

Research Article

Optimizing Decision Making on Business Processes Using a Combination of Process Mining, Job Shop, and Multivariate Resource Clustering

Hanung Nindito Prasetyo ^{1,2}, **Riyanarto Sarno** ¹, **Dedy Rahman Wijaya**,²
Raden Budiraharjo,^{1,3} **Indra Waspada**,^{1,4} **Kelly Rossa Sungkono**,¹
and **Abdullah Faqih Septiyanto**¹

¹Department of Informatics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

²Department of Information System Diploma, Telkom University, Bandung, Indonesia

³Department of Information Systems, Institut Teknologi Nasional, Bandung, Indonesia

⁴Department of Informatics, Universitas Diponegoro, Semarang, Indonesia

Correspondence should be addressed to Riyanarto Sarno; riyanarto@if.its.ac.id

Received 10 November 2022; Revised 18 January 2023; Accepted 19 January 2023; Published 27 February 2023

Academic Editor: V. E. Sathishkumar

Copyright © 2023 Hanung Nindito Prasetyo et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The current business environment has no room for inefficiency as it can cause companies to lose out to their competitors, to lose customer trust, and to experience cost overruns. Business processes within the company continue to grow and cause them to run more complex. The large scale and complexity of business processes pose a challenge in improving the quality of process model because the effectiveness of time and the efficiency of existing resources are the biggest challenges. In the context of optimizing business processes with a process mining approach, most current process models are optimized with a trace clustering approach to explore the model and to perform analysis on the resulting process model. Meanwhile, in the event log data, not only the activities but also the other resources, such as records of employee or staff working time, process service time, and processing costs, are recorded. This article proposes a mechanism alternative to optimize business processes by exploring the resources that occur in the process. The mechanism is carried out in three stages. The first stage is optimizing the job shop scheduling method from the generated event log. Scheduling the time becomes a problem in the job shop. Utilizing the right time can increase the effectiveness of performance in order to reduce costs. Scheduling can be defined as the allocation of multiple jobs in a series of machines, in which each machine only does one job at a time. In general, scheduling becomes a problem when sequencing the operations and allocating them into specific time slots without prolonging the technical and capacity constraints. The second stage is generating the resource value that is recorded in the event log from the results of analysis of the previous stage, namely, job shop scheduling. The resource values are multivariate and then clustered to determine homogeneous clusters. The last stage is optimizing the nonlinear multipolynomials in the homogeneous cluster formed by using the Hessian solution. The results obtained are analyzed to get recommendations on business processes that are appropriate for the company's needs. The impact of long waiting times will increase service costs, but by improving workload, costs can be reduced. The process model and the value of service costs resulting from the mechanism in the research can be a reference for process owners in evaluating and improving ongoing processes.

1. Introduction

The current business environment has no room for inefficiency as it can cause companies to lose out to their competitors, to lose customer trust, and to experience cost

overruns. Therefore, companies must now focus on continuous monitoring and supervision and continuously make adjustments to their business operations to ensure that the company is working at an optimal level [1]. By monitoring the business processes carried out, the company will gain

insights, be able to identify bottlenecks, anticipate problems, record policy violations, and recommend preventive actions, and be able to streamline processes to make them more efficient [2]. Real-time data processing retrieval is very important. How to understand data processing? Process mining becomes an important mechanism in improving the efficiency of company business processes. In this context, process mining is an enabler that can uncover the root causes of process inefficiency by reconstructing and visualizing current business processes as they are with their various variations [3]. In principle, process-mining techniques use event log data to find the process model and check the suitability of the process model with predetermined procedures and the ability to improve the model with a variety of information about bottlenecks, decisions, and resource use [4]. In business processes, the resource use will certainly affect the related processes whether they can run smoothly or not [5]. The initial mechanism of the process mining is the process discovery. The patterns of the process model obtained are explored to uncover situations that lead to errors and inefficiencies or exacerbate the risks of ongoing processes. This mechanism is often used to perform an analysis related to the resulting process model. The optimization of this model mostly focuses on activity trace that is recorded in the event log and focuses on optimization of the trace to obtain an optimal process model.

In the last 1.5 decades, the concept of process modelling has undergone significant development, and process mining has become a concept and its implementation is an effort to evaluate business processes. In several studies, it is stated that the largest part of real-life processes is complex and ad hoc and produces unstructured events [6]. Various efforts have been made to carry out the process mining for these conditions. The existing workflow discovery techniques had difficulty in dealing with event logs that were generated from highly flexible environments because the models that were mined from those logs often suffered from problems of high complexity. This condition then gave rise to various new techniques to obtain the expected process model. One technique that is still widely used is the trace clustering [7]. By using this technique, the existing event log is broken down into homogeneous subsets based on the activity trace, and then, a process model is created for each subset. Several studies use the same basic concepts, such as studies in [8–22], in which the optimization method used is trace clustering with a procedural approach. In general, the thing to do is to use a trace profile and then use a reference matrix profile to do clustering based on the size of the distance. Then, there is also a variation in the use of trace clustering with a model approach as done in [23–28]. These studies present an approach that aims to overcome the difficulties experienced by extracting only useful data and presenting it in an understandable way. The mechanism is divided into two stages, namely, preprocessing and activity sequence clustering. In general, the steps taken are the active trace-clustering approach, generating a cluster trace based on a machine learning approach with a number of iterations until a cluster is obtained and then measuring the clustering results.

There are other studies that divide traces by using the time-based clustering approach [29]. The clustering approach consists of two steps: The first step is extracting the distance of activity pair for each log-trace. The second step is partitioning the logs given into subsequent trace clusters that have a similar structure. There is also a similar study but with a different time approach that is conducted in [30]. In their research, it is stated that when business processes are based on time, they are not always stable, in which certain times can affect the process activities. The trace clustering is based on the identification of similar times. This means that the clusters are formed both in structural similarity and in temporal proximity. There is another interesting approach in terms of modelling optimization with a trace clustering approach but by improving the quality of the resulting clusters as done in research studies by [31–40].

There is also an approach that carries out a two-stage clustering mechanism, in which, in the first stage, the clustering is based on the employability indicators, and in the second stage, there is a grouping of the clusters obtained by using the AXOR algorithm that is based on a trace profile [41]. There is also a mechanism that combines trace clustering and text mining [42]. The use of text mining aims to improve the process discovery technique to extract more useful insights from data processing. The latest approach to track clustering is a study conducted in [43] by detecting the business areas that are most correlated with the clusters found.

Based on the explanation above, it can be seen that most of the trace clustering is based on vectors, then focuses on the model, and then on the context. Thus, in general, the clustering in previous studies exhibits clustering that is centered on flexible, complex, and specific process behaviors based on strict sequences of events. Basically, the trace clustering mechanism carried out is to obtain optimization of the resulting process model. Several studies that have started to use alternatives to the trace clustering approach are research studies by [44, 45] and [46], in which, in these studies, the process mining is carried out to use multiple perspectives. In their research studies, it is said that a multiperspective view of the process is needed, by looking at other things beyond the control perspective that determine the activity sequence of a process. The multiperspective view consists of control flow perspective, resources perspective, data perspective, time perspective, and function perspective. These five perspectives are often considered in the literature on business process management, process model, and process mining [44]. In the event log data, not only the activities but also the other resources, such as records of employee or staff working time, process service time, processing costs, and workload, are recorded. These indicators become important parts in measuring the processes that occur within the company [44]. In measuring process efficiency, it is not enough to just model the process through process discovery, but other indicators are needed that allow the process evaluation results to be more reliable [47].

Therefore, this study proposes an alternative in optimizing the process based on the resources recorded in the event log based on process mining, especially process

discovery. The resources recorded in the event log are resources items that define the patterns and relate to the processes. This is in line with the concept stating that several interacting process perspectives, especially activities, data, resources, time, and other necessary indicators, can be considered further together. These things are not owned by the process model resulted from the trace clustering. The business process optimization is carried out in three stages. The first stage is optimizing the job shop scheduling method from the generated event log. The second stage is generating the resource value that is recorded in the event log from the results of analysis of the previous stage, namely, job shop scheduling. The resulting resource value, which is in the form of multivariate, is then clustered to determine a homogeneous cluster by using K-means clustering. The final stage is to perform nonlinear optimization on homogeneous clusters formed using the Hessian approach. The Hessian approach is used as a nonlinear multivariate solution approach to obtain a minimum service cost value. The use of three stages is based on the fact that appropriate steps are needed to optimize resources. The job shop model is used to analyze time based on scheduling that occurs. Utilization of the right time can improve performance effectiveness in order to reduce costs. Then, the use of multivariate resources clustering aims to explore the resources recorded in the event log.

