Accredited Ranking SINTA 2
Decree of the Director General of Higher Education, Research and Technology, No. 158/E/KPT/2021
Validity period from Volume 5 Number 2 of 2021 to Volume 10 Number 1 of 2026

Published online on: **http://jurnal.iaii.or.id**

JURNAL RESTI
(Rekayasa Sistem dan Teknologi Informasi)
Vol. 7 No. 5 (2023) 1233 - 1238 ISSN Media Electronic: 2580-0760

Process Mining on New Student Admission Process in Telkom University using Genetic Miner

Supra Yogi[1], Angelina Prima Kurniati[2], Ichwanul Muslim Karo Karo[3]
[1,2]School of Computing, Telkom University, Bandung Indonesia
[3]Computer Science Departement, Medan State University, Medan Indonesia
[1]yogisupra@student.telkomuniversity.ac.id, [2]angelina@telkomuniversity.ac.id, [3]ichwanul@unimed.ac.id

Abstract

The selection process for new students at Telkom University, also known as SMB Telkom University has been running for years and already has its process flow. However, the existing process flow can be further improved to better reflect the actual field processes and become more accurate. Process mining can enhance this process flow by creating a new process flow based on event logs or previously executed processes. One of the algorithms in process mining is genetic process mining, where process mining is performed multiple times over several generations and genetic algorithms such as crossover and mutation are applied to generate a more accurate process model compared to other process mining algorithms such as heuristic and inductive mining. After conducting experiments, the best process model that was produced was at the 100th generation which has a fitness point of 0.755910819 and precision point of 0.742857143, after examining the parameters and the resulting Petri net or process flow that was produced it was concluded that the process model obtained from the application of Genetic Process Mining to SMB Telkom University is not very good because the resulting Petri net has several duplicate activities and appears to be non-linear. This could be due to several factors i.e., incompatible, or inaccurate data.

Keywords: *new student admission; telkom university; process mining; genetic process mining*

1. Introduction

Telkom University's New Student Admission (SMB) is a registration process to become a Telkom University student, which takes place every year at the beginning of the odd semester. There are several SMB pathways at Telkom University, such as the Jalur Prestasi Akademik (Academic Achievement Path), Undangan Seleksi Mitra (Partnered Selection Invitation), Ujian Tulis Gelombang (Written Exam Batch), Jalur Beasiswa (Scholarship Pathway), and many more. Among all the SMB pathways, the Jalur Prestasi Akademik (JPA) and the Ujian Tulis Gelombang (UTG) have the highest number of applicants each year. This is because both pathways are regular pathways that can be followed by all high school graduates [1].

The SMB process conducted by Telkom University has long used an online registration system via smb.telkomuniversity.ac.id, eliminating the need for applicants to come to the campus in person. Although the online registration system has made the process easier, there are still several stages in the JPA and UTG registration process that can be streamlined since applicants can register for more than one pathway. To simplify this registration process, an evaluation of the current SMB process model needs to be conducted.

Among these pathways, the two most popular and widely chosen pathways each year are the Academic Achievement Pathway (JPA), where applicants are selected based on their academic records in vocational school/high school, and the Written Exam Batch (UTG), which is a pathway that assesses applicants' abilities through a written test [1].

Process Mining is a Data Mining field that focuses on learning and improving existing process models, either by enhancing the current process model or creating a new process model using event logs or recorded activity data[2].

Since the SMB Telkom University is a process that produces an event log, process mining can be applied to the SMB Telkom University process[3].

Genetic Miner is one of the process mining algorithms that utilizes genetic algorithms[4] to model process models based on the execution event log of the existing process model. This algorithm initially creates an initial population of process models and then searches for individuals with the highest fitness. These individuals

Accepted: 05-07-2023 | Received in revised: 22-10-2023 | Published: 28-10-2023

undergo crossover or mutation, resulting in new individuals, until a generation with high fitness and an optimized process model is achieved. The Genetic Miner algorithm aims to solve problems in process models such as duplicate activities, hidden activities, noise, and incomplete data[5].

An example of one of the SMB registration pathways at Telkom University is the UTG pathway, which is commonly chosen by participants who wish to apply to Telkom University. This can be seen in Figure 1.

