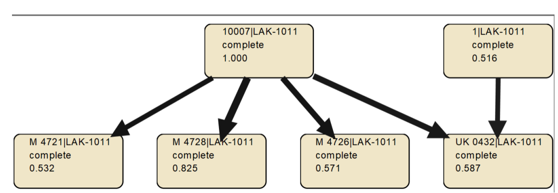

Fig. 4. Filtered model on attributes that creates ambiguity, created by fuzzy miner.

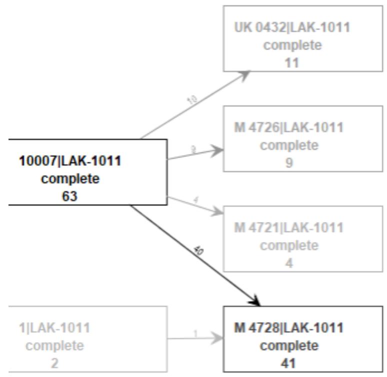

Fig. 5. Filtered Production Line, with Heuristic Miner.

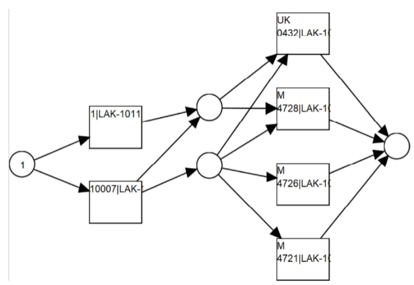

Fig. 6. Filtered Production Line, With Alpha Miner.

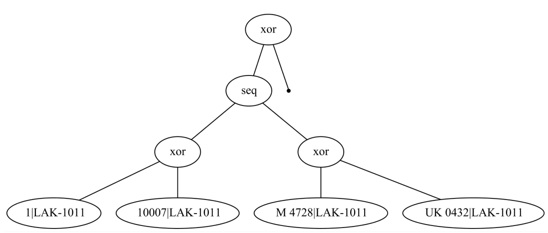

Fig. 7. Filtered Process Tree by Inductive Miner

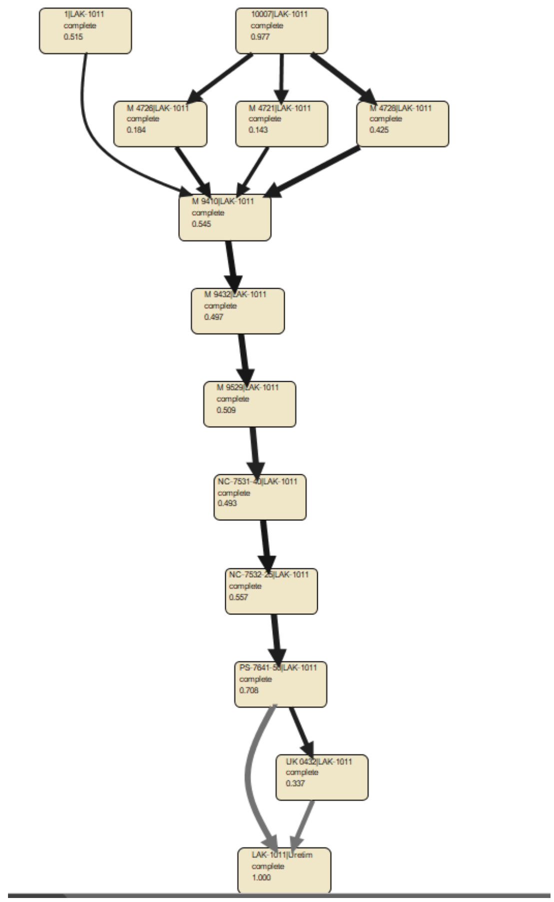

Fig. 8. Directed Followed Graph 5.2.2.1, produced by Fuzzy Miner 5.2.3.3.

