

## Article

# Behavioral and Performance Analysis of a Real-Time Case Study Event Log: A Process Mining Approach

Naveed Anwer Butt <sup>1</sup>, Zafar Mahmood <sup>1</sup>, Muhammad Usman Sana <sup>2</sup>, Isabel de la Torre Díez <sup>3,\*</sup>,  
Juan Castanedo Galán <sup>4,5,6</sup>, Santiago Brie <sup>4,7,8</sup> and Imran Ashraf <sup>9,\*</sup>

<sup>1</sup> Department of Computer Science, Faculty of Computing and Information Technology, University of Gujrat, Gujrat 50700, Pakistan

<sup>2</sup> Department of Information Technology, University of Gujrat, Gujrat 50700, Pakistan

<sup>3</sup> Department of Signal Theory, Communications and Telematics Engineering, University of Valladolid, Paseo de Belén 15, 47011 Valladolid, Spain

<sup>4</sup> Higher Polytechnic School, Universidad Europea del Atlántico, Isabel Torres 21, 39011 Santander, Spain

<sup>5</sup> Department of Projects, Universidad Internacional Iberoamericana, Arecibo, PR 00613, USA

<sup>6</sup> Department of Projects, Universidade Internacional do Cuanza, Cuito EN 250, Bié, Angola

<sup>7</sup> Department of Project Management, Universidad Internacional Iberoamericana, Campeche 24560, Mexico

<sup>8</sup> Fundación Universitaria Internacional de Colombia, Bogotá 11001, Colombia

<sup>8</sup> Department of Information and Communication Engineering, Yeungnam University, Gyeongsan 38541, Republic of Korea

\* Correspondence: isator@tel.uva.es (I.d.l.T.D.); imranashraf@ynu.ac.kr (I.A.)

**Abstract:** Project-based organizations need to procure different commodities, and the failure/success of a project depends heavily on procurement management. Companies must refine and develop methods to simplify and optimize the procurement process in a highly competitive environment. This paper presents a methodology to help managers of project-based organizations analyze procurement processes to determine the optimal framework for simultaneously addressing multiple objectives. These goals include minimizing the time between the generation and required approval for a purchase, identifying unnamed activities, and allocating the budget efficiently. In this paper, we apply process mining algorithms to a dataset consisting of event logs on Oracle Financials-based enterprise resource planning (ERP) procurement processes in ERP systems and demonstrate interesting results leading to project procurement intelligence (PPI). The provided log data is the real-life data consisting of 180,462 events referring to seven activities within 43,101 cases. The logged procurement processes are filtered and analyzed using the open-source process mining frameworks PrOM and Disco. As a result of the process mining activities, a simulation of the discovered process model derived from the event log of the entire procurement process is presented, and the most frequent potential behaviors are identified. This analysis and extraction of frequent processes from corporate event logs help organizations understand, adapt, and redesign procurement operations and, most importantly, make them more efficient and of higher quality. This study shows that after the successful formulation of guiding principles, data refinement, and process structure optimization, the case study results are considered significant by the organization's management.

**Keywords:** process mining; event log analysis; real-life application; procurement process; Petri net; heuristic miner



**Citation:** Butt, N.A.; Mahmood, Z.; Sana, M.U.; Díez, I.d.l.T.; Galán, J.C.; Brie, S.; Ashraf, I. Behavioral and Performance Analysis of a Real-Time Case Study Event Log: A Process Mining Approach. *Appl. Sci.* **2023**, *13*, 4145. <https://doi.org/10.3390/app13074145>

Academic Editor: João M. F. Rodrigues

Received: 22 February 2023

Revised: 20 March 2023

Accepted: 21 March 2023

Published: 24 March 2023



**Copyright:** © 2023 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

Traditionally, organizations are divided into departments such as production, sales, etc. These departments have their data collection and analysis systems. However, other departments need to know what work is being done, leading to confusion. Modern information systems have emerged to overcome these problems. Most companies and organizations store their data in information systems (IS), such as executive information systems and enterprise resource planning (ERP) systems. ERP systems are growing in

popularity because they view the entire organization as one system and its departments as sub-systems [1]. These systems overcome the traditional problems as all the organization's information is stored centrally and available to every department, bringing many benefits to the organization, such as process integration, data transparency, automation to increase productivity, etc. ERP is a business process management software that helps companies integrate internal and external management information, such as finances, procurement, and customer relationship management. Many systems store all relevant events in a structured form, usually as logs, also known as audit trails or event logs [2,3].

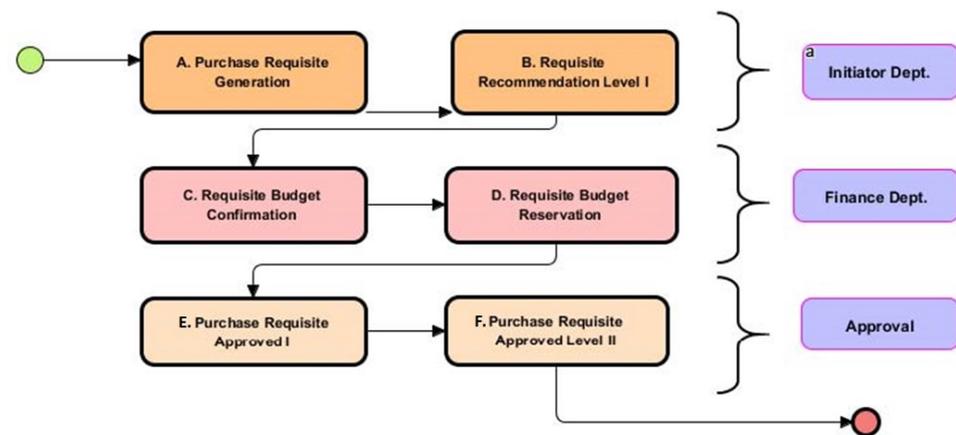
Over the past few decades, auditors have faced many problems, such as the inability to detect and prevent accounting errors because the systems or tool kits provided to auditors cannot detect all errors and fraud. Many system owners also need more information about what is happening because data mining tools are used in many areas to support business decisions, but they are bad for the process. Organizations spend a lot of money on process modeling. Manual modeling of processes means models need to be updated. Process mining makes sense at this point because it automatically creates process models from log data that can be updated at any time [4].

Process mining aims to extract knowledge from event logs using various tools, strategies, and methods to identify, monitor, and improve real-world processes. Throughput, bottlenecks, and variance are just a few examples of process performance metrics that can be used to analyze processes with process mining [5]. Process mining technology is very suitable for extracting information about existing processes from ERP systems. When a real-time process is executed in an ERP system, the generated data are used to reform the process model [6,7]. Discovery is a major application domain of process mining, which aims to discover process models by analyzing event logs and extracting knowledge from them. This is the only prior information that is available at this stage to analyze event logs through multiple algorithms, automatically generate Petri net models, and accurately capture the actual control flow of business operations.

Conformance is a specific type of analysis that verifies the accuracy of a discovered or ideal model by detecting deviations (who, what, when, and where) based on a comparison between the discovered model and the event log. It also detects the strength level of the model (how close it is to the ideal business process or model). Conformance checks reiterate the notion that logs are reflected in the process model and examine bottlenecks and time stamps associated with each event and process. Process mining also covers three different perspectives. These are used to answer questions such as 'how', 'what', and 'who'. The sequence of activities or the flow of a process is what the term 'process perspective' means. The organizational perspective is used to answer the question of 'who is running the processes and how are they related'.

The goal of this perspective is to structure the organization through the visibility of relationships between performers or between performers and tasks. The characteristics of the case are the focus of the case perspective. It helps to clarify 'what happened to this particular operation?' [8,9]. The purpose is to study and improve business processes using process mining tools and techniques at Oracle Financials based on the ERP procurement data of Pakistani organizations. Procurement is procuring goods and services, preparing and processing requirements, and final acceptance and approval for payment, as shown in Figure 1.

The primary contribution of this study is to show how process mining techniques can be used to compute the alignments between event logs and process models and highlight both low-level and high-level deviations. This study applies process mining techniques in the field of the project procurement process. Given an event log and a Petri net, these metrics yield intuitive insights into the conformance between the log and the net, even if the log is non-fitting.



**Figure 1.** Procurement life cycle.

The main objective of this research is to thoroughly analyze the provided ERP procurement log data to find an effective and efficient process model that shows the actual process flow of the organization along with frequently correlated sub-processes, optimum time frames calculated for the sub-processes for process flow, and detection of anomalies. Basically, procurement is the act of buying goods and services. This process includes the creation and processing of claims and the final receipt and approval of payments. It typically includes supply planning, standard determination, supplier research and selection, value analysis, financing, price negotiation, procurement, and other related functions. Process mining techniques and algorithms can be used to generate useful knowledge for organizations and help organizations improve business processes from ERP systems.