The remainder of this article is structured as follows: Section 2 explains materials and methods including experimental setup, dataset, and our proposed method. Section 3 presents the results and discussion of the analysis obtained from the three approaches used. Section 4 evaluates the results and discussion. Finally, Section 5 contains the discussion of the analysis, and Section 6 contains conclusions and opportunities for this research in the future.

2. Materials and Methods

2.1. Process Mining and Process Discovery. Process mining is a relatively new science in the last decade. Basically, process mining is a technique that helps companies analyze and monitor internal processes to identify their activities and resources. In the past, companies would manage this process through a series of interviews that produced an overall ideal summary of the business. Process mining allows companies to automatically monitor and track business processes in real-time without human intervention [48]. The process mining system uses various forms of technology, including data mining, deep process analysis, and artificial intelligence (AI)-based business intelligence, to provide actionable insights into the enterprise processes [49]. Another important aspect of process mining is that it enables business processes to identify process risks and bottlenecks, to determine potential improvements that can improve efficiency, and to monitor data to predict future events [50]. The process mining mechanism starts by capturing event logs that are obtained from the daily system run by the company. Then, in the event log, a process-mining mechanism is carried out, which consists of three stages, namely, process discovery, conformance checking, and enhancement [48]. The three

stages are carried out according to the needs of the analysis, and they do not have to be sequential but can be random.

Process discovery is a data-based automated mechanism for finding, mapping, and documenting existing business process activities. Then, an analysis of the data that are obtained automatically is carried out so that it can be recommended automatically in the modelling process or in the workflow [51]. The discovery and analysis of a company's business processes can be used to identify key problem areas, not only at the start of a digital transformation initiative but also when improving the performance of existing processes. Along with business process modelling, process discovery is an important part in improving the quality of business process management [52].

2.2. Production Process Analysis and Scheduling.

Production is one of the most important parts of a company related to the transformation of various inputs into outputs in accordance with the predetermined quality standards. Meanwhile, the production process is a series of steps used to convert inputs into outputs. Basically, the production process flow is based on the concept of scheduling. Scheduling can be divided into three types, namely, job shop, flow shop, and project [53]. Apart from that, there is also the problem of scheduling open shop, mixed shop, and general shop [54]. Scheduling is the process of managing resources to complete tasks that involves work, resources, and time. According to [55], scheduling is the allocation of resources to carry out a series of tasks based on time. Flow shop scheduling or so-called flow shop is a production system that produces its products with similar process flow or sequence. The overall process flow of the product is constant. Resource management follows the process flow of the product. Flow shop has two types, namely, continuous flow shop and intermittent flow shop [56]. In the continuous flow shop, materials continuously move (in and/or out) in one process without waiting for the process to finish. The example of continuous flow shop can be found in the chemical industry. Meanwhile, in the flow shop, new material moves from one process, when the process is finished, the material moves to the next process.

In job shop scheduling, the operations performed by each job are often on a different route. In this case, [57] identified several objectives of scheduling activities as follows:

- (1) Increasing resource usage or reducing waiting time
- (2) Reducing the number of jobs waiting in the queue
- (3) Reducing some work delays that have a deadline for completion in order to minimize late fees (cost of delay)
- (4) Assisting in making decisions regarding enterprise capacity planning and the type of capacity required

In addition to the above discussed concepts, to complete the scheduling analysis, a dispatching rule approach is needed. Dispatching rule is a method used for both static and dynamic job shop scheduling cases [58]. Priority rules provide guidance for the sequence of work to be carried out

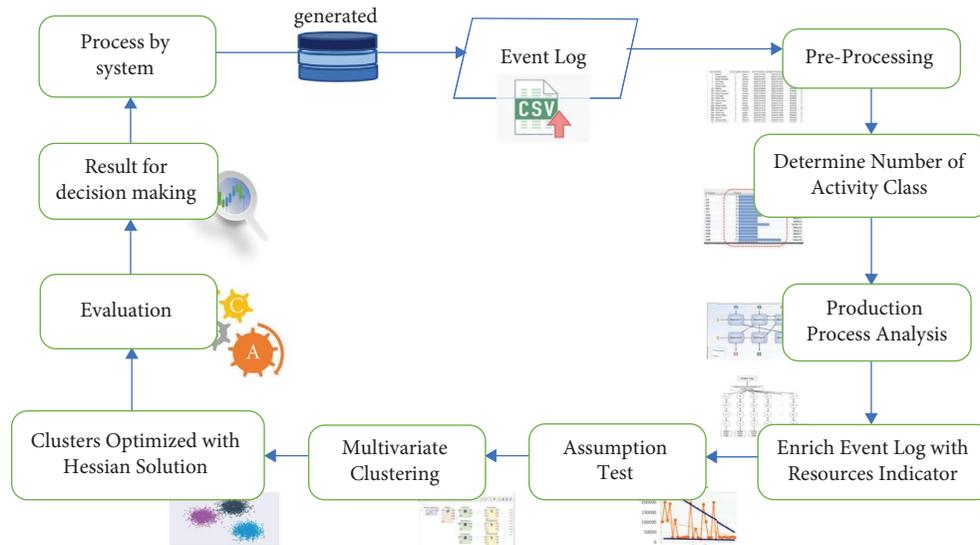


FIGURE 1: Research flow.

[59]. The rules are particularly applicable for process-focused facilities. Priority rules are applied to reduce completion time, number of jobs processed in one unit of time, and processing delays due to resource availability. The scheduling problems that will be used in this study are the flow shop and job shop. This is due to the case context that is used more to the flow shop and job shop conditions.

2.3. Method. The flow of the study conducted is shown in Figure 1.

The research steps shown in Figure 1 are described in detail.

2.3.1. *Generating Event Log.* The first step is generating an event log based on system activity records according to the explored business processes. In this study, a dataset about repair services is used and has been adapted to the needs of the analysis [60]. The dataset used is an event log data about the repair process. The business process of improvement services is the focus of process modelling in this study.

2.3.2. *Preprocessing.* After getting the event log, a pre-processing step is carried out to tidy up the event log obtained in preparation for generating indicators.

2.3.3. *Determining the Number of Activity Classes.* In this step, activities are grouped by class. In the event log, different classes of activity can occur from one case to another. The initial step in the framework of this study is determining the number of activity classes. The number of activity classes determines the process path based on the number of activities. In the process that is recorded in the event log, the number of classes may have different activities depending on the system record. This step can be visually seen in Figure 2. The step is necessary to see how many forms of activity paths occur.

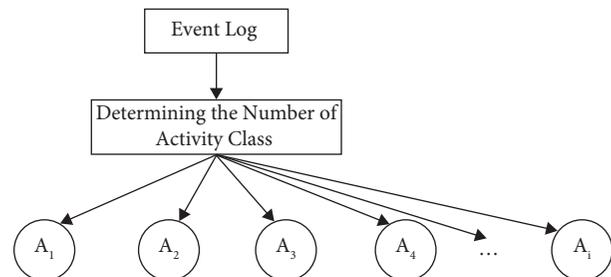


FIGURE 2: Determining the number of the activity classes model.

After determining the number of activity classes, the next step is generating a resource indicator from the event log that has been divided by the activity class. The resource indicator is based on the objectives desired by the analyst. In this study, the basic indicators produced are workload and waiting time. These two indicators are important things that affect the process [61, 62]. Then, as a complement, another indicator is selected, namely, service cost [63]. Thus, there are three indicators of concern in this study, namely, workload, waiting time, and service cost.

2.3.4. *Production Process Analysis.* In this step, a job shop analysis is performed in the event log by the activity class.