Figure 1. New Student Admission Steps[1]

Another example of the path for SMB registration is Ujian Tulis Gelombang (Written Exam Batch) that have the following stages [1]: 1. Participants create an account. 2. Participants choose what major they want to be in, 3. Participants order the PIN to pay the admissions fee, 4. Participants paid for said PIN, 5. Participants verify their paid PIN, 6 Participants print their participant card, 7. Participants attend the written exam and 8. Participants wait for their exam results.

2. Research Methods

Process mining is a field that combines data science with process management, focusing on analyzing process models based on event logs. The main goal of process mining is to improve processes using the available data[6].

Process mining consists of three types: discovery, conformance checking, and enhancement[2]. Discovery is a process mining technique where a process model is created based on the available event logs[7]. Conformance checking involves comparing the process model generated from the event logs with the previously used process model[8]. Enhancement refers to improving the existing process model using an ideal process model derived from reading the event logs[9].

One of the objectives of process mining is to automatically create a process model that depicts the activities performed in the event logs. This process model illustrates how a process should be conducted based on the recorded data of its execution[10].

Although simple process mining algorithms like the alpha algorithm can create efficient process models, they still have some limitations. These limitations include the inability to handle hidden activities, noise (input errors in the event logs), and duplicate activities[5].

The general steps in process mining are as follows[11]: First, prepare the dataset for process mining. The dataset needs to be cleaned from empty data and then formatted into an event log. Once the event log is ready, process discovery is applied[12], which outputs a process model that describes the flow of activities from the event log. This process model is then evaluated by executing events on the process model[8] and measuring the trace fitness and precision of the process model against the event log[9].

2.1 Genetic Process Mining

To overcome the limitations of traditional process mining algorithms like the alpha algorithm, Genetic Miner is used. Genetic Miner is a technique in process mining that utilizes genetic algorithms[4]. By using this algorithm, the resulting models can handle noise, hidden activities, and duplicate activities [2].

An overview of the stages of genetic process mining can be seen in Figure 2.

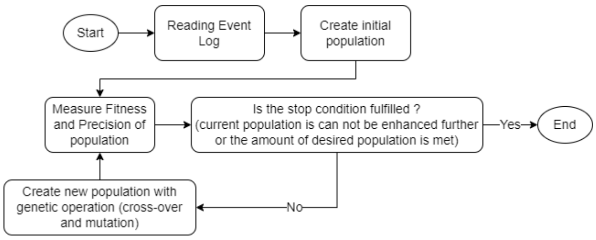

Figure 2. Genetic Mining Stages[5]

The first step in genetic mining is the creation of the initial generation. In this study, the initial generation consists of process models generated by applying the inductive miner. The goal is for the genetic miner to improve the quality of the basic process models produced by the inductive miner[5].

Next, two individuals or the best process models are selected based on their fitness to become parents for the next generation. After selecting the two parents, genetic operations, namely crossover and mutation[4], are performed. Crossover involves randomly exchanging some elements between the two parents, while mutation involves changing one element of a parent. This process is repeated until the desired population size is reached, creating a new population.

The steps of selecting parents, performing genetic operations, and creating a new population are repeated until there is no change in the parents of the population or a stopping parameter is reached. For example, this process can be performed for 100 generations. An example of the crossover genetic operation in process mining can be seen in Figure 3.

DOI: https://doi.org/10.29207/resti.v7i5.5341

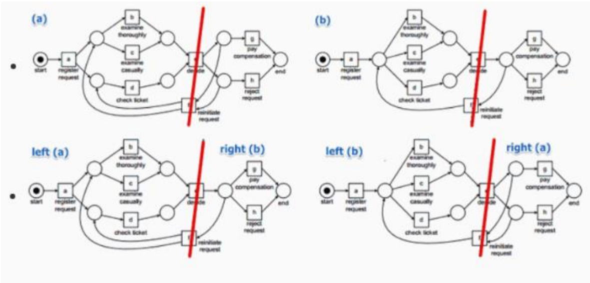

Figure 3. Cross-over Operation in Genetic Mining[5]

2.2 Conformance Checking

2.2.1 Trace Fitness Measurement

In process mining, several metrics can measure the quality of the process model generated by the process mining steps. One of them is fitness, which is a metric that describes the conformity of the process model with the available event log.[8], [13], [14]

The calculation of fitness can be outlined as follows: it compares the number of correct executions performed by the process model when executed with the event log to the total number of executions in the event log. If the fitness value approaches 1, the model can be considered to represent the event log well[8].