Various process mining tools and techniques are applied to log data to find the technique that provides the best solution. In process mining, log data needs to be filtered and loaded into process mining software (such as PrOM, DISCO, etc.) for actual mining and process model reconstruction. After this process, the model can be used for its intended purpose. PrOM is a general-purpose open-source framework, which means it supports various plugins for various process mining techniques, such as the Alpha algorithm and its extensions. Another process mining tool is Disco [10]. It can easily transform and filter data, and it can handle large event logs and complex process models. Disco is used to automatically map logs with CSV and XLS extensions to XES or MXML format (powered by PrOM) to optimize performance and control deviations without informing the algorithm. In this work, the Disco performance view is used to find time delays in the procurement process.

The PrOM plugin Conformance Checker is utilized for the initial validation stage, making it easier to compare the event log with the model. This comparison shows whether the algorithm's output, including aggregation and decomposition processes, is the desired outcome. Before this comparison, the models are manually transformed into Petri nets. As this is a process conformance-based case study for the procurement process, we have not compared it with any other study. Different process mining algorithms and techniques, i.e., fuzzy miner,  $\alpha$ -algorithm, heuristic miner, genetic algorithm [11], and colored Petri nets (CPN) have been applied to log data aimed at the discovery of processes from event logs used to find the optimal solution.

Three qualitative and competence metrics, discovery, consistency, and enhancement, are introduced to evaluate the utility of the models. These metrics provide analysts with fast and reliable feedback on how representative the current model is relative to observed actual behavior. In discovery, process models are created from event logs without using a priori models, which means no additional information is used. In discovery, the alpha algorithm is used for model building. The generated model is called the initial process model, which can be crafted by hand. The conformance checking has an a priori model. In this phase, conformance-checking techniques are used to compare the observed event log

with the initial process model to detect and locate discrepancies between reality and the model. During enhancement, the process model and event log are kept consistent, and a given process model can be improved/refactored with some additional perspectives.

The following are the significant contributions to research that are made by this research work.

- What methodology should be followed to apply process mining to ERP systems?
- Which procedures should be followed for extracting and processing event logs from ERP systems?
- What process mining methods are best for identifying process models from log data?

The rest of this article is divided into five sections. Section 2 provides an overview of the important research works. The methodology used in this study is described in Section 3, while discovery analysis is presented in Section 4. The study is concluded in Section 5.

## 2. Related Work

Over the past few decades, the majority of attention has been dedicated to developing new techniques and algorithms, primarily focusing on the discovery of flow-controlled prospecting [12,13]. One of the primary causes of the maturation of process mining techniques is the simple availability of event data. Very few studies have a practical focus. This study employs past techniques instead of proposing a specific approach to investigate the inefficiencies of service providers' processes. These findings are corroborated by a case study of procurement services, which utilized multiple process mining techniques. However, only a few examples of real-world applications in the literature demonstrate the effectiveness of process mining. The case study of invoice processing services is examined by [8,14–16] using process mining techniques. The authors of [8] utilized the process mining technique of a heuristic miner for verification and network analysis.

The most significant benefit of this research is how different process mining techniques incorporate different perspectives on event logs. This case study is similar, but the analysis method differs from the different perspectives. Through the application of process mining techniques, they re-create processes and discover deviations and a lot of other pertinent information. The case study in [14] has a variety of diversity in the human processes supported by systems. They must rely on data from actual reconstructions to figure out what happened. The issues in the auditing process domain have been addressed [2,9,15,17–23]. The authors discussed the importance of ACs, which are necessary to produce thorough audit results but are currently neglected because they are not readily available in process models.

The proposed method is limited to data stored electronically in ERP systems. The appropriate depiction of audit-important data in the context of process audits is, thus, a worthy research topic. The investigation and audit of business processes that lead to financial entries is a significant obstacle in the auditing process, which is discussed in [15]. The study [18] investigates the effect of process mining on the internal audit process. In this study, the volume of event data from internal auditors provides an unprecedented opportunity to assess the value of process mining. This facilitates the identification of a baseline against which to discover the information used by standard audit trails. It has been discussed in [17] how process mining and the reconstruction of mined processes can be utilized to bridge the gap between automated transaction processing and other audit methods. More research is necessary to find solutions to the issues of selecting instances for processing, automating the aggregation, visualizing the results, and the difficulty of creating algorithms.

The authors of [18] investigate the procedure for procuring services for major European providers to determine if process mining can augment internal audits. They discovered that the techniques of process mining reveal failures in internal control that the auditors failed to recognize. The lack of generalizability is a deficiency of this case study approach. However, it only affects the particular outcome rather than the overall conclusion. In this context, data mining techniques have been employed by [19,20,24–26]. The authors proposed a

more advanced method of process management called the 'procedure tree' (PT) for RFID data mining in [19]. They can effectively manage the massive data associated with RFID and efficiently utilize the suggested PT during the process of real-time management.

The study [7] proposed a method that attempts to give software engineers an automatic process to construct mined models from systematic event logs that describe requirements; this process includes addressing technological difficulties and problem-solving. The authors proposed that the system utilize the ActiTrac algorithm to cluster generated models; this would lead to a more refined description of the models, which would decrease the likelihood of error and reduce the need for additional analysis during the creation phase of models. The authors proposed a new approach in [24] that avoids the over-generalization of business processes in ERP by employing process mining and cluster analysis. They employed the Euclidean distance and K-means in their study of event log data. For process discovery, they employed the heuristic algorithm, and for the verification of model conformance, a Petri net model was employed. The experimental results demonstrate that the trace clustering approach can be employed to avoid the over-generalization of ERP processes and generate accurate and specific models. As a result, the process mining model is concise and straightforward. The study [20] attempted to implement the CRISP-DM methodology to increase the transparency of the techniques associated with process mining in the ERP context. Additionally, the healthcare sector has also experienced a significant boost from the mining process. Similarly, refs. [16,27] utilized process mining to investigate complex care pathways. A methodological approach to the utilization of process mining in this scenario is derived from the outcomes of studies conducted by a significant number of patients to track deviations from recommendations. The methodology focuses on the sequence clustering of applications to discover different utilization scenarios. In [28], the authors proposed a systematic and automated method for identifying tasks and extracting data for process mining in enterprises to relieve manual labor and improve data quality.

The study [29] summarizes the research on process mining to analyze the warehouse management process. The actual business process model is produced by the heuristic miner algorithm, which is employed to process event log data. The analysis of the results demonstrates the deviation from the company's established process. One significant component of legacy modernization that maintains system maintenance is the proposal of an incremental process mining algorithm used to mine the structures of processes evolutionarily in legacy systems [30]. A scheme has been proposed by [31] that incorporates predictive analytics and big data analytics into a new framework. The proposed framework accomplishes their strategic objectives regarding operational decisions, allowing organizations to create horizontal processes. The lack of necessity of accounting for information such as AC controls to derive a comprehensive audit result is because this information is not present in process models. To address this deficiency, a method for automatically augmenting process model enrichment with audit-relevant information about ACs is presented [32]. Successfully navigating issues and difficulties associated with analyzing event logs is crucial to any process mining process [25]. Several categories of data quality issues documented in event logs have been identified. Such concerns limit the usefulness of specific process mining approaches and diminish the value of knowledge that is gained. According to the authors, the findings will facilitate systematic logging procedures, repair methods, and analytical methods.

According to [33], no research has been conducted on the 'preliminary variables' of success in process mining or how to 'quantify the effectiveness of process mining activities'. The investigation comprised three successful outcomes, five successful criteria, and a validated, pre-determined model of process mining success. There are several negative aspects to this approach, the primary one being that the a priori model is primarily derived from theory and literature that is similar to it and has inherent limitations that are addressed by a thorough validation of the model. The study [21] serves as the basis for a methodology that process mining can employ to analyze complex event logs. Additionally, the literature demonstrates the subtlety and versatility of approaches to process mining in the financial

services industry. The case study illustrates several limitations of the process mining methods. Since the mining process is based on actual data, this is its primary attribute. The primary benefit of the mining process is that it is based on actual data, but this also has a negative aspect. Second, the process mining techniques are a clandestine struggle involving vast amounts of data that reflect the behavior of the unstructured process. It states that the practice of using a relevant filter that is utilized can be a means of extracting crucial information from the event logs; this indicates the necessity of additional research in this area to advance and modify PM methods.