2.3.5. *Enriching Indicators from the Event Log.* In this step, the event log indicators are generated, which are resulted from the production process analysis. The steps are as follows:

- (a) *Event log indicator generation:* In the previous step, it has been stated that after determining the activity class, the indicator generation is carried out. The indicator generation is based on need. In this study, the indicators generated are workload, waiting time, and service costing. These 3 things are the reasons for

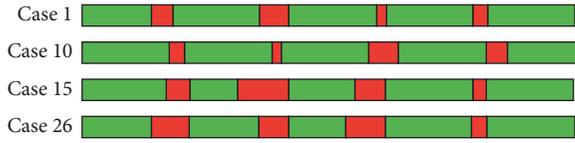


FIGURE 3: Overview of waiting time in progress.

using related resources as multivariate variables in research. This is reinforced by several studies that always focus on these three things alternately, such as the research studies conducted in [64–66] and [67]. This means that in this case there are very few studies that simultaneously use all the three.

- (b) The basic dimensions of the indicators generated in this study are workload, waiting time, and service costing. Activity indicator is one of the important indicators because the process that occurs in each case can be determined by the number of activities carried out. From one case to another, the number of activities in the process may differ. It all depends on whether the case is complicated or not.

Workload analysis measures the number of activities performed by each resource in a given period. Workload analysis shows the number of activities started, activities completed, and activities in progress. Using the activity frequency, workload analysis provides information about how much is the workload of each resource. There are many ways to find the workload value in each case in the event log. One way is using the definition from [68] that workload is the number of activities that has been carried out during a certain period. The definition of “how busy” the resource is generally means that it only looks at the resource quantitatively not qualitatively. This is because it only looks at the amount of work completed in a certain period, and it is not based on a standard assessment, in which the workload assessment can differ from one company to another. Another method is stated in Gabriel’s research that uses the factor evaluation system (FES) [69]. The FES mechanism uses clear measurements because it is based on pre-determined standards and weights. Therefore, the assessment is based on achievements in the workload. This formula was developed and adapted in this study.

There are many definitions related to waiting time in different contexts that depend on the case being discussed. Waiting can be caused by a person or machine that is required for the process, even if the person or machine is working at full capacity. This is shown in detail in Figure 3.

Figure 3 shows that the green color shows the working time of item or document, while the red color shows the waiting time that the work item or document will be worked on in the next process. In an event log process, there may be different waiting times between one case and the other. Thus, in general, the waiting time calculation equation is as follows:

$$WT_i = (T_C - T_S)_i - \sum_{j=1}^n (t_C - t_S)_j \quad (1)$$

In general, the total service cost is obtained based on the combined sum of the workload cost equation $((T_C - T_{S_i}))$ and the waiting time cost equation $(\sum_{j=1}^n ((t_C - t_S)_j))$. This can be seen in Figure 3. The calculation of service costs is based on the calculation of each workload cost and waiting time. The calculation concept is based on the project triangle approach [70]. In the project triangle concept, implementation always experiences problems, namely, cost, time, and quality. It is impossible in its implementation that a product or service will be able to apply these three components. Only two components apply out of three options. If a product or service wants a fast time and has good quality, then the value of the goods will be expensive. Otherwise, if you want cheap goods and fast time, the desired quality of goods will not be obtained [71]. In this study, quality becomes the default feature so that only two components are compared. This means that, in this case, cost and time are two inversely proportional components. Thus, in compiling the equation, we apply the inverse ratio between cost and time as the basis of the equation. The following is a detailed calculation of each equation.

(1) Activity cost: The activity cost equation is calculated in each case. The value is based on the standard cost of each activity that has been set. This means that the activity cost depends on the number of activities in each case because there may be differences in activity costs from one case to another. The activity cost equation is represented as follows:

$$NACC_i = \sum_{i=1}^n n_i(A).C_i(A). \quad (2)$$

(2) Workload cost: The workload cost value is calculated as follows:

$$WLC_i = \frac{SC(A)}{CI_i}. \quad (3)$$

(3) Waiting time cost: The impact of waiting on cost-effectiveness is highly scenario dependent. The scenario used in this study is a scenario where the cost per unit time decreases with the waiting time. Thus, in this context, the longer the waiting time (meaning that the longer the service duration), the lower the service fee. The simple logic used in this case is process cost versus waiting time, which is represented as follows :

$$CWT_i = \frac{SC(A)}{WT_i}. \quad (4)$$

(4) Service costing: The determination of service costing is based on the concept of quality. In principle, of course, it is not true that if the waiting time experienced by customers in completing the process is very long, the costs are also large. Supposedly, a short service process will have high costs consequences, while a long service process results in low costs. This is the basis for this research. Based on the project triangle concept, it will never be possible that there will be product quality along with fast turnaround times and low costs [70]. What happens is that there are only 2 possibilities for the 3 elements [71]. This means that by determining that

the process being carried out prioritizes product quality, the service cost is inversely proportional to the speed of the process completion time. Based on the calculation of these resource components and the basic concept, the total cost per case can be calculated as follows:

$$TSC_i = \sum_{j=1}^n n_i(A).Ci(A) + \left(\frac{SC(A)}{CI_i}\right) + \left(\frac{SC(A)}{WT_i}\right). \quad (5)$$

Total service costing (TSC_i) is an accumulation of activity costs, workload costs, and waiting time costs.

2.3.6. Multivariate Clustering of Resources. Clustering the event log with the indicator has been selected. In this case, after obtaining the indicator value from the generated event log, multivariate clustering will be carried out. In this context, the approach used is K-means. K-means clustering algorithm provides easy implementation, very flexible, and easy to use adaptation [72, 73]. In addition to the time needed to do learning faster, it also uses simple principles that can be explained in nonstatistical terms. This is the reason for using K-means clustering compared to the other method approaches.

2.3.7. Optimized Resource Based on Multivariate Clustering. After getting the clustering results, the last step is optimizing the selected clusters by using the Hessian method and then performing a comparison analysis of the selected clusters between class activities. The reason for using the Hessian method is because the objective function formed is a nonlinear multipolynomial function. Therefore, a nonlinear programming problem solving method is needed to get the optimum solution. The method to get solutions to nonlinear problems is by direct or indirect solutions. One of the direct solutions in the context of this study is using the Hessian solution approach without constraints. This is done because of the characteristics of the resulting data [74].

2.3.8. Evaluation. The most commonly used performance evaluation metrics such as accuracy, sensitivity, and precision are calculated and compared for the resulting clusters. Sensitivity and precision are used to ensure the performance of K-means clustering. The formulation is as follows:

$$\begin{aligned} \text{accuracy} &= \frac{(tp + tn)}{(tp + tn + fp + fn)}, \\ \text{precision} &= \frac{tp}{(tp + fp)}, \\ \text{recall} &= \frac{tp}{(tp + fn)}, \\ F - \text{measure} &= 2 \cdot \left(\frac{(\text{precision} \times \text{recall})}{(\text{precision} + \text{recall})} \right), \end{aligned} \quad (6)$$

where tp , tn , fp , and fn are true positive, true negative, false positive, and false negative, respectively.

2.3.9. Result Analysis. In this step, the analysis of the results obtained is carried out.

3. Results and Discussion

The first step in this research is to do a sampling on the event log to facilitate the simulation process [75]. The dataset of event log obtained is shown in Table 1.

Table 1 shows the event log in which activities are in the repair process, such as registering, analyzing defects, repairing (simple), and repairing (complex). Each case ID shows that the number of activities may differ from one case to another. In terms of determining the activity class from the event log, there are many types of variants or classes. The type of activity class can be detected through ProM as shown in Table 2.

The number of events in each case causes different variants in each service process provided. The number of activities between one customer and another customer may vary depending on the importance and designation of the repair services carried out. In principle, each variant is a normal activity. The difference in the number of activities does not mean that there is an abnormal activity, although in the process modelling, it is often found that there are "abnormal" activities. The idea in this study is to explore the process model that occurs through multiperspectives. Thus, in this case, the number of classes is grouped to facilitate exploration. Therefore, the next step is grouping Case ID in the class based on the number of activities as shown in Tables 3–6.

