



## Performance Evaluation of Inductive Miner for Internship Program Workflow

Afina Lina Nurlaili<sup>1</sup>, Muhsin<sup>2</sup>, Dhian Satria Yudha Kartika<sup>3</sup>

<sup>1</sup>Informatics, Universitas Pembangunan Nasional “Veteran” Jawa Timur, Surabaya, Indonesia

<sup>2</sup>Telecommunication Engineering, Institut Teknologi Telkom Surabaya, Surabaya, Indonesia

<sup>3</sup>Information System, Universitas Pembangunan Nasional “Veteran” Jawa Timur, Surabaya, Indonesia

E-mail: [afina.lina.if@upnjatim.ac.id](mailto:afina.lina.if@upnjatim.ac.id)

### ARTICLE INFO

### ABSTRACT

#### Article history:

Received: August 23, 2021  
Revised: September 29, 2021  
Accepted: October 18, 2021

#### Keywords:

Process discovery,  
Process mining,  
Alpha,  
Heuristic,  
Inductive

Modern organizations involve a high number of activities in their operations, which can be very complex. Process model is used to simplify these complex activities. Various algorithms for discovering process models have been developed in process mining. Process mining can extract important information from each activity in obtained cases. There are three process mining algorithms commonly used: Alpha, Heuristic Miner, and Inductive Miner. Each of the algorithms has its own characteristics. This paper compares these algorithms for the Internship Program. Based on the obtained evaluation, Alpha algorithm can't describe the process based on PKL event log well. It is because loop processes exist. It is also shown that the PKL process is not implementing the SOP well yet. On the other hand, Heuristic Miner neglects minor processes which do not frequently happen and does not describe it in the process model. Inductive miner combines the working principle of Alpha Miner and Heuristic Miner. Overall, the process model that is formed uses the Alpha algorithm which is closest to reality because it has a fitness of 0.96.

Copyright © 2021 Jurnal Mantik.  
All rights reserved.

## 1. Introduction

Currently, many organizations have problems with time efficiency [1]. They try to find the most optimized workflow by reducing the duration of some activities or removing unimportant activities [2]. The most promising technique to solve this problem is creating a business process procedure in form of a Standard Operating Procedure (SOP) [3]. SOP is the main guidance as a reference of activities. SOP is created in form of text and process models ensuring easy understanding [4]. Activities of SOP is documented in event log [2]. This event log is usually stored in an information system, so that will be easier to monitor and evaluate [5]. Most event log consist of traces, cases, and activities. In addition, some event log also has timestamps and resources.

The process model of SOP and event log usually have some mismatches. These mismatches lead to fraud, including corruption and financial loss. It can be severe for organizations or processes efficiency. Process mining is a powerful method to evaluate business process based on event log [6]. Process mining begins by process discovery which visualizes process model. After that conformance checking is conducted between process model and event log to evaluate its performance. Finally, the evaluation is used to enhance process model.

Process discovery becomes a vital part of process mining since it provides data for the next part [7]. The wrong visualization from process discovery causes wrong interpretation of SOP and event log. Moreover, conformance checking could be inaccurate and process model cannot be enhance effectively [1]. We use three algorithms of process discovery: Alpha, Heuristic Miner, and Inductive Miner to evaluate internship business process. Then we analysed each algorithm in fitness, precision, simplicity, and generalization. These algorithms are evaluated on internship program workflow in higher education institution.



## 2. Method

### 2.1 Conformance Checking Evaluation

#### a. Fitness

Fitness between event log and process model represents quality of process model. Fitness is focused on traces of event log which can be represented successfully in the process model. It values within the range of 0 - 1. Fitness is formulated as

$$F = \frac{A + B}{2} \tag{1}$$

with

$$A = \frac{tm}{tt} \tag{2}$$

and

$$B = \frac{1}{2} \frac{ta - te}{ta} + \frac{1}{2} \frac{ta - tr}{ta} \tag{3}$$

where c is the number of traces in event log which can be represented in the process model, t is the total trace of event log, m is the number of activities in event log which cannot be represented in the process model, r is the number of activities of trace in event log and e is total activity in event log. In case of all traces can be captured perfectly in the process model, it will be valued as 1. Nevertheless, if all traces in event log cannot be captured in process model from the start until last activity, then the value fitness of process model will be 0.