Ref. [34] employed process mining techniques, collaboration analysis, and frequent sub-graph mining in real-world cases to identify relevant behavioral trends. The objective of this investigation was to identify frequent sub-processes that are not anticipated as a novelty in the process model. The proposed method comprises two primary steps. First, an instant graph is generated for each trace, followed by hierarchical cluster analysis. The proposed method is unsuccessful in finding parallel behavior. To reduce the burden of manual labor and improve data quality, research is currently being conducted in the field of process mining to address these issues by presenting a systematic and automated process for identifying jobs and extracting data for enterprise PM. Additionally, a system that attempts to automate software engineers with the technique of constructing mined models from systematic event log requirements contains solutions to problems intended to benefit people and technical difficulties. Some limitations also exist in the current endeavor, for example, including limitations on data in some cases, etc. The suggested approaches can be improved in several ways, including further research to validate and test the idea and studies to see if the suggested strategy is portable. Finally, yet importantly, the literature analysis shows that there are still many practical applications for process mining that have not yet been studied, and there is a need for more research on real-world case studies that show the effectiveness of process mining.

### 3. Methodological Framework for Applying Process Mining in Practice

This work aims to conduct a case study to illustrate the practical advantages of process mining and offer recommendations for its practical implementation. Since synthesis frequently found a process behavior much more unstructured, there have been protocols illustrating how many of the given algorithms have trouble handling actual events logs [35–37].

Furthermore, because several control flow-mining algorithms are available, several preliminary visualizations of the process can be acquired for this function. Based on these stages of exploration, business experts can improve the iterative framework of the process and the time frames that ensure input data for further analysis. The feedback loop, the range for the adjustment, and the term process are essential and, therefore, explicitly shown in Figure 2.

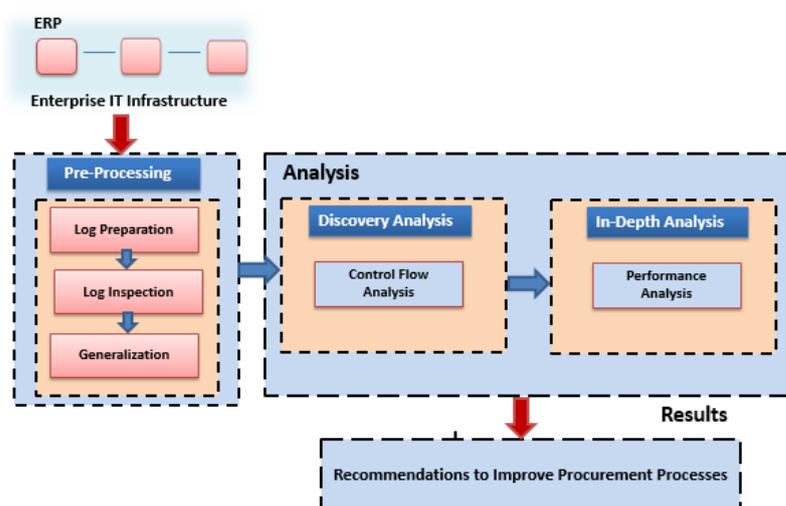


Figure 2. The process mining methodology framework.

Several event logs can be established for examination in exceptional cases, such as the case study provided in the next section, because a single event log often immediately includes three aspects. However, it is beneficial to create several event logs to study the various viewpoints, especially in non-process situations. One can begin the fundamental analysis once one has determined the different analysis dimensions and decided to create several event logs from the execution data. The fundamental discovery analysis and analysis of the comprehensive compliance and performance analysis are the two primary divisions of the analysis phase.

This study distinguishes between the organization's control flow and the potential for case data during the discovery phase. Investigating the activity flow patterns inside the business process is part of the control flow perspective. Additionally, data from an organizational perspective might be evaluated, for instance, by process teams. To uncover specific patterns while looking for patterns in process executions, it can be helpful to investigate the underlying data element. Usually, a discovery scan highlights various points of interest suitable for further assessment. In the case study, management is usually interested in the throughput (downturn) times and a performance analysis run for the entire process executions. In addition, it is usually worth exploring the performance more closely. The effects of control flow or other aspects of the executions of the synthesis process are then studied using subgroups of traces.

Finally, the result phase is the closing phase of the process mining framework. The analysis's findings serve as a valuable platform for efforts to optimize the business, as mentioned earlier, such as process modifications or even process re-engineering. Management can define new objectives based on new findings gained from the mining process to solve such an identified inefficient process measurement.

### 3.1. Case Study

To demonstrate the utility of process mining analysis in practice, a case study is described in the field of the education sector. This case study addresses a company's request for an analysis of the procurement business process in the Oracle finance system to pinpoint the circumstances under which the process is ineffective and offer a recommendation for process improvement. The analyzed company runs thirty-four ERP systems, i.e., payroll systems, management information systems, online request systems, central registry systems, store and inventory systems, overtime systems, etc. Due to the large number of human-centric business activities for which event log analysis is precious, this sector is important to the mining process.

### 3.2. Data Sources and Collection

In our research work, the log data were extracted from an ERP system of an organization in the form of raw historical data. The business process selected for analysis was the procurement process. Therefore, the organization's procurement cycle was the input of this work. The provided log data were the real-life data in the form of an Excel spreadsheet consisting of 180,462 events referring to 7 activities within 43,101 cases with 'DATEEND' between 14 May 2004 and 16 September 2013. Figure 3 shows the three main characteristics of this log data: case ID showing multiple linked events, an activity that occurs during the event, and a timestamp of the sequence of events in a case.



**Table 1.** Global statistics of log data.

SR#	Attributes	Statistics
1	Total no. of events	180,462
2	Total no. of cases	43,101
3	Total no. of event classes	9
4	Maximum events per case	35
5	Average events per case	4
6	Minimum events per case	1
7	No. of start events	4
8	No. of end events	7

**Table 2.** Activities with their corresponding event IDs.

SR#	Activity	Event ID
1	Purchase requisite generation	1
2	Purchase requisite generation	2
3	Requisite recommendation I	3
4	Requisite budget confirmation	4
5	Requisite budget reservation	4
6	Purchase requisite approved I	5
7	Purchased requisite approved II	5
8	Unnamed activities	4, 6
9	Mark for procurement	7

An activity forms one step in the process, and the names of these activities represent the level of detail for the process steps. There may be many steps in a process, and some may occur more than once in a case, but it is not necessary for them to happen every time. As mentioned above, 7 activities are recorded in the event log, each of which takes place during an event. Table 3 shows the activities of the process, their occurrences, and their relative occurrences.

**Table 3.** Activities statistics.

SR#	Activity	Frequency	Event ID
1	Requisite recommendation I	48,631	26.95%
2	Purchase requisite generation	47,846	26.51%
3	Requisite budget confirmation	40,932	22.68%
4	Unnamed activities	33,136	18.36%
5	Purchase requisite approved I	9590	5.31%
6	Requisite budget reservation	313	0.17%
7	Purchased requisite approved II	13	0.01%
8	Mark for procurement	1	0%

### Evaluation of Event Log

In this section, the four quality problems of the event log identified in our log data are presented as they manifest in an event log, and how these problems and their effects on the application of process mining can be addressed.

#### Missing Attribute Values

Many essential attributes can be absent from an event log, or specific characteristics may have no value. Such attributes can either belong to a trace (e.g., the identifier of the case, etc.) or an event (e.g., the name of the task to which the event refers or the time stamp of the event). The process mining methods can be affected by event logs with missing features or values. For example, control-flow discovery techniques are affected by such missing task information or time stamps. To deal with these issues, a solution is to remove the affected events/traces from the event log [25]. In this case, many unnamed activities are found while analyzing log data. These activities have case IDs and time stamps, but the

activity name needs to be added, confusing what actual activity is performed. Out of the total of 43,101 cases, 27,383 cases containing 115,690 events have this quality issue. Among these events, 33,136 events have missing activity names. Thus, to avoid this confusion, these unnamed activities were removed. Table 4 represents the log data statistics after removing unnamed activities.