The first analysis carried out as a part of the mechanism in this study is analyzing the flow shop and job shop of the production system concept. The basis for doing this optimization is to see if in the repair process, there may be queues that can cause process delays. Therefore, if an analysis of the production system is performed, it can be seen that the queuing process can be improved through a job shop analysis approach, in order to see the optimal scheduling time as an improvement in providing future improvement processes. The first step in this stage is to determine the job and machine in the scheduling analysis based on the event log used. In job shop scheduling, every process that must be passed is considered as a machine. This is illustrated in Table 7.

Then, based on the machine notation that has been set, the next step is arranging input in the form of a job shop, as shown in Table 8. The analysis shown is the treatment carried out in class 1, which has a total of 5 activities (in this study, all treatments are only presented in class 1, while the overall results of the treatment are shown at the end).

The next step is analyzing the production system in the event log as shown in Table 9, which has been compiled by setting the rules. The collection of data processing obtained is data resources, process flow, and the processing time of each job. The rules used in the process are as follows:

- (i) The company operates every day for 12 hours
- (ii) The completion of the order process is based on the FPTO (finish production time of order) schema.

TABLE 1: Event log used in the research.

Case ID	Activity	Resource	Start timestamp	Complete timestamp
1	Register	System	02/01/70 18.23	02/01/70 18.23
1	Analyze defect	Tester3	02/01/70 18.23	02/01/70 18.30
1	Repair (complex)	SolverC1	02/01/70 18.31	02/01/70 18.49
1	Test repair	Tester3	02/01/70 18.49	02/01/70 18.55
1	Inform user	System	02/01/70 19.10	02/01/70 19.10
1	Archive repair	System	02/01/70 19.10	02/01/70 19.10
100	Register	System	04/01/70 08.28	04/01/70 08.28
100	Analyze defect	Tester4	04/01/70 08.28	04/01/70 08.36
100	Repair (complex)	SolverC1	04/01/70 08.52	04/01/70 09.09
100	Inform user	System	04/01/70 09.20	04/01/70 09.20
1000	Register	System	22/01/70 12.23	22/01/70 12.23
1000	Analyze defect	Tester6	22/01/70 12.23	22/01/70 12.28
1000	Repair (simple)	SolverS3	22/01/70 12.35	22/01/70 12.49
1000	Test repair	Tester6	22/01/70 12.49	22/01/70 12.58
1001	Register	System	22/01/70 13.51	22/01/70 13.51
—	—	—	—	—

TABLE 2: The number of activity class from event log.

Case ID	Events	Started	Finished
1	6	02.01.1970 18:23:00	02.01.1970 19:10:00
—	—	—	—
100	6	04.01.1970 08:28:00	04.01.1970 09:28:00
101	6	04.01.1970 09:00:00	04.01.1970 10:30:00
—	—	—	—
1000	6	22.01.1970 00:23:00	22.01.1970 13:08:00
1001	5	22.01.1970 13:51:00	22.01.1970 15:00:00
1009	5	22.01.1970 01:59:00	22.01.1970 02:46:00
1010	8	22.01.1970 18:08:00	22.01.1970 19:10:00
—	—	—	—
1040	5	23.01.1970 20:24:00	23.01.1970 21:21:00
1041	11	23.01.1970 19:15:00	23.01.1970 21:05:00

TABLE 3: The cluster of activity class 1.

Case ID	Events	Started	Finished
1001	5	22.01.1970 13:51:00	22.01.1970 15:00:00
1002	5	23.01.1970 03:40:00	23.01.1970 04:40:00
1004	5	22.01.1970 16:27:00	22.01.1970 17:10:00
1005	5	22.01.1970 06:31:00	22.01.1970 07:16:00
—	—	—	—

TABLE 4: The cluster of activity class 2.

Case ID	Events	Started	Finished
1	6	02.01.1970 18:23:00	02.01.1970 19:10:00
100	6	04.01.1970 08:28:00	04.01.1970 09:28:00
1000	6	22.01.1970 00:23:00	22.01.1970 13:08:00
101	6	04.01.1970 09:00:00	04.01.1970 10:30:00
—	—	—	—

Finish production time is met when all the items ordered have passed the processing stage.

(iii) An activity exactly visits one machine once

TABLE 5: The cluster of activity class 3.

Case ID	Events	Started	Finished
1003	8	22.01.1970 00:43:00	22.01.1970 13:51:00
1010	8	22.01.1970 18:08:00	22.01.1970 19:10:00
1033	8	23.01.1970 16:18:00	23.01.1970 17:08:00
1034	8	22.01.1970 22:56:00	23.01.1970 00:27:00
—	—	—	—

TABLE 6: The cluster of activity class 4.

Case ID	Events	Started	Finished
1008	11	22.01.1970 09:25:00	22.01.1970 10:47:00
1016	11	23.01.1970 06:32:00	23.01.1970 08:19:00
1024	11	23.01.1970 10:32:00	23.01.1970 11:58:00
1041	11	23.01.1970 19:15:00	23.01.1970 21:05:00
—	—	—	—

TABLE 7: Determining process name and machine notation.

No.	Process name	Machine notation
1	Register	M1
2	Analyze defect	M2
3	Repair (simple)	M3
4	Repair (complex)	M4
5	Test repair	M5
6	Inform user	M6
7	Archive repair	M7

(iv) Each machine can only process one activity at a time and is always ready at all times without any breakdowns or under repair

(v) An activity cannot be interrupted. In other words, once an activity has taken place, it must be completed before any other activity is processed on the same machine.

(vi) After an activity has been processed on a machine, the activity will be transferred directly to the next machine

TABLE 8: Job shop event log arrangement in activity class 1.

Case ID	Job shop						
	M1 (register)	M2 (analyze defect)	M3 (repair simple)	M4 (repair complex)	M5 (test repair)	M6 (inform user)	M7 (archive repair)
1001	0	5	—	48	6	0	—
1002	0	10	20	—	0	0	—
1004	0	7	11	—	9	0	—
1005	0	9	14	—	10	0	—
1006	0	6	—	40	10	0	—
1007	0	6	20	—	5	0	—
—	—	—	—	—	—	—	—

- (vii) Each activity has a certain type of activity, which is carried out by a certain machine and in a certain processing time
- (viii) Human and machine resources are set to be available 12 hours per day for 7 days per week
- (ix) No machine breaks down or stops working
- (x) Service processes are based on orders and are carried out sequentially
- (xi) The standard fee used for one full service is 100,000 (in this case, we modify it in Indonesian rupiah)

Many methods can be used to analyze the job shop scheduling performance, such as FCFS (first come-first served), SPT (short processing time), LPT (long processing time), and EDD (earliest due date). The choice of method is very dependent on the needs. Each method has advantages and disadvantages. In the analysis of scheduling carried out, the three methods used are FCSC, SPT, and LPT, as shown in Table 9.

In addition, it can also be seen in Figure 4.

Based on Table 9, it is found that the job shop scheduling with SPT rule has a shorter flow time compared to other rules. This can also be seen in Figure 4. In principle, the selection of which scheduling to use is left to the user because each priority rule has its own advantages and possibilities. Different evaluation results on different order data may occur. However, in this study, the SPT rule is chosen. The graphs show that the SPT rule minimizes makespan 28818 and average flow time 873.27. Therefore, the next step is using the event log based on the job shop scheduling analysis and the selected SPT rule. These results are in accordance with the research studies conducted in [76, 77]. The next mechanism is performing clustering in the event log by previously generating resource values for workloads, waiting times, and service costing. The values are shown in Table 10,

TABLE 9: Scheduling performance with 3 rules.

	Average flow time	Average number of jobs in system	Makespan
FCFS	1126.24242	14.3831269	37166
SPT	873.272727	11.1524768	28818
LPT	1789.0303	22.8475232	59038

which are based on the calculation results of equations (3)–(5).

The process of performing clustering is based on the fact that the resulting multiple perspectives or resources may vary or are not homogeneous. Before doing clustering, the data have been tested by using one of the indicators of the classical assumption (the heteroscedasticity test) with a 5% confidence level. The data having been tested show that the data have heteroscedasticity or nonhomogeneous conditions. The clustering method uses K-means with the Davies Bouldin approach. Davies Bouldin's value obtained is shown in Table 11.