#### b. Precision

Precision measures if an algorithm is proper to solve the following case. Precision value calculated

$$P = \frac{ttp}{t} \tag{4}$$

where ttp is the number of true positive and t is the total true positive and false negative. These values are referred to sets of cases with same traces in event log which can be captured in the process model. Perfect precision (P=1) leads to overfitting. It means the process model is very specific. This very specific characteristic made the process model cannot be used in other cases, even with very small differences.

## 3. Results and Analysis

### 3.1 Event Log

This paper uses event logs for the analysis process. Event log can store various kinds of information [11]. The event log is the result of recording activity every time someone doing something on the information system. Event logs that have been collected and then analyzed so that it can be used as a guideline for making business process models of a company for the future [9]. To be able to apply process mining techniques, the first thing to do is retrieve event logs from data sources (eg: databases, transaction logs, audit trails, etc.). XES is a standard format for process mining that is supported by most process tools mining. XES was created in 2010 by the IEEE Task Force on Process Mining. Next to XES (eXtensible Event Stream), another format supported by the ProM toolkit is an MXML file (Mining eXtensible Markup Language).

The event log has several terms that are often used as a basis analysis, namely case, trace, activity, time, resource [12]. Case is a series of activities that connected to each other starting from the initial activity to the final activity to solve the problem certain. Trace is a collection of cases that have the same path or sequence similarity activity. Activities are work that has been carried out in a business process. Time is the execution time recorded for each activity. Resources are people or a device in charge of completing each activity. We use Internship Process event log from XYZ universities. In this paper event log consists of 26 activities, 7 traces, and 50 cases.

Trace ID	Case ID	Activity
1	1	A
1	1	B
1	1	C
1	1	E
1	1	F
1	1	G
1	1	H
1	1	J
1	1	K
1	1	L
1	1	M
1	1	N
1	1	N
1	1	O
1	1	O
1	1	P
1	1	Q
1	1	R
1	1	S

Fig 1. Fragment of Event log

### 3.2 Process Discovery Using Alpha Algorithm

Alpha Algorithm is the first original algorithm on process discovery [13]. It works by connecting one activity to other activities based on event log, so the whole traces of activities can be successfully constructed. However, it cannot detect trace on one-length-loop. In this one-length-loop, the activities have trace to itself. This becomes the main weakness of alpha algorithm. Weakness of Alpha Algorithm is shown in Table 1.

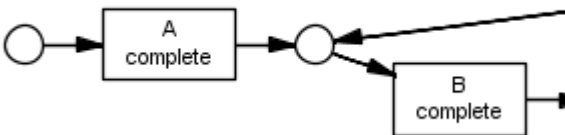
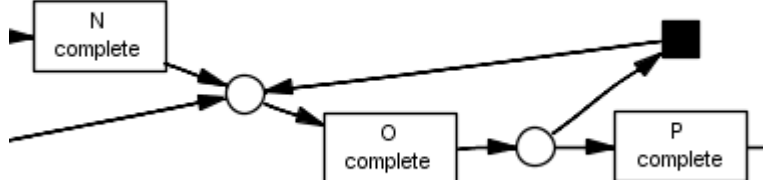
Table 1  
Weakness of Alpha