**Table 4.** Statistics of log data after removing missing attributes.

SR#	Attribute	Statistics
1	Total no. of events	147,326
2	Total no. of cases	43,101
3	Total no. of event classes	7
4	Maximum events per case	35
5	Average events per case	3
6	Minimum events per case	1
7	No. of start events	3
8	No. of end events	6
9	Total no. of variants	332

### Incomplete Traces

In this issue, prefix and/or suffix events corresponding to a trace in the event log are missing, although they occur in reality. Due to these incomplete traces, there may be problems with the results produced by the process mining algorithms, as different relations may infer the start or end of the process. There are some algorithms to deal with this kind of noise, e.g., fuzzy miner. Another solution is to filter the log data to remove incomplete traces [22]. This study uses the endpoints filter during the analysis to remove incomplete traces. This filter selects the cases based on their start and end activities. This activity-based filter filters incomplete cases or trims the cases to cut out the parts of the process. Out of 147,326 events corresponding to 43,101 cases, there were 10,320 events corresponding to 33,926 cases with incomplete traces. After applying the endpoints filter to the data to remove the incomplete traces, the obtained filtered log consisted of 44,106 events corresponding to 9175 cases. As this filter is activity-based, not based on the time stamp, we applied another filter named ‘filter log using simple heuristic’ on the filtered log file. This filter combines many configurable log filters. The event-type filter, which allows choosing the kind of events or tasks we want to take into account while mining the log, is the initial log filter.

The ‘start event filter’ filters the log in such a way that only traces or cases that start with the selected tasks are kept. A frequency threshold of 80% was applied to select the most frequent start events. They cover 80% of the traces here. The third filter applied in our simple heuristic filter was the ‘end events filter’, which filters the log so that only the traces or cases that end with the indicated tasks are kept. The frequency threshold was set to 80 to select the most frequent traces%. The fourth filter was the ‘event filter’, which filters all unselected events from the log. Now, upon inspecting the resultant log, there were fewer cases, and all the cases started with the activity ‘purchase requisite generation’ and ended with the tasks ‘purchase requisite approved I’ and ‘purchase requisite approved II’. After applying this filter on 44,106 events corresponding to 9175 cases, the resultant log consisted of 38,746 events corresponding to 8276 cases. Table 5 presents the statistics of the log after the removal of incomplete cases.

**Table 5.** Statistics of log data after removing incomplete traces.

SR#	Attribute	Statistics
1	Total no. of events	38,746
2	Total no. of cases	8276
3	Total no. of event classes	6
4	Maximum events per case	35
5	Average events per case	5
6	Minimum events per case	4
7	No. of start events	1
8	No. of end events	2
9	Total no. of variants	204

### Repetition of Activities

In log data, there can be events with the same activity name and event IDs within the same case. This can affect the results of process mining algorithms either by producing inaccurate results or by producing complex results. For instance, duplicate tasks in process discovery are represented by a single node, leading to a large fan-in or fan-out. This issue is resolved by considering these repeating events as one event [25]. During the analysis of the data, which consist of 38,746 events corresponding to 8276 cases, many events occurred repeatedly. This issue was resolved by considering those repeated activities as one activity with the same activity names with identical event IDs. After removing repetition, the resultant log consisted of 28,373 events corresponding to 7023 cases. This resultant log contained 5 activities, which are represented in Table 6.

**Table 6.** Activity statistics after removing repetition of an activity.

SR#	Activity	Frequency	Relative Frequency
1	Purchase requisite generation		25.41%
2	Requisite budget confirmation	7026	24.76%
3	Requisite recommendation I	7023	24.75%
4	Purchase requisite approved I	7012	24.71%
5	Requisite budget reservation	95	0.33%
6	Purchase requisite approved II	11	0.04%

Table 7 represents the events IDs corresponding to activities mentioned in Table 6. Here, their ID, their frequency, and relative frequency are given.

**Table 7.** Statistics after removing repetition of an activity.

SR#	Event ID	Frequency	Relative Frequency
1	4	7121	25.1%
2	2	7023	24.75%
3	3	7023	24.75%
4	5	7023	24.75%
5	1	186	0.66%

### 3.4. Generalization of Data

During the data analysis, two other situations of repetition were found where events could not be removed. Instead, they needed to be generalized. The first case was the activities with different event IDs but the same activity names; this issue was resolved by keeping the one with a high occurrence. In these data, event IDs 1 and 2 have the same activity name 'purchase requisite generation'. Table 7 shows that 2 has a higher occurrence than 1, i.e., 25.1%. Thus, the activity with ID 2 was kept. The second issue was related to those cases with the same event ID but a different activity name, and its solution was to keep the one with a high occurrence. For example, ID 5 has two activity names, 'purchase

requisite approved I' and 'purchase requisite approved II'. Table 6 shows that 'purchase requisite approved I' covers 24.75% of the data, and 'purchase requisite approved II' covers 0.04% of the data. The first approval was kept because of its high occurrence. There was another similar case in the data where ID 4 has two different activities, 'requisite budget confirmation' and 'requisite budget reservation'.

In comparison with Event ID 5, both the activities have different connotations, so we did not generalize; instead, we kept both these activities. A summary of highlighted issues in real-life log data is given in Table 8. This analysis aims to provide insight into how the identified quality problems exist in the actual data used for process mining. After removing noise and outliers from the event log, it was now in the form where process mining techniques and algorithms could be applied. Table 9 shows the global statistics of the noise-free log. As mentioned above, there were 4 activities in our final log data to gain insight into these activities.

**Table 8.** Evaluation of event log issues in the procurement process.

SR#	Issues	No. of Cases	No. of Events
1	Missing attribute values	27,383	33,136
2	Exposing less frequent traces	1	1
3	Incomplete traces	33,926	103,220
4	Repetition activities	1988	13,038
5	Generalization	197	978

**Table 9.** Statistics of log data after removal of noise.

SR#	Attributes	Statistics
1	Total no. of events	28,187
2	Total no. of cases	7023
3	Total no. of event classes	5
4	Maximum events per case	5
5	Average events per case	4
6	Minimum events per case	4
7	No. of start events	1
8	No. of end events	1
9	Total no. of variants	2

Table 10 presents these activities with their event IDs, frequency, and relative frequency. The following table shows two activities across event ID 4 because they have different meanings and cannot be removed or ignored.

**Table 10.** Activity statistics of log data after removal of noise.

SR#	Event ID	Activity	Frequency	Relative Frequency
1	4	Requisite budget confirmation	7026	24.92%
2	2	Purchase requisite generation	7023	24.91%
3	3	Requisite recommendation I	7023	24.91%
4	5	Purchase requisite approved I	7023	24.91%
5	4	Requisite budget reservation	95	0.34%

After the detailed analysis of log data and after removing noise from it, in the next section, different process mining algorithms are applied to visualize the organization's procurement process flow.

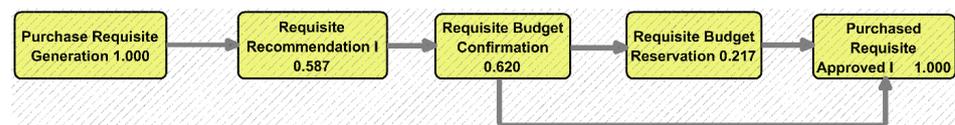
#### 4. Discovery Analysis

As mentioned in the methodological framework, the detected event log studied is an exploratory analysis to find interesting observations for further analysis. According

to the exploratory analysis, this organization is actually in contact with different activities. The next step is to discover, from the control-flow perspective, the actual processes recorded based on events. Typically, this analysis begins with the visualization of the underlying process.

#### 4.1. Control Flow Analysis

In control flow analysis, Petri nets are generated that model the concurrency and synchronization in the organization. All the statistics and pre-processing results obtained by applying a series of cleaning and inspecting methods are visualized by developing these Petri nets. They are a visual communication aid to model the system's behavior [22]. Many discovery algorithms aim to model the underlying processes from logs, and three process mining algorithms are applied, the alpha algorithm, fuzzy model, and heuristic miner, to discover the control flow of the procurement monitoring process. Discovering a control-flow perspective model only involves case IDs and their respective activities and marks the most frequent behavior underlying the log. The goal of the control flow perspective is to characterize all possible paths in terms of Petri nets. Figure 4 shows Petri nets generated by applying the three discovery algorithms mentioned earlier on refined event logs.



**Figure 4.** Visualization (dependency graph) for the control flow log using different algorithms.

#### 4.2. In-Depth Analysis: Tracking Process Inefficiencies

Did the thorough analysis and study go further and deeper into the data with intriguing ideas? Since the focus on performance is in terms of throughput time, it was decided to create an event log reference to assess the behavior of unwanted processes better.

##### Performance Analysis

After the process discovery, the resultant process models can analyze the performance. The performance analysis phase answers questions such as 'Are there any bottlenecks in the process' and 'What is the effect of pre-processing on the optimization of the process?'. It can be used to give insights into the deviations that occur on the other level than control flow, such as delays in the process. Process mining provides a wide range of performance techniques [2]. PrOM dotted chart analysis, PrOM sequence analysis, and Disco's performance view can provide valuable insights into the deviations.