Based on Davies Bouldin's value, the number of the selected cluster is 4. Cluster number 4 is selected because in the 4th iteration, Davies Bouldin's value drops to 0.410, which means that it stops at the 3rd iteration and that the event log is divided into four clusters. The results of K-means clustering show that the order of clusters from the best cluster is cluster 1, cluster 2, cluster 3, and cluster 0. This is shown in Tables 12–15.

From each table, it can determine the objective function model with the Hessian solution approach. The objective function model is generated by using an approach in the form of multi-nonlinear polynomial. After going through the linearity test, by using a nonlinear multipolynomial equation approach, the objective function for cluster 0 in class 1 is obtained as equation (6). Cluster 0 is chosen because it is the best cluster compared to other clusters in class 1.

$$S_c(W_L, W_T) = 104893.48 + 25.59W_L - 223.62W_T - 0.008W_LW_T - 0.04W_L^2 + 1.93W_T^2. \quad (7)$$

In this context, the resulting objective function has no constraints. Therefore, the solution approach that can be

used to find the minimum service costing value is the Hessian optimization solution approach. By using the

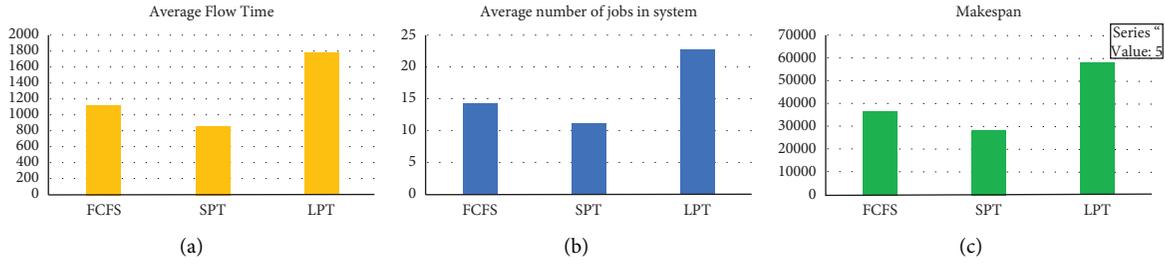


FIGURE 4: The graphs of job shop scheduling performance based on FCFS, SPT, and LPT.

TABLE 10: Resource values in event log.

Case ID	Workload	Waiting time	Service costing
1001	290	29	103793
1002	280	28	103929
1004	280	16	106607
1005	290	12	108678
1006	280	21	105119
1007	300	21	105095
—	—	—	—
1039	310	24	104489
1040	290	22	104890

TABLE 11: Davies Bouldin’s value in K-means clustering iteration.

No	Number of K	Davies Bouldin’s value
1	2	0.014
2	3	0.297
3	4	0.435
4	5	0.410

TABLE 12: Cluster 0 class 1.

Case ID	Workload	Waiting time	Service costing
1001	290	29	103793.10
1002	280	28	103928.57
—	—	—	—

TABLE 13: Cluster 3 class 1.

Case ID	Workload	Waiting time	Service costing
1004	280	16	106607.14
1006	280	21	105119.05
—	—	—	—

Hessian optimization solution for nonlinear multi-polynomial equations, the minimum cost function value is 102363.07 at $W_L = 320$ and at $W_T = 57.27$.

The same process is carried out on the other clusters, and the results are shown in Table 16.

TABLE 14: Cluster 2 class 1.

Case ID	Workload	Waiting time	Service costing
1005	290	12	108678.16
1011	290	14	107487.68
—	—	—	—

TABLE 15: Cluster 1 class 1.

Case ID	Workload	Waiting time	Service costing
1014	290	1	200344.83
1019	290	1	200344.83
—	—	—	—

TABLE 16: Optimization score for each class.

Class	Waiting time	Workload	Service cost
Class 1	57.21	320	102363.07
Class 2	31.20	151.04	104610.99
Class 3	19.55	322.93	109008.58
Class 4	25.5	143.82	106708.98

TABLE 17: Detail evaluation of each class for clustering performance.

Class	Accuracy	Precision	Recall	F-measure
Class 1	0.82	0.94	0.90	0.92
Class 2	0.85	0.92	0.90	0.91
Class 3	0.83	0.93	0.90	0.92
Class 4	0.82	0.93	0.90	0.92

Table 16 shows the results of the comparison of the minimum costs for each class.

4. Evaluation

In this section, measurements are made to see the performance of K-means clustering. Table 17 shows the performance of the 4 classes produced in this study.

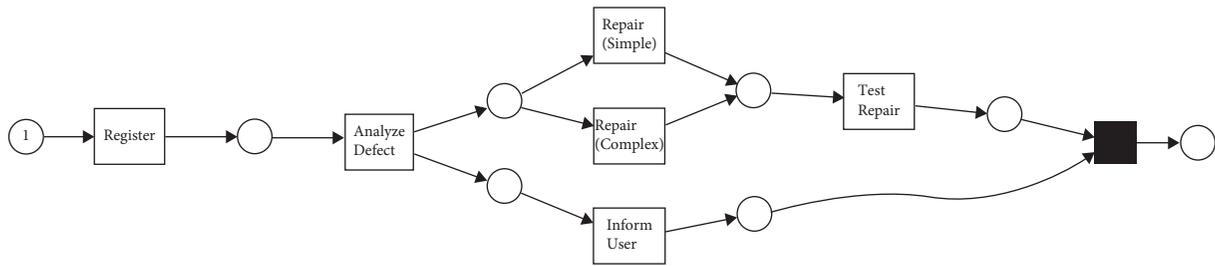


FIGURE 5: Process model with Petri net that has a minimum service cost of 102363.07.

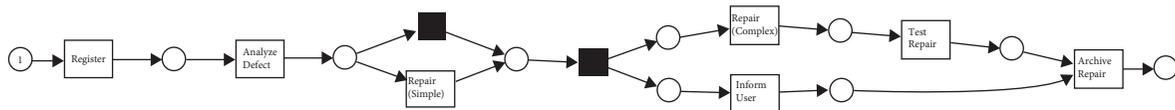


FIGURE 6: Process model with Petri net that has a minimum service cost of 104610.99.

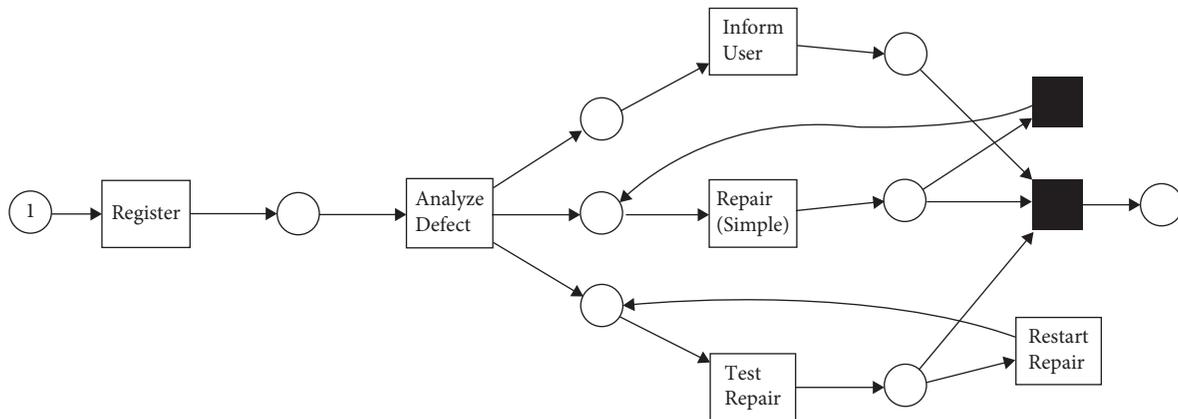


FIGURE 7: Process model with Petri net that has a minimum service cost of 109080.58.

The performance results show that the F-measure value reaches 0.92 in almost all classes. This shows that the clustering performance is good enough.