Weakness of Alpha	Result
Length-one-loop	<p>The diagram shows a Petri net with a transition labeled 't:1'. An arrow points from 't:1' to a place labeled 'A complete'. Another arrow points from 'A complete' back to 't:1', forming a loop. Additionally, there are two other places: 'P complete' and 'Q complete'. An arrow points from 'P complete' to a transition, and an arrow points from that transition to 'Q complete'. Another arrow points from 'Q complete' back to the transition. There is also a place 'B complete' with an arrow pointing to a transition, and an arrow from that transition to 'C complete'. Another arrow points from 'C complete' back to the transition, and an arrow from 'D complete' also points to the transition.</p>
Length-two-loop	<p>The diagram shows a Petri net with a transition labeled 't:1'. An arrow points from 't:1' to a place labeled 'A complete'. Another arrow points from 'A complete' back to 't:1', forming a loop. There are also places for 'O complete', 'N complete', 'P complete', and 'Q complete'. An arrow points from 'P complete' to a transition, and an arrow points from that transition to 'Q complete'. Another arrow points from 'Q complete' back to the transition. There are also arrows from 'O complete' and 'N complete' pointing to the transition.</p>

### 3.3 Process Discovery Using Heuristic Miner

The Heuristic Miner is one basic improvement of the Alpha algorithm [14] This algorithm focuses on the frequency of the process. The process that happens frequently will be more used compared to the less frequent process. It means that this algorithm neglects processes with low frequency. So that this algorithm can be effectively used to describe processes that recurring or frequently happened. Process model generated using Heuristic Miner is shown in Table 2.

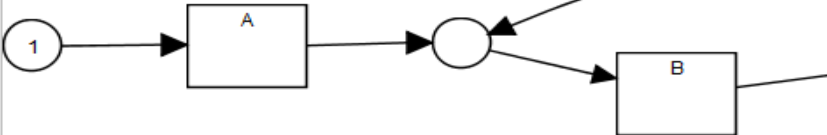
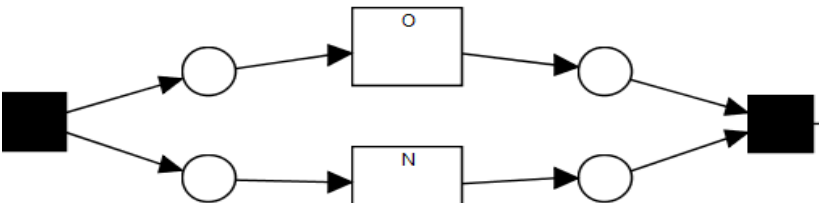
**Table 2**  
Process model using Heuristic Miner

Heuristic	Result
Model Process using Heuristic Miner in process A	
Model Process using Heuristic Miner in process O	

### 3.4 Process Discovery Using Inductive Miner

The Inductive Miner has dual approach working principles [10]. It focuses on both begin and end of the whole process and then makes grouping of processes. The tracing is also made within the group. This algorithm can describe more detail compared to Heuristic Miner but still consider less frequent processes by making it sometimes like a parallel process. Process model generated using Inductive Miner is shown in Table 3.

**Table 3**  
Process Model using Inductive Miner

Weakness	Result
Model Process using Inductive Miner in process A	
Model Process using Inductive Miner in process O	

The result of fitness and precision for Alpha, Heuristic Miner, and Inductive Miner algorithm is shown in Table 4 Alpha has highest fitness compared to other two algorithm. It means that Alpha algorithm can describe process in a very good manner. On the other hand, its precision is very bad at 0.25. The highest precision is achieved by Inductive Miner due to its dual approach characteristics. It made higher precision, but the fitness value is reduced just a small number.

**Table 4**  
Fitness and Precision each algorithm

	Fitness	Precision
Alpha	0.96	0.25
Heuristic Miner	0.92	0.62
Inductive Miner	0.94	0.89

#### 4. Conclusion

Process discovery in process mining is used to create process model automatically based on event log. Various process discovery methods have been introduced which are the Alpha algorithm, Heuristic Miner, and Inductive Miner. The Alpha algorithm focuses on causality each activity. Heuristic Miner focuses on the frequency of the process. Inductive miner focuses on both begin and end of the whole process and then makes grouping of processes. Fitness is done by measuring the fit between the event log and the process model. Precision measures the accuracy of an algorithm used for a case. In this study, the Alpha algorithm has a higher fitness of 0.96 which indicates that the process model formed is close to reality. However, the point in this research is that the Alpha, Heuristic, and Inductive Miner algorithms each have characteristics that are seen in the case internship program.