In this work, the Disco performance view is used to find the time delays in the procurement process. Figure 5 shows the performance map of the data that show the mean execution time between activities. The mean time depicts the average time of execution for each activity. It can be seen in the map that there is more time consumption between the activities 'requisite budget confirmation' and 'purchase requisite approved I', causing the delay in the process. After the pre-processing of procurement log data, there remained 2 variants in our log data. These two sequences are shown by mined models of the log data, as mentioned in Table 11. A variant is a specific sequence of activities, and multiple cases may follow the same sequence through the process. The two variants in our event log are:

- Variant 1: In this variant, there are 6925 cases, and in each case, there are 4 activities involved. This variant covers 98.6% of the log, and the sequence of activities in this variant is purchase requisite generation → requisite recommendation I → requisite budget confirmation → purchase requisite approved I (2 → 3 → 4 → 5).
- Variant 2: In this variant, there are 98 cases, and each consists of 5 activities. This variant covers 1.4% of the log data and the sequence of activities according to their event IDs; this variant contains purchase requisite generation → requisite recommendation I

→ requisite budget confirmation → requisite budget reservation → purchase requisite approved I (2 → 3 → 4 → 4 → 5). Table 11 represents these 2 most frequent sequences of our log.

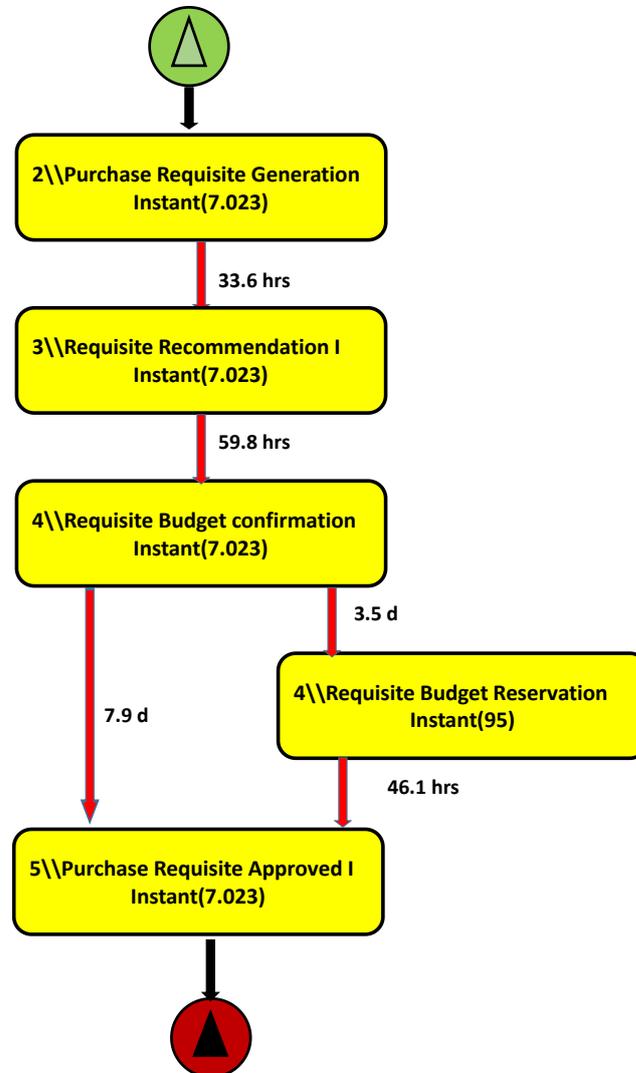


Figure 5. Performance view of mean time duration between activities.

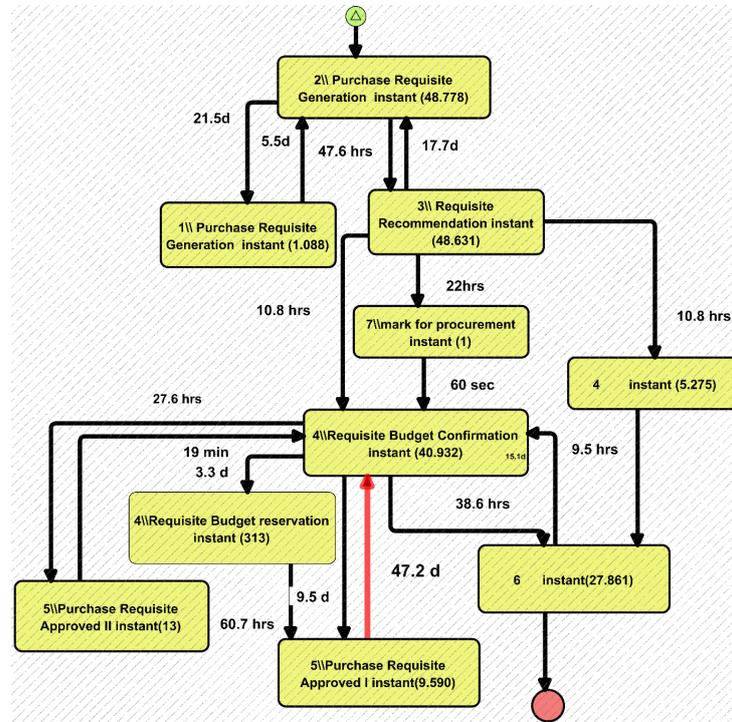
Table 11. Two variants of log.

Sequence	Occurrence	% Occurrence	Mean	Max.	Min.	Std. Dev.
2→3→4→5	6925	98.6	7023.75	7026	7023	47
2→3→4→4→5	98	1.4	7047.5	7121	7023	49

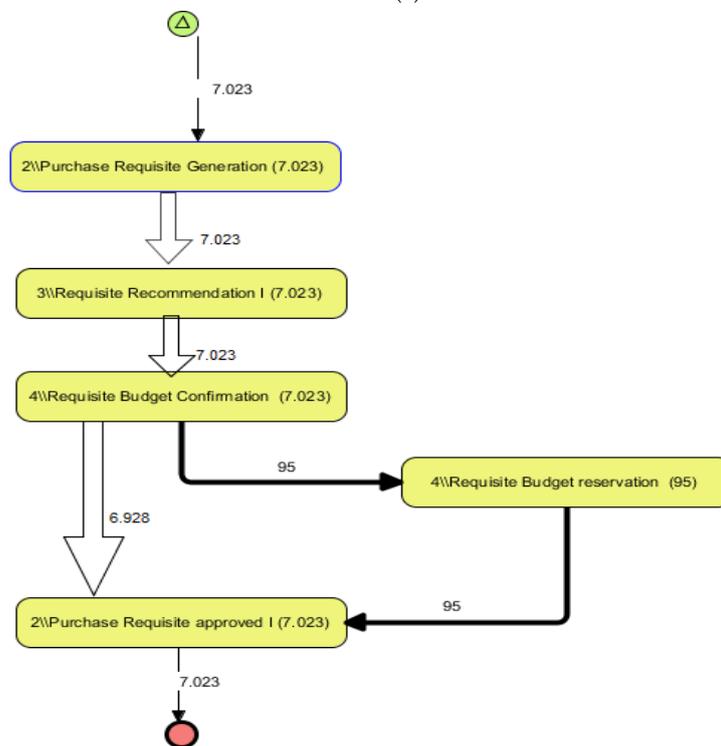
### Alpha Miner

The alpha algorithm [39] aims at reconstructing causality from a set of sequences of events. It constructs Petri nets with unique properties (workflow nets) from event logs, and each transition of a Petri net corresponds to observed tasks. Figure 6a illustrates how complex the model is to understand the actual flow of the process. It represents the total of 789 variants or sequences in which many sequences have the repetition of activities in a single sequence (as in the sequence 2 → 1 → 2 → 3 → 2 → 3 → 4 → 5 → 4 → 5). Furthermore, there are many incomplete traces (such as 2 → 3 → 4 → 6), less frequent traces (e.g., the sequence 2 → 3 → 4 → 7 occurs only one time in the whole data), some unnamed activities (such as activity 6 and some activities with ID 4 that have no activity

name), and generalizability issues related to activities (e.g., the presence of two activities across the same event ID 5, and the fact that the activity ‘purchase requisite generation’ has two IDs, 1 and 2). Three discovery algorithms were also applied to these data to generate mined models.



(a)



(b)

Figure 6. (a) Process map of original noisy data, and (b) process map of data after preprocessing.