5. Discussion

Based on the simulation, it is found that the job shop schedules with SPT rule as an initial reference to generate resource value from the process. This mechanism shows that the patterns of running services can be seen from the types of transactions that occur. This can provide input for the process owner to consider which type of transaction can take precedence, in order to get the desired time optimization. From the analysis results of the job shop, the optimal time sequence is obtained, and the values (workload, waiting time, and service costs) needed in the next analysis are generated and then are clustered based on the number of classes. The calculation results show that cluster 0 is the cluster that has the most optimal objective function value compared to other clusters in class 1. From four classes, four process models (that are accompanied by a minimum cost value) are obtained, which are described by using a Petri net with

inductive miner algorithm. Figures 5–8 show the process model of each class.

The results obtained indicate that the model in Figure 7 has the lowest cost compared to other models. When referring to the minimum service cost, the process owner will definitely choose the model in Figure 7. However, in this case, further analysis is needed to determine which process model is better. Note that the described process model comes from 4 classes. Figure 5 is a process model generated from class 1. Class 1 has an optimal cost value of 102363.07 with $W_L = 320$ and $W_T = 57.21$. This value can be optimized in the process by eliminating the restart repair activity. While in Figure 7, which comes from class 3, the repair (complex) activity is eliminated. This causes the process cost to be optimal at 109008.58 at points $W_L = 322.93$ and $W_T = 19.55$. In addition, what can be used as a comparison is that the process model in Figure 7 has a lower waiting time value compared to the process model in Figure 6. The consequence of a low waiting time results in a larger workload value. This has the consequence that service costs become more expensive. While, the models in Figures 6 and

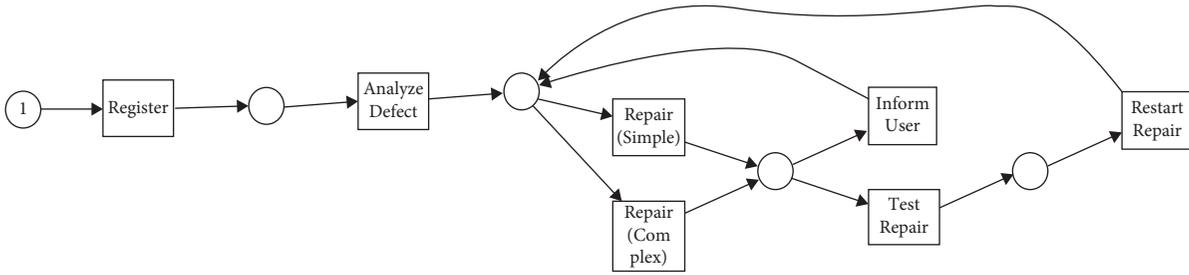


FIGURE 8: Process model with Petri net that has a minimum service cost of 106708.98.

TABLE 18: Notations.

Notation	Descriptions
n	Number
i	Case
j	Trace
WT_i	Waiting time per case
TC	End duration
TS	Initial duration
tc	End duration per trace
ts	Initial duration per trace
$n_i (A)$	Number activity per case
$C_i (A)$	Activity standard cost per trace
$NACC_i$	Number of activity per case
WLC_i	Workload cost per case
$SC (A)$	Total activity standard cost
CI_i	Consistency index per case
CWT_i	Waiting time cost per case
W_T	Waiting time per case
W_L	Workload
S_C	Service costing

8 have the same number of activities. The difference lies in the last activity, where the model in Figure 6 eliminates the repair archive activity, while the model in Figure 8 eliminates the restart repair activity. The process model in Figure 6 has an optimal process value at service costs of 104610.99 which has a value of $W_L = 151.04$ and $W_T = 31.20$. While, the process model in Figure 8 has an optimal service cost of 106708.98. With this analysis, the process model that occurs in the cluster can be a reference for process owners in evaluating and improving ongoing processes.

The weakness of this research is that almost half of the simulation activities are still carried out manually in Excel, while the rest use applications, such as POM-QM, ProM, DISCO, and RapidMiner. To overcome this, it is necessary to build an algorithm as a basis for compiling a special analytical tool, so that it can be a solution to the optimization problem in this study. Then, the case study used is a study that allows job shop analysis where the process flow does not occur constantly or each operation can be carried out by more than one agent. In addition, we have limitations in terms of available datasets. Apart from that, we also have limitations in conducting simulations such as the dispatching rule mechanism. Therefore, further research refinement steps are needed.

6. Conclusions and Future Works

The study that has been conducted provides good alternative results because it makes a major contribution to the analysis of the process in the case used in it. The mechanism is carried out in several stages, namely, production system analysis, multivariate clustering on class activities, and cluster optimization that provides a new perspective in analyzing the process. The production system analysis stage is able to produce a more optimal scheduling analysis in handling processing time, in which this is not a concern in most studies on the process mining. The resulting job shop scheduling analysis provides notes for the process owner to pay attention to which activity sequence that causes the process time to be more optimal with the SPT Rule. This will be an input in the management planning process within the company.

Then, the process of generating resource indicators becomes an advantage in the stages carried out. Generating the value of resource indicators makes the analysis stronger because it not only focuses on the sequence of activities but also pays attention to multiperspective aspects. The process that occurs is rich in information, so it must be a concern. Multivariate clustering stages on resource indicators are able to cluster the event log to be more homogeneous based on the resulting indicators. The resulting information can add indicators of the analysis, in addition to clustering based on the sequence of activities.

The last stage in this research is the optimization of clusters resulting from the K-means clustering process. With the results of clustering performance which is quite good at 0.92, resource indicator data can be used as the basis for optimization calculations. Optimization results provide information on which processes are optimal based on service cost indicators. Cost optimization becomes a significant force in seeing which process is the most optimal. It provides meaningful input into the overall process analysis. The approach taken in this research provides a new perspective in viewing and analyzing processes within the company. The collaboration between multiperspective aspects of process mining and production systems will open up new insights and perspectives on related research studies. There are still many multiperspective aspects of the process that have not been touched upon and require further study. The notational

explanation of all the formulas in this article is shown in Table 18.

Data Availability

The event log data used to support the findings of this study are deposited in the Mendeley Data repository (DOI: 10.17632/w9hpjv84k9.1).

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors would like to thank their colleagues from the Computer Science Doctoral Program with scheme no. IF186401 at the FTEIC Intelligent Information Management laboratory, Institut Teknologi Sepuluh Nopember, Surabaya, as pleasant colleagues during the research discussion and PPM unit of Telkom University that has provided assistance and support in this research process.