#### 5. References

- [1] Y. Amelia Effendi and R. Sarno, "Conformance Checking Evaluation of Process Discovery Using Modified Alpha++ Miner Algorithm," Proc. - 2018 Int. Semin. Appl. Technol. Inf. Commun. Creat. Technol. Hum. Life, iSemantic 2018, pp. 435–440, 2018, doi: 10.1109/ISEMANTIC.2018.8549770.
- [2] F. M. Maggi, C. Di Francescomarino, M. Dumas, and C. Ghidini, "Predictive Monitoring of Business Processes," 2013, [Online]. Available: <http://arxiv.org/abs/1312.4874>.
- [3] S. Huda, R. Sarno, and T. Ahmad, "Increasing accuracy of process-based fraud detection using a behavior model," Int. J. Softw. Eng. its Appl., vol. 10, no. 5, pp. 175–188, 2016, doi: 10.14257/ijseia.2016.10.5.16.
- [4] A. Awad, G. Decker, and M. Weske, "Efficient compliance checking using BPMN-Q and temporal logic," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 5240 LNCS, pp. 326–341, 2008, doi: 10.1007/978-3-540-85758-7-24.
- [5] E. Files and D. S. Models, Discovering Simulation Models - Tutorial, vol. 34, no. 3. 2010.
- [6] A. Rozinat, R. S. Mans, M. Song, and W. M. P. van der Aalst, "Discovering colored Petri nets from event logs," Int. J. Softw. Tools Technol. Transf., vol. 10, no. 1, 2008, doi: 10.1007/s10009-007-0051-0.
- [7] M. Song and W. M. P. van der Aalst, "Towards comprehensive support for organizational mining," Decis. Support Syst., vol. 46, no. 1, pp. 300–317, 2008, doi: 10.1016/j.dss.2008.07.002.
- [8] D. Rahmawati, M. A. Yaqin, and R. Sarno, "Fraud detection on event logs of goods and services procurement business process using Heuristics Miner algorithm," in 2016 International Conference on Information Communication Technology and Systems (ICTS), 2016, pp. 249–254, doi: 10.1109/ICTS.2016.7910307.
- [9] Y. A. Effendi and R. Sarno, "Discovering process model from event logs by considering overlapping rules," Int. Conf. Electr. Eng. Comput. Sci. Informatics, vol. 2017-Decem, no. September, pp. 19–21, 2017, doi: 10.1109/EECSI.2017.8239193.
- [10] R. Ghawi, "Process Discovery using Inductive Miner and Decomposition," no. October, 2016, [Online]. Available: <http://arxiv.org/abs/1610.07989>.
- [11] A. K. Alves de Medeiros, A. J. M. M. Weijters, and W. M. P. van der Aalst, "Genetic process mining: A basic approach and its challenges," Bus. Process Manag. Work. (BPM 2005), vol. 3812, no. task C, pp. 203–215, 2006, doi: 10.1007/11678564\_18.

- [12] R. Sarno, Kartini, W. A. Wibowo, and A. Solichaha, “Time based Discovery of parallel business processes,” *Proceeding - 2015 Int. Conf. Comput. Control. Informatics Its Appl. Emerg. Trends Era Internet Things, IC3INA 2015*, pp. 28–33, 2016, doi: 10.1109/IC3INA.2015.7377741.
- [13] D. Informatika, F. Teknik, U. T. Madura, and D. Informatika, “a More Efficient Deterministic Algorithm in Process,” vol. 14, no. 3, pp. 971–995, 2018.
- [14] R. Sarno, Y. A. Effendi, and F. Haryadita, “Modified Time-Based Heuristics Miner for Parallel Business Processes,” *IRECOS (International Rev. Comput. Software)*, vol. 11, no. 3, pp. 249–260, 2016, doi: <https://doi.org/10.15866/irecos.v11i3.8717>.