Figure 6b represents the process map after the pre-processing of data. As discussed above, after pre-processing, there are only 2 variants in the log data, shown in the following process map. It shows that the procurement process can occur in two most frequent manners, which are: 2 → 3 → 4 → 5 and 2 → 3 → 4 → 4 → 5. In the map, the arrows show the dependencies and frequency of the performed activities, and the thicknesses of the arrows represent the frequency of occurrence; the more the thickness, the more common its occurrence. We can see in the process map that variant 1 (2 → 3 → 4 → 4 → 5) has a high occurrence, which means this flow is primarily followed in the procurement process.

Figure 7a illustrates that the generated Petri net did not reflect the correct flow because of the limitations of the alpha algorithm. In addition to the general issue of log completeness, it cannot produce the correct model. It produces very complex models, and the frequencies are not considered in this algorithm; therefore, it is susceptible to noise and can easily misclassify a relation. As our data were extensive, it did not give reliable results.

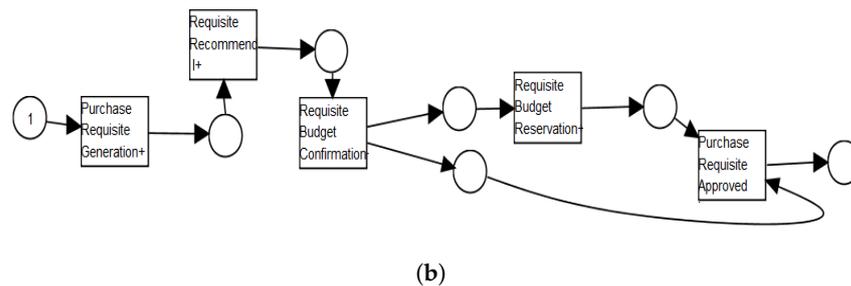
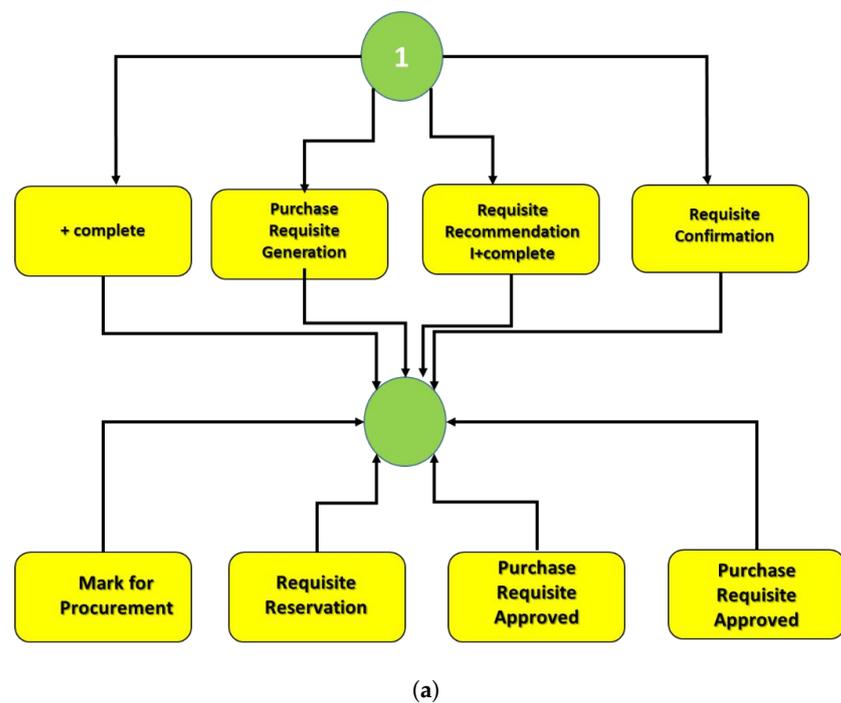


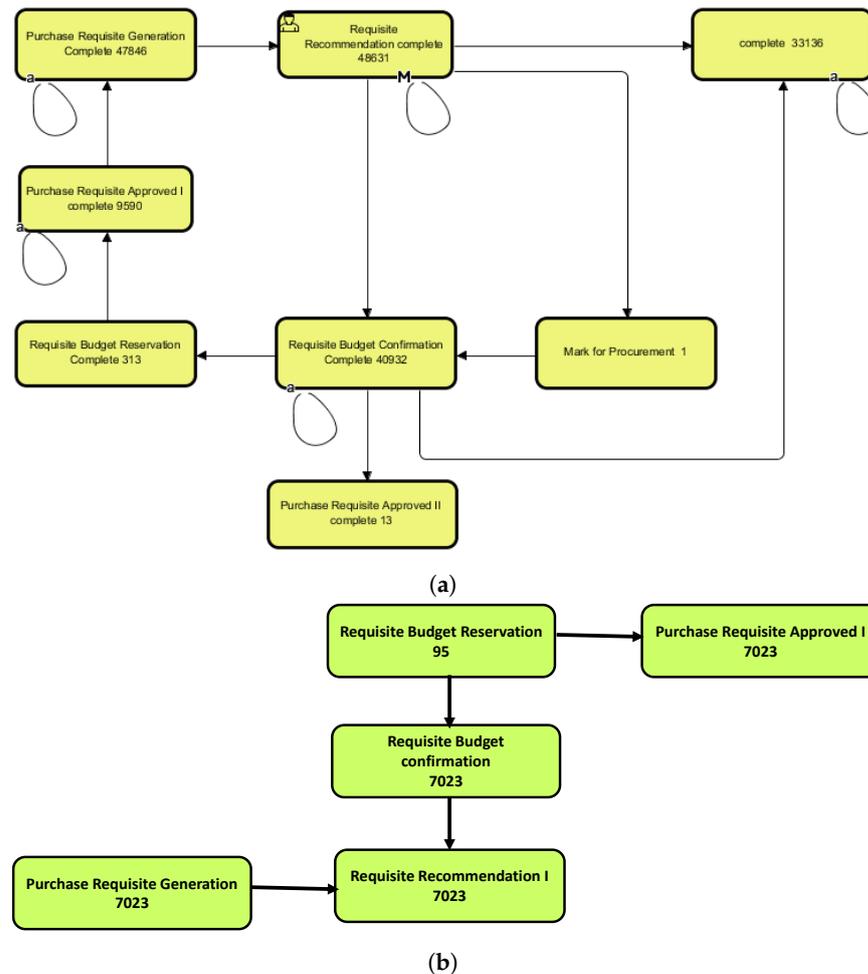
Figure 7. (a) Petri net generated by alpha miner on the original log data and (b) Petri net generated by alpha miner after preprocessing.

### Heuristic Miner

The heuristic miner algorithm [10] should be applied to real-world data that contain a limited number of distinct events. It can handle noise and convey the primary behavior, which excludes all details and exceptions and is recorded in an event log. Heuristic miner generates a heuristic net that can be converted into other process models, such as a Petri net, for further analysis.

To avoid the constraints and solve the problem of the alpha algorithm, the heuristic miner algorithm was applied to the log data as it is more sophisticated and adequate

than the alpha algorithm. Using this algorithm, we wanted to generate a model that would be less sensitive to the incompleteness of the log data and the log containing the noise. Frequencies are considered in the heuristic miner algorithm compared to the alpha algorithm. Figure 8 shows the heuristic net model created by applying this algorithm, and frequencies are also shown in this model. Although the resultant model represents a more sophisticated view of the process flow than the alpha algorithm, the produced model cannot correctly deal with mixed and complex data. Moreover, due to missing connections or activities, the results produced by heuristic mining give less meaningful information about the process.



**Figure 8.** (a) Petri net generated by fuzzy miner on the original noisy data, and (b) Petri net generated by heuristic net on refined data.

### Fuzzy Miner

The third discovery algorithm that was applied to overcome the limitations of the heuristic miner was the fuzzy miner [40]. The fuzzy miner is one of the younger process discovery algorithms. It is suitable for mining less structured processes with many activities and highly unstructured and conflicting behavior and interactively simplifies the model; i.e., it shapes spaghetti-like models into more concise ones. This algorithm is more sophisticated than the heuristic miner because it can deal with more complex structures that need to be more easily comprehensible at first glance. Figure 9 shows the fuzzy model of the log data. In this generated fuzzy model, the arrows' thickness represents the absolute frequency of occurrences. It shows all the activities as well as their casual dependencies. However, it can be seen in this model that some exceptional behaviors and loops show the repetition of activities in the data, which means there is a need to pre-process the data.