References

- [1] J. Hillmann, "Disciplines of organizational resilience: contributions, critiques, and future research avenues," *Review of Managerial Science*, vol. 15, no. 4, pp. 879–936, 2021.
- [2] R. Accorsi, M. Ullrich, and W. M. P. Van Der Aalst, "Process mining," *Informatik-Spektrum*, vol. 35, no. 5, pp. 354–359, 2012.
- [3] S. Aguirre, C. Parra, and J. Alvarado, "Combination of process mining and simulation techniques for business process redesign: a methodological approach," *Lecture Notes in Business Information Processing*, vol. 162, pp. 24–43, 2013.
- [4] J. vom Brocke, A. van der, W. M. P. Van Der Aalst, W. Kremser et al., "Process science: the interdisciplinary study of continuous change," *SSRN Electronic Journal*, vol. 9, 2021.
- [5] R. Dijkman, O. Turetken, G. R. van Ijzendoorn, and M. de Vries, "Business processes exceptions in relation to operational performance," *Business Process Management Journal*, vol. 25, no. 5, pp. 908–922, 2019.
- [6] B. F. A. Hompes, J. C. A. M. Buijs, W. M. P. Van Der Aalst, P. M. Dixit, and J. Buurman, "Detecting change in processes using comparative trace clustering," *CEUR Workshop Proceedings*, vol. 1527, pp. 95–108, 2015.
- [7] M. Song, C. W. Günther, and W. M. P. Van der Aalst, "Trace clustering in process mining," in *Proceedings of the International Conference on Business Process Management*, pp. 109–120, Milano, Italy, September 2008.
- [8] R. P. J. C. Bose and W. M. P. Van Der Aalst, "Context aware trace clustering: towards improving process mining results," in *Proceedings of the 2009 SIAM International Conference on Data Mining*, pp. 397–408, Sparks, Nevada, USA, May 2009.
- [9] R. P. J. C. Bose and W. M. P. Van Der Aalst, "Trace clustering based on conserved patterns: towards achieving better process models," *Business Process Management Workshops*, vol. 43, pp. 170–181, 2010.
- [10] A. K. A. De Medeiros, A. Guzzo, G. Greco et al., "Process mining based on clustering: a quest for precision," *Business Process Management Workshops*, vol. 4928, no. 1, pp. 17–29, 2008.
- [11] J. De Weerd, S. K. L. M. Vanden Broucke, J. Vanthienen, and B. Baesens, "Leveraging process discovery with trace clustering and text mining for intelligent analysis of incident management processes," in *Proceedings of the 2012 IEEE Congress on Evolutionary Computation, CEC*, Brisbane, QLD, Australia, June 2012.
- [12] J. De Weerd, S. Vanden Broucke, J. Vanthienen, and B. Baesens, "Active trace clustering for improved process discovery," *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 12, pp. 2708–2720, 2013.
- [13] M. Fani Sani, M. Boltenhagen, and W. van der Aalst, "Prototype selection using clustering and conformance metrics for process discovery," *Business Process Management Workshops*, vol. 397, pp. 281–294, 2020.
- [14] C. Di Francescomarino, M. Dumas, F. M. Maggi, and I. Teinmaa, "Clustering-based predictive process monitoring," *IEEE Transactions on Services Computing*, vol. 12, no. 6, pp. 896–909, 2019.
- [15] G. Greco, A. Guzzo, L. Pontieri, and D. Sacca, "Discovering expressive process models by clustering log traces," *IEEE Transactions on Knowledge and Data Engineering*, vol. 18, no. 8, pp. 1010–1027, 2006.
- [16] S. Jablonski, M. Röglinger, S. Schöning, and K. Wyrtyki, "Multi-perspective clustering of process execution traces," *Enterprise Modelling and Information Systems Architectures*, vol. 14, no. 2, pp. 1–22, 2019.
- [17] A. Meinheim, C. D. S. Garcia, J. C. Nievola, and E. E. Scalabrin, "Combining process mining with trace clustering: manufacturing shop floor process-an applied case," in *Proceedings of the International Conference on Tools with Artificial Intelligence, ICTAI*, Boston, MA, USA, November 2018.
- [18] S. Montani and G. Leonardi, "Retrieval and clustering for business process monitoring: results and improvements," *Lecture Notes in Computer Science*, vol. 7466, pp. 269–283, 2012.
- [19] R. Neubaer, M. Fantinato, and M. Peres, "Interactive trace clustering," in *Proceedings of the XV Simpósio Brasileiro de Sistemas de Informação*, pp. 45–48, September 2019.
- [20] S. J. van Zelst and Y. Cao, "A generic framework for attribute-driven hierarchical trace clustering," *Business Process Management Workshops*, vol. 397, pp. 308–320, 2020.
- [21] J. Xu and J. Liu, "A profile clustering based event logs repairing approach for process mining," *IEEE Access*, vol. 7, pp. 17872–17881, 2019.
- [22] F. Zandkarimi, J. R. Rehse, P. Soudmand, and H. Hoehle, "A generic framework for trace clustering in process mining," *2020 2nd International Conference on Process Mining (ICPM)*, vol. 2020, pp. 177–184, 2020.
- [23] P. De Koninck and J. De Weerd, "Scalable mixed-paradigm trace clustering using super-instances," in *Proceedings of the 2019 International Conference on Process Mining (ICPM)*, Aachen, Germany, June 2019.
- [24] P. De Koninck, K. Nelissen, S. vanden Broucke, B. Baesens, M. Snoeck, and J. De Weerd, "Expert-driven trace clustering with instance-level constraints," *Knowledge and Information Systems*, vol. 63, no. 5, pp. 1197–1220, 2021.
- [25] J. Evermann, T. Thaler, and P. Fettke, "Clustering traces using sequence alignment," *Business Process Management Workshops*, vol. 256, pp. 179–190, 2016.
- [26] F. Prathama, B. N. Yahya, D. D. Harjono, and M. Er, "Trace clustering exploration for detecting sudden drift: a case study in logistic process," *Procedia Computer Science*, vol. 161, pp. 1122–1130, 2019.

- [27] G. M. Veiga and D. R. Ferreira, "Understanding spaghetti models with sequence clustering for ProM," *Lecture Notes in Business Information Processing*, vol. 43, pp. 92–103, 2010.
- [28] P. Wang, W. Tan, A. Tang, and K. Hu, "A novel trace clustering technique based on constrained trace alignment," *Lecture Notes in Computer Science*, vol. 10745, pp. 53–63, 2018.
- [29] T. Stocker, "Time-based trace clustering for evolution-aware security audits," in *Proceedings of the International Conference on Business Process Management*, pp. 471–476, Clermont-Ferrand, France, August 2011.
- [30] D. Luengo and M. Sepúlveda, "Applying clustering in process mining to find different versions of a business process that changes over time," in *Proceedings of the International Conference on Business Process Management*, pp. 153–158, Clermont-Ferrand, France, August 2011.
- [31] M. Boltenhagen, T. Chatain, J. Carmona et al., "Generalized alignment-based trace clustering of process behavior to cite this version :HAL id : hal-02176771," 2019, <https://www.lsv.fr/Publis/PAPERS/PDF/BCC-atpn19.pdf>.
- [32] T. Chatain, J. Carmona, and B. van Dongen, "Alignment-based trace clustering," *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 10650 LNCS, vol. 10650, pp. 295–308, 2017.
- [33] A. Koschmider and D. S. V. Moreira, "Change detection in event logs by clustering," in *Proceedings of the on the Move to Meaningful Internet Systems. OTM 2018 Conferences: Confederated International Conferences: CoopIS, C&TC, and ODBASE, 2018, Proceedings, Part I*, pp. 643–660, Springer International Publishing, Valletta, Malta, October 2018.
- [34] M. Dobrota, B. Delibašić, and P. Delias, "A skiing trace clustering model for injury risk assessment," *International Journal of Decision Support System Technology*, vol. 8, no. 1, pp. 56–68, 2016.
- [35] Q. T. Ha, H. N. Bui, and T. T. Nguyen, "A trace clustering solution based on using the distance graph model," *Lecture Notes in Computer Science*, vol. 9875, pp. 313–322, 2016.
- [36] A. Koschmider, D. Siqueira, and V. Moreira, "Change detection in event logs by clustering approach to structure event logs to provide solutions for the research questions RQ1 and RQ2 from figure 2,".
- [37] D. Lee, J. Park, I. R. Pulshashi, and H. Bae, "Clustering and operation analysis for assembly blocks using process mining in shipbuilding industry," *Lecture Notes in Business Information Processing*, vol. 159, pp. 67–80, 2013.
- [38] S. Schuhmann, J. R. Rehse, S. Baumann, and P. Fettke, "Interactive process clustering with t-SNE," *CEUR Workshop Proceedings*, vol. 2673, pp. 82–86, 2020.
- [39] A. Seeliger, T. Nolle, and M. Mühlhäuser, "Finding structure in the unstructured: hybrid feature set clustering for process discovery," in *Proceedings of the International Conference on Business Process Management*, pp. 288–304, Sydney, NSW, Australia, September 2018.
- [40] Y. Sun, B. Bauer, and M. Weidlich, "Compound trace clustering to generate accurate and simple sub-process models," in *Proceedings of the Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 10601 LNCS, Malaga, Spain, November 2017.
- [41] H. Ariouat, A. H. Cairns, K. Barkaoui et al., "A two-step clustering approach for improving educational process model discovery Nasser Khelifa to cite this version :HAL Id : hal-01500511," 2017, <https://hal.science/hal-01500511/document>.
- [42] P. Nguyen, A. Slominski, V. Muthusamy, V. Ishakian, and K. Nahrstedt, "Process trace clustering: a heterogeneous information network approach," in *Proceedings of the 16th SIAM International Conference on Data Mining 2016*, pp. 279–287, SDM, Miami FL, USA, May 2016.
- [43] T. Lehto and M. Hinkka, "Discovering business area effects to process mining analysis using clustering and influence analysis," in *Proceedings of the International Conference on Business Information Systems*, pp. 236–248, Colorado Springs, CO, USA, June 2020.
- [44] F. Mannhardt, M. De Leoni, and H. A. Reijers, "The multi-perspective process explorer," *BPM (Demos)*, Ivm, Armonk, New York, NY, USA, 2015.
- [45] A. Senderovich, "Service analysis and simulation in process mining," *Business Process Management Workshops*, vol. 202, pp. 578–581, 2015.
- [46] T. B. H. Tu and M. Song, "Analysis and prediction cost of manufacturing process based on process mining," in *Proceedings of the 2016 International Conference on Industrial Engineering, Management Science and Application (ICIMSA)*, pp. 1–5, Jeju, South Korea, May 2016.
- [47] W. Zhao, L. Yang, H. Liu, and R. Wu, "The optimization of resource allocation based on process mining," *Lecture Notes in Computer Science*, vol. 9227, pp. 341–353, 2015.
- [48] W. Van Der Aalst, "Process mining," *Communications of the ACM*, vol. 55, no. 8, pp. 76–83, 2012a.
- [49] K. Okoye, S. Islam, U. Naeem, and M. S. Sharif, "The application of a semantic-based process," *Proceedings of SAI Intelligent Systems Conference*, vol. 1, pp. 1381–1403, 2018.
- [50] F. Caron, J. Vanthienen, and B. Baesens, "Comprehensive rule-based compliance checking and risk management with process mining," *Decision Support Systems*, vol. 54, no. 3, pp. 1357–1369, 2013.
- [51] W. M. P. Van der Aalst, "Process discovery: an introduction," in *Process Mining*, pp. 125–156, Springer, Berlin, Germany, 2011.
- [52] J. C. A. M. Buijs, B. F. Van Dongen, and W. M. P. van Der Aalst, "On the role of fitness, precision, generalization and simplicity in process discovery," in *Proceedings of the OTM Confederated International Conferences On the Move to Meaningful Internet Systems*, pp. 305–322, Rome, Italy, September 2012.
- [53] S. S. Nezhad and R. G. Assadi, "Preference ratio-based maximum operator approximation and its application in fuzzy flow shop scheduling," *Applied Soft Computing*, vol. 8, no. 1, pp. 759–766, 2008.
- [54] P. Brucker, "Scheduling algorithms," *Journal of the Operational Research Society*, vol. 50, p. 774, 1999.
- [55] S. C. Graves, "A review of production scheduling," *Operations Research*, vol. 29, no. 4, pp. 646–675, 1981.
- [56] A. Baskar and A. Xavier, "A simple model to optimize general flow shop scheduling problems with known break down time and weights of jobs," *Procedia Engineering*, vol. 38, no. 1, pp. 191–196, 2012.
- [57] S. Wang and M. Liu, "Two-stage hybrid flow shop scheduling with preventive maintenance using multi-objective tabu search method," *International Journal of Production Research*, vol. 52, no. 5, pp. 1495–1508, 2014.
- [58] S. Zhang and S. Wang, "Flexible assembly job-shop scheduling with sequence-dependent setup times and part sharing in a dynamic environment: constraint programming model, mixed-integer programming model, and dispatching rules," *IEEE Transactions on Engineering Management*, vol. 65, no. 3, pp. 487–504, 2018.