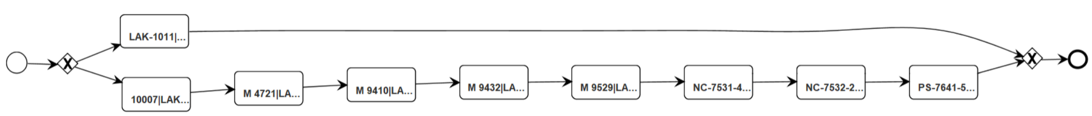

Fig. 9. Business Process Model Notation 5.2.2.3 created by Inductive Miner

Fig. 10. Process Tree 5.2.2.3 produced by Inductive Miner 5.2.3.4

Stok Kodu	Stok Ismi	Sıra No	Miktar	Fire Mik.	Br	Br.Maliyet	T.Maliyet	Stok Sevi.	Sip.Gereği
LAK-1011	**RELEASE LAK**		**1,0000**		**KG**	**0,0000**	**0,0000**	**0,0000**	**1,0000**
001	BASKILI UZUN BIDON (430MMX0,32MM)	0010	0,0500	0,0000	AD	0,00	0,00	1.078,0000	0,0000
M 4728	TEGO GLİDE A116	0007	0,0060	0,0000	KG	0,00	0,00	55,4725	0,0000
M 9410	HEKZAN-HAVA K-	0006	0,0600	0,0000	KG	0,00	0,00	15.718,1950	0,0000
M 9432	TOLUOL-RAPIT-	0004	0,0700	0,0000	KG	0,00	0,00	961,9800	0,0000
M 9529	IPA	0005	0,0540	0,0000	KG	0,00	0,00	15.779,7760	0,0000
NC-7531-40	**MUM PASTA**	**0002**	**0,0500**	**0,0000**	**KG**	**0,00**	**0,00**	**1.228,3080**	**0,0000**
M 7529	CERETAN MX 9820	0001	0,0200	0,0000	KG	0,00	0,00	1.700,0000	0,0000
M 7551	CERAFLOUR 916	0005	0,0001	0,0000	KG	0,00	0,00	11,5000	0,0000
M 7553	CERAFLOUR 1000	0006	0,0001	0,0000	KG	0,00	0,00	9,5000	0,0000
M 9529	IPA	0003	0,0249	0,0000	KG	0,00	0,00	15.779,7760	0,0000
NC-7925-50-I	**MALEİK VERNİK**	**0004**	**0,0050**	**0,0000**	**KG**	**0,00**	**0,00**	**105,5010**	**0,0000**
M 7925	ILREZ 200	0001	0,0025	0,0000	KG	0,00	0,00	5.686,0000	0,0000
M 9529	IPA	0003	0,0025	0,0000	KG	0,00	0,00	15.779,7760	0,0000
NC-7532-25	**NC-MUM PASTA**	**0003**	**0,0500**	**0,0000**	**KG**	**0,00**	**0,00**	**155,6100**	**0,0000**
M 7540	CRODAMIDE ER POWDER	0001	0,0125	0,0000	KG	0,00	0,00	950,0000	0,0000
M 9529	IPA	0003	0,0275	0,0000	KG	0,00	0,00	15.779,7760	0,0000
NC-7925-50-I	**MALEİK VERNİK**	**0002**	**0,0100**	**0,0000**	**KG**	**0,00**	**0,00**	**105,5010**	**0,0000**
M 7925	ILREZ 200	0001	0,0050	0,0000	KG	0,00	0,00	5.686,0000	0,0000
M 9529	IPA	0003	0,0050	0,0000	KG	0,00	0,00	15.779,7760	0,0000
PS-7641-50	**PAS VERNİK**	**0001**	**0,7100**	**0,0000**	**KG**	**0,00**	**0,00**	**61,2600**	**0,0000**
M 7635	CRAYAMID 941	0001	0,3550	0,0000	KG	0,00	0,00	12.750,0000	0,0000
M 9410	HEKZAN-HAVA K-	0005	0,1420	0,0000	KG	0,00	0,00	15.718,1950	0,0000
M 9517	IZOBUTONOL	0004	0,0355	0,0000	KG	0,00	0,00	871,9000	0,0000
M 9529	IPA	0003	0,1633	0,0000	KG	0,00	0,00	15.779,7760	0,0000
M 9590	ÖZEL SOLVENT	0006	0,0142	0,0000	KG	0,00	0,00	239,0900	0,0000
Toplam Maliyet	**0,00**								
Maliyet Katkısı	**0,00**								
Malzeme Maliyeti	**0,00**								
Hammadde Toplamı	**1,0501**								

Fig. 11. Recipe for the production of LAK-1011

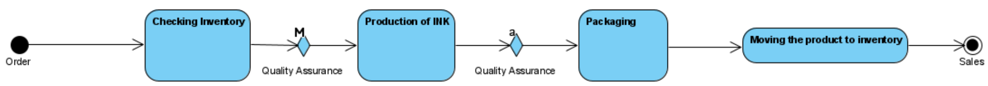

Fig. 12. Model created by the company and edited, for Production