All generated process models will be evaluated for their fitness. Firstly, the fitness within one case from the log is calculated using Formula 1.

$$Fitness(\sigma, N) = \frac{1}{2}\left(1 - \frac{m}{c}\right) + \frac{1}{2}\left(1 - \frac{r}{p}\right) \quad (1)$$

m is the number of missing tokens, c is the number of used tokens r is the number of tokens left even though output is reached and p is the amount of produced tokens[7].

After that, the fitness of the process model is calculated by combining the fitness values of each case in the log, using Formula 2.

$$Fitness(L, N) = \frac{1}{2}\left(1 - \frac{\sum_{\sigma \epsilon L} L(\sigma) \times m_{N,\sigma}}{\sum_{\sigma \epsilon L} L(\sigma) \times c_{N,\sigma}}\right) + \frac{1}{2}\left(1 - \frac{\sum_{\sigma \epsilon L} L(\sigma) \times r_{N,\sigma}}{\sum_{\sigma \epsilon L} L(\sigma) \times p_{N,\sigma}}\right) \quad (2)$$

$L(\sigma)$ is the amount of trace σ in log L.

2.2.2 Precision Measurement

One other metric to measure the quality of the generated process model is precision. Precision is a measure of how well the process model can predict the execution outcomes from the event log, or it can be referred to as the accuracy of the generated process model[8], [13], [14].

In essence, the calculation of precision compares the successfully predicted process executions to the total predictions made by the process model[8]. Similar to fitness, as the precision value approaches 1, the better the generated process model.

The calculation of precision can be summarized in Formula 3.

$$Prec(L, N) = \frac{TP}{TP + FN + FP} \quad (3)$$

TP (True Positives) represents the number of correctly aligned activities between the observed process and the model. FP (False Positives) represents the number of activities that are aligned in the observed process but are not present in the model. FN (False Negatives) represents the number of activities that are expected to be present in the model but are not aligned with the observed process[8].

3. Results and Discussions

3.1. Data Structure

The data used is the Telkom University New Student Admission (SMB) data from the year 2021. The data consists of 6 columns. The "user" column contains a unique ID that identifies the applicant and is used to track the activities performed by the applicant. The "tanggal_aktivitas" column contains the date and time of the activities performed by the user, which is used to identify the stages of the process. The "aktivitas" column contains the specific activity performed by the user on the given date and time. Finally, there are the "keterangan 1" to "keterangan 3" columns, which are additional columns used to provide further details about an activity, such as whether the applicant passed or failed the "cek_kelulusan" activity. List of Activities at SMB Telkom University can be seen in Table 1.

Table 1. List of Activities at SMB Telkom University

Activities	Description
Registration	Activities where participants create new accounts for SMB
Account Activation	Participant activate their newly created account by verifying their email
PIN request	Participant request new PINs to pay their admission fee
PIN payment	Participants pay for their admission fee
PIN verification	Participants verify their PIN with a code that shows up after payment
Participant number activation	Participant activate their participant number to verify they are registered
Report upload	Participants upload their high school report card
Program selection	Participants select what program or major they want
Exam location selection	For participants that choose the Exam path here, they choose where they want to attend the exam
Participant card printing	Before participants attend the written exam they first have to print their participant card to verify their identity
Taking the exam	Participant from the exam path attend their exam
Checking result	Participants check their SMB result

3.2. Data Preprocessing

The received data for the research initially was raw and not suitable for application because there were several aspects of the data that needed to be cleaned and formatted according to the algorithm to be used[15] - [20].

Firstly, data cleansing was performed to remove data without values. This data had several rows without ID or activity dates, so those rows had to be deleted as they lacked the necessary information. After this preprocessing step, the number of rows reduced from 1,300,690 to 1,219,650.