- [59] M. E. Leusin, E. M. Frazzon, M. Uriona Maldonado, M. Kück, and M. Freitag, "Solving the job-shop scheduling problem in the industry 4.0 era," *Technologies*, vol. 6, no. 4, p. 107, 2018.
- [60] B. F. van Dongen, "BPI challenge 2015. BPI challenge 2015. 4TU.ResearchData. Collection," in *Proceedings of the 11th International Workshop on Business Process Intelligence (BPI 2015)*, Innsbruck, Austria, August 2015.
- [61] R. R. Chen, S. Kumar, J. Singhal, and K. Singhal, "Note: the value and cost of the customer's waiting time," *Manufacturing & Service Operations Management*, vol. 23, 2020.
- [62] M. Park, M. Song, T. H. Baek, S. Y. Son, S. J. Ha, and S. W. Cho, "Workload and delay analysis in manufacturing process using process mining," *Lecture Notes in Business Information Processing*, vol. 219, pp. 138–151, 2015.
- [63] B. Meidyani, R. Sarno, and A. L. Nurlaili, "Time and cost optimization using scheduling job shop and linear goal programming model," in *Proceedings of the 2018 International Conference on Information and Communications Technology, ICOIAC 2018*, pp. 555–560, Yogyakarta, Indonesia, March 2018.
- [64] X. Huang, R. Yu, D. Ye, L. Shu, and S. Xie, "Efficient workload allocation and user-centric utility maximization for task scheduling in collaborative vehicular edge computing," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 4, pp. 3773–3787, 2021.
- [65] N. Kumar and A. Mondal, "Work-in-progress: pricing mechanism and workload scheduling to optimize social welfare and cost for fog computing systems," in *Proceedings of the 2019 IEEE Real-Time Systems Symposium (RTSS)*, pp. 544–547, Hong Kong, China, December 2019.
- [66] Q. Qian, P. Guo, and R. Lindsey, "Comparison of subsidy schemes for reducing waiting times in healthcare systems," *Production and Operations Management*, vol. 26, no. 11, pp. 2033–2049, 2017.
- [67] A. Badshah, A. Ghani, S. Shamshirband, and A. T. Chronopoulos, "Optimising infrastructure as a service provider revenue through customer satisfaction and efficient resource provisioning in cloud computing," *IET Communications*, vol. 13, no. 18, pp. 2913–2922, 2019.
- [68] J. Nakatumba and W. M. P. Van Der Aalst, "Analyzing resource behavior using process mining," *Business Process Management Workshops*, vol. 43, pp. 69–80, 2010.
- [69] G. Sophia and R. Sarno, "Process mining and factor evaluation system method for analysing job performance," in *Proceedings of the 2018 International Symposium on Advanced Intelligent Informatics: Revolutionize Intelligent Informatics Spectrum for Humanity*, pp. 153–156, Yogyakarta, Indonesia, August 2019.
- [70] M. J. Wolf and F. S. Grodzinsky, "Good/fast/cheap: contexts, relationships and professional responsibility during software development," in *Proceedings of the 2006 ACM Symposium on Applied Computing*, pp. 261–266, Dijon, France, April 2006.
- [71] T. Stafford, "Editorial preface: 'fast, cheap or good; pick any two,'" *Journal of Global Information Technology Management*, vol. 15, no. 3, pp. 1–4, 2012.
- [72] D. S. Maylawati, T. Priatna, H. Sugilar, and M. A. Ramdhani, "Data science for digital culture improvement in higher education using K-means clustering and text analytics," *International Journal of Electrical and Computer Engineering*, vol. 10, no. 5, pp. 4569–8708, 2020.
- [73] A. M. Shabut, M. Hoque Tania, K. T. Lwin et al., "An intelligent mobile-enabled expert system for tuberculosis disease diagnosis in real time," *Expert Systems with Applications*, vol. 114, pp. 65–77, 2018.
- [74] H.-H. Lue, "On principal Hessian directions for multivariate response regressions," *Computational Statistics*, vol. 25, no. 4, pp. 619–632, 2010.
- [75] M. F. Sani, S. J. Van Zelst, and W. M. P. Van der Aalst, "The impact of event log subset selection on the performance of process discovery algorithms," in *Proceedings of the European Conference on Advances in Databases and Information Systems*, pp. 391–404, Bled, Slovenia, September 2019.
- [76] M. R. Garey, D. S. Johnson, and R. Sethi, "The complexity of flowshop and jobshop scheduling," *Mathematics of Operations Research*, vol. 1, no. 2, pp. 117–129, 1976.
- [77] J. K. Lenstra, A. H. G. R. Kan, and P. Brucker, "Complexity of machine scheduling problems," *Annals of discrete mathematics*, vol. 1, pp. 343–362, 1977.