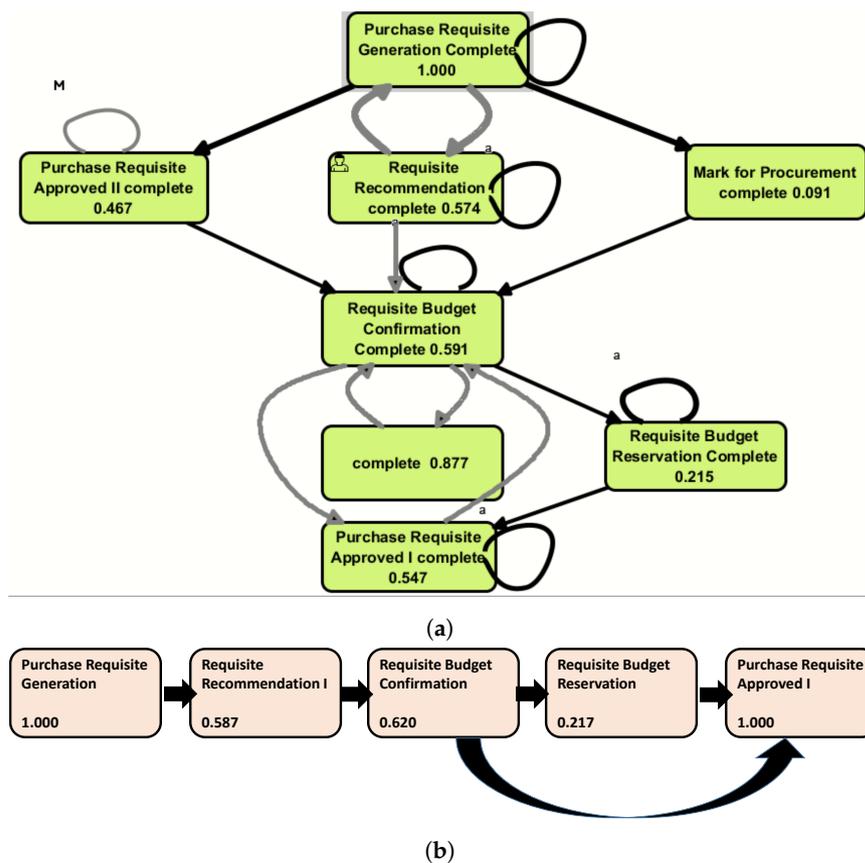


Figure 9. (a) Mined model generalized by fuzzy miner on original data and (b) Petri net generated by fuzzy miner on refined data.

These Petri nets now fully explained the process flow in optimized form. After the control flow analysis, the performance analysis was carried out to find the issues related to the time stamp in order for us to analyze the things that still had an impact on the process flow. Figure 9 shows the performance map of the procurement process in Disco, and from it, we recognize the activity path that consumes the maximum time and affects the performance. In the Disco tool, several options related to time are shown in Table 12. In this table, the first column shows the activity paths across which time is measured, and in this column, ‘C’ represents the activity ‘requisite budget confirmation’, and ‘R’ represents the ‘requisite budget reservation’. Here, the total duration shows the highly impacted areas for delays in our process by giving the cumulative times (time taken by adding up the overall cases) for each path between activities. Alternatively, the mean duration gives the average time spent between activities. The maximum duration measures the most considerable execution time and delays in flow, and the minimum duration gives the minimum execution time taken between activities. These measures show that the maximum time consumption and delay are in the path ‘requisite budget confirmation purchase requisite approved I’.

Table 12. Activity paths showing their time measures from a performance view.

Activity Paths	Total Duration	Mean Duration	Min. Duration	Max. Duration
2 → 3	26.9 years	33.6 h	Instant	12.5 m
3 → 4(C)	48 years 59.8 h	Instant	41.4 wks	
4 → 4(C → R)	47.4 weeks	3.5 d	60 s	41 days
4(C) → 5	150.1 years	7.9 d	Instant	16.5 m
4(R) → 5	26.1 weeks	46.1 h	60 s	14.1 d
4(R) → 5	26.1 weeks	46.1 h	60 s	14.1 d

#### 4.3. Results: Process Improvement Measures

Finally, organizational management evaluates the case study findings as significant by comparing actual behavior recorded in the event log data with expectations and requirements. Beyond these guiding principles of different approaches and processes, other improvements can also be observed.

First, the discovery process mining results are based on the quality of the input data. For example, using verb-object names is useful for activity description and data interpretation. Additionally, start and end timestamps of activities should be tracked to improve the performance analysis. The results of this study suggest improvements in data quality. Secondly, after data refinement, the process structure is optimized. The marked timing deviations help the purchase departments understand problems in their processes and optimize them. It is suggested to solve these problems and further improve procurement efficiency. Finally, multiple inefficiencies have been explored, providing an excellent opportunity for the administrative staff to draw attention to them and improve the business processes through better training and counseling. For example, the disadvantage of frequent retransmission could be highlighted to reduce process inefficiency.

### 5. Conclusions and Future Work

This study performs process mining in real-world case studies and acknowledges its usefulness. First, a discovery analysis reveals the most beneficial opportunities from the management's perspective. This promotes a more focused and valuable framework for additional analysis. Second, different perspectives take advantage of the ability to look for deficiencies instead of simply innovative ways to discover process model squint. Based on the results of our analysis, process mining can serve as a starting point for the management involved in the formulation of concrete measures to perform process mining on a real-life event log. Additionally, the case study exposes that the highlighted deviations in time help the procurement department understand the problems in the procedure and improve it. It is recommended to address these issues to enhance the procurement process's effectiveness. Because of the need for more data quality in event logs, the applications that utilize process mining in real life are frequently complex and have their limits. The case study demonstrates the importance of data preparation for data analysis in process mining. Specifically, researchers in the field of process mining need to address these concerns.

This study focused on the issues regarding the quality of the log data associated with procurement in an ERP system. Five types of issues with quality have been identified in the logs through the use of different process mining tools. Additionally, this investigation demonstrates that utilizing appropriate filters to extract knowledge from such event logs can be very helpful. Further research is required to understand the flaws and obstacles in the organization's process. Additionally, the process can be evaluated by applying a conformance check to find discrepancies between the intended and actual process. Moreover, the reasons for the deviations and the identified time-related issues can be found by applying data mining techniques, and these issues can be removed by applying simulation techniques. The comparison of the results between raw and preprocessed data reveals the need for data preprocessing. In addition, the performance analysis of the refined process model helps identify the time-related issues and delays affecting the optimization.

**Author Contributions:** Conceptualization, N.A.B. and Z.M.; Data curation, Z.M. and M.U.S.; Formal analysis, N.A.B. and M.U.S.; Funding acquisition, I.d.l.T.D.; Investigation, J.C.G. and S.B.; Methodology, M.U.S.; Project administration, I.d.l.T.D. and J.C.G.; Resources, I.d.l.T.D.; Software, S.B.; Supervision, I.A.; Validation, S.B. and I.A.; Visualization, J.C.G.; Writing—original draft, N.A.B. and Z.M.; Writing—review & editing, I.A. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the European University of Atlantic.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Data used in this study may be obtained from the authors upon reasonable request.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Reference