Furthermore, to avoid any issues during program execution, the activity dates were standardized to the same format. An example of the cleaned data can be seen in Figure 4.

	iduser	tanggal_aktivitas	aktivitas	keterangan_1	keterangan_2	keterangan_3
0	270971881	2020-09-17 12:09:00	buat_akun	-	-	-
1	270971882	2020-09-17 12:09:00	buat_akun	-	-	-
2	270971883	2020-09-17 12:09:00	buat_akun	-	-	-
3	270971884	2020-09-17 12:10:00	buat_akun	-	-	-
4	270971885	2020-09-17 12:16:00	buat_akun	-	-	-
...	...	...	...	...	...	...
1048570	484516268	2021-06-29 15:01:00	upload_raport	upload_raport	upload_raport	upload_raport
1048571	484516155	2021-06-29 13:02:00	upload_raport	upload_raport	upload_raport	upload_raport
1048572	484510620	2021-06-29 13:07:00	upload_raport	upload_raport	upload_raport	upload_raport
1048573	484507509	2021-06-29 15:01:00	upload_raport	upload_raport	upload_raport	upload_raport
1048574	484516029	2021-06-29 13:10:00	upload_raport	upload_raport	upload_raport	upload_raport

Figure 4. Data Sample

Secondly, a trace event log was generated using ProM tools. This tool provides an overview of the steps performed by each user for each case, as shown in Figure 4. In the trace result, it can be observed that some cases had repeated activities, such as uploading a report. To optimize the results for the genetic miner, these repeated activities were merged using the "Merge subsequent events" technique, which combines all repeated activities into a single activity.

Thirdly, the trace event log also revealed that 43% of the applicants only performed the account creation and activation steps. Since these activities are considered incomplete in the New Student Admission process and represent nearly half of the data, they were removed to ensure that the process mining analysis is not affected.

Lastly, the "START" and "END" conditions were added to the beginning and end of each activity to indicate the start and end points of the activities performed. After the preprocessing steps were completed, the data, which is now clean and can be seen in Figure 5, can be used for the next stage of analysis.

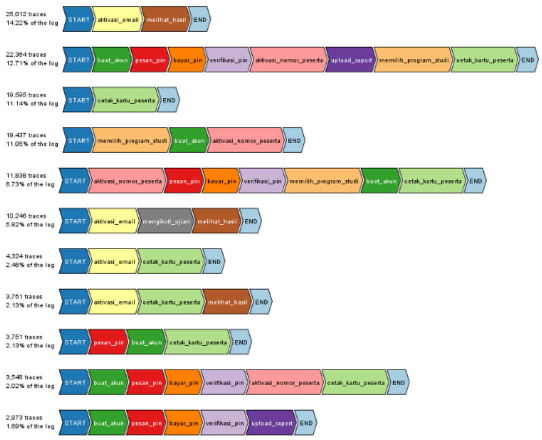

Figure 5. Event Log Sample

3.3 Genetic Mining Result

The parameters used for the genetic miner are as follows: population size of 20, elite count of 5, and crossover and mutation probabilities of 25%. The testing is performed by varying the number of generations. The models are created 10 times with the stopping parameter set to different numbers of generations, ranging from 50 generations to 150 generations, with an increment of 10 generations each time. The fitness and precision values generated by each generation can be seen in Table 2.

Table 2. Result of Fitness and Precision Measurement

Generation	Fitness	Precision
50	0.709447005	0.7
60	0.713701788	0.727272727
70	0.720682723	0.655172414
80	0.690683123	0.677419355
90	0.693407398	0.830508475
100	0. 755910819	0.742857143
110	0.734470529	0.666666667
120	0.722620443	0.857142857
130	0.743254572	0.898876404
140	0.705672749	0.803921569
150	0.750353577	0.84

As seen in Table 2 process model at the 100^{th} generation has the best trace fitness value at 0.755910819 this means that the process model at the 100^{th} generation at least has 75% of traces of activities correctly traced, which means the process model is not overly fitted as the fitness value does not reach 1 and can be considered moderately fit for the event log because the fitness value is at least more than 0.5.

On the other hand, if the goal is to obtain the best precision model, then the model in the 130th generation is considered the best with a precision value of 0.898876404. This means that 89% of the steps identified by the