1. Berdik, D.; Otoum, S.; Schmidt, N.; Porter, D.; Jararweh, Y. A survey on blockchain for information systems management and security. *Inf. Process. Manag.* **2021**, *58*, 102397. [[CrossRef](#)]
2. Bozkaya, M.; Gabriels, J.; van der Werf, J.M. Process diagnostics: A method based on process mining. In Proceedings of the 2009 International Conference on Information, Process, and Knowledge Management, Cancun, Mexico, 1–7 February 2009; pp. 22–27.
3. van der Aalst, W.M. Process mining: A 360 degree overview. In *Process Mining Handbook*; Springer: Cham, Switzerland, 2022; pp. 3–34.
4. van Cruchten, R.; Weigand, H. Towards Event Log Management for Process Mining-Vision and Research Challenges. In Proceedings of the Research Challenges in Information Science: 16th International Conference, RCIS 2022, Barcelona, Spain, 17–20 May 2022; Springer: Cham, Switzerland, 2022; pp. 197–213.
5. Van Der Aalst, W. Service mining: Using process mining to discover, check, and improve service behavior. *IEEE Trans. Serv. Comput.* **2012**, *6*, 525–535. [[CrossRef](#)]
6. Reinkemeyer, L. Status and future of process mining: From process discovery to process execution. In *Process Mining Handbook*; Springer: Cham, Switzerland, 2022; pp. 405–415.
7. Sonawane, S.B.; Patki, R.P. Process mining by using event logs. *Int. J. Comput. Appl.* **2015**, *975*, 8887.
8. Van Der Aalst, W.M.; Reijers, H.A.; Weijters, A.J.; van Dongen, B.F.; De Medeiros, A.A.; Song, M.; Verbeek, H. Business process mining: An industrial application. *Inf. Syst.* **2007**, *32*, 713–732.
9. Jans, M.J.; Alles, M.; Vasarhelyi, M.A. *Process Mining of Event Logs in Auditing: Opportunities and Challenges*; SSRN: Amsterdam, The Netherlands, 2010; SSRN 1578912.
10. Weijters, A.; van Der Aalst, W.M.; De Medeiros, A.A. *Process Mining with the Heuristics Miner-Algorithm*; Tech. Rep. WP; Technische Universiteit: Eindhoven, The Netherlands, 2006; Volume, 166, pp. 1–34.
11. Van der Aalst, W.M.; De Medeiros, A.A.; Weijters, A.J. Genetic process mining. In Proceedings of the Applications and Theory of Petri Nets 2005: 26th International Conference, ICATPN 2005, Miami, FL, USA, 20–25 June 2005; Springer: Berlin/Heidelberg, Germany, 2005; pp. 48–69.
12. Ebrahim, M.; Golpayegani, S.A.H. Anomaly detection in business processes logs using social network analysis. *J. Comput. Virol. Hacking Tech.* **2022**, *18*, 127–139. [[CrossRef](#)]
13. Rodríguez-Quintero, J.F.; Sánchez-Díaz, A.; Iriarte-Navarro, L.; Maté, A.; Marco-Such, M.; Trujillo, J. Fraud Audit Based on Visual Analysis: A Process Mining Approach. *Appl. Sci.* **2021**, *11*, 4751. [[CrossRef](#)]
14. Štolfa, J.; Kopka, M.; Štolfa, S.; Koběřský, O.; Snášel, V. An application of process mining to invoice verification process in sap. In Proceedings of the Innovations in Bio-Inspired Computing and Applications: Proceedings of the 4th International Conference on Innovations in Bio-Inspired Computing and Applications, IBICA, Ostrava, Czech Republic, 22–24 August 2013; Springer: Cham, Switzerland, 2014; pp. 61–74.
15. Werner, M.; Gehrke, N.; Nuttgens, M. Towards automated analysis of business processes for financial audits. *AISel* **2013**, *24*, 375–389.
16. Rebuge, Á.; Ferreira, D.R. Business process analysis in healthcare environments: A methodology based on process mining. *Inf. Syst.* **2012**, *37*, 99–116. [[CrossRef](#)]
17. Werner, M.; Gehrke, N.; Nuttgens, M. Business process mining and reconstruction for financial audits. In Proceedings of the 2012 45th Hawaii International Conference on System Sciences, Maui, HI, USA, 4–7 January 2012; pp. 5350–5359.
18. Jans, M.; Alles, M.; Vasarhelyi, M. Process mining of event logs in internal auditing: A case study. In Proceedings of the 2nd International Symposium on Accounting Information Systems, Italy, 2011.
19. Kwon, K.; Kang, D.; Yoon, Y.; Sohn, J.S.; Chung, I.J. A real time process management system using RFID data mining. *Comput. Ind.* **2014**, *65*, 721–732. [[CrossRef](#)]
20. Sastry, S.H.; Babu, P.; Prasada, M. Implementation of CRISP methodology for ERP systems. *arXiv* **2013**, arXiv:1312.2065.
21. De Weerd, J.; Schupp, A.; Vanderloock, A.; Baesens, B. Process Mining for the multi-faceted analysis of business processes—A case study in a financial services organization. *Comput. Ind.* **2013**, *64*, 57–67. [[CrossRef](#)]
22. Wil Van Der Aalst, M.; Stahl, C. *Modeling Business Processes: A Petri Net-Oriented Approach*; MIT Press: Cambridge, MA, USA, 2011.
23. Jans, M.; Van Der Werf, J.M.; Lybaert, N.; Vanhoof, K. A business process mining application for internal transaction fraud mitigation. *Expert Syst. Appl.* **2011**, *38*, 13351–13359. [[CrossRef](#)]
24. Sarno, R. Avoiding Over-generalization ERP Business Processes Using Process Mining Trace Clustering. *Int. J. Adv. Comput. Technol.* **2014**, *6*, 176.
25. Bose, R.J.C.; Mans, R.S.; Van Der Aalst, W.M. Wanna improve process mining results? In Proceedings of the 2013 IEEE Symposium on Computational Intelligence and Data Mining (CIDM), Singapore, 16–19 April 2013; pp. 127–134.
26. Kolkas, M.K.; El-Bakry, H.M.; Saleh, A.A. Integrated data mining techniques in enterprise resource planning (ERP) systems. *Int. J. Inf. Sci. Intell. Syst.* **2014**, *2*, 131–152.

27. Mans, R.S.; Schonenberg, M.; Song, M.; van der Aalst, W.M.; Bakker, P.J. Application of process mining in healthcare—A case study in a dutch hospital. In Proceedings of the Biomedical Engineering Systems and Technologies: International Joint Conference, BIOSTEC 2008 Funchal, Madeira, Portugal, 28–31 January 2008; Springer: Berlin/Heidelberg, Germany, 2009; pp. 425–438.
28. Li, J.; Wang, H.J.; Bai, X. An intelligent approach to data extraction and task identification for process mining. *Inf. Syst. Front.* **2015**, *17*, 1195–1208. [[CrossRef](#)]
29. Rojas, E.; Munoz-Gama, J.; Sepúlveda, M.; Capurro, D. Process mining in healthcare: A literature review. *J. Biomed. Inform.* **2016**, *61*, 224–236. [[CrossRef](#)] [[PubMed](#)]
30. Kalsing, A.C.; Iochpe, C.; Thom, L.H.; do Nascimento, G.S. Re-learning of Business Process Models from Legacy System Using Incremental Process Mining. In Proceedings of the Enterprise Information Systems: 15th International Conference, ICEIS 2013, Angers, France, 4–7 July 2013; Springer: Cham, Switzerland, 2014; pp. 314–330.
31. Babu, M.P.; Sastry, S.H. Big data and predictive analytics in ERP systems for automating decision making process. In Proceedings of the 2014 IEEE 5th International Conference on Software Engineering and Service Science, Beijing, China, 27–29 June 2014; pp. 259–262.
32. Schultz, M. Enriching process models for business process compliance checking in ERP environments. In Proceedings of the Design Science at the Intersection of Physical and Virtual Design: 8th International Conference, DESRIST 2013, Helsinki, Finland, 11–12 June 2013; Springer: Berlin/Heidelberg, Germany, 2013; pp. 120–135.
33. Mans, R.R.; Reijers, H.; Berends, H.; Bandara, W.; Prince, R. Business process mining success. In Proceedings of the 21st European Conference on Information Systems, Utrecht, The Netherlands, 5–8 June 2013; Association for Information Systems (AIS): Atlanta, Georgia, 2013; pp. 1–13.
34. Diamantini, C.; Genga, L.; Potena, D. Behavioral process mining for unstructured processes. *J. Intell. Inf. Syst.* **2016**, *47*, 5–32. [[CrossRef](#)]
35. Goedertier, S.; De Weerd, J.; Martens, D.; Vanthienen, J.; Baesens, B. Process discovery in event logs: An application in the telecom industry. *Appl. Soft Comput.* **2011**, *11*, 1697–1710. [[CrossRef](#)]
36. Günther, C.W. *Process Mining in Flexible Environments*; Information Systems IE&IS: Eindhoven, The Netherlands, 2009.
37. Veiga, G.M.; Ferreira, D.R. Understanding spaghetti models with sequence clustering for ProM. In Proceedings of the Business Process Management Workshops: BPM 2009 International Workshops, Ulm, Germany, 7 September 2009; Springer: Berlin/Heidelberg, Germany, 2010; pp. 92–103.
38. Stoop, J. Process Mining and Fraud Detection. Master's Thesis, University of Twente, Enschede, The Netherlands, 2012.
39. 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](#)]
40. Günther, C.W.; Van Der Aalst, W.M. Fuzzy mining—Adaptive process simplification based on multi-perspective metrics. In Proceedings of the Business Process Management: 5th International Conference, BPM 2007, Brisbane, Australia, 24–28 September 2007; Springer: Berlin/Heidelberg, Germany, 2007; pp. 328–343.

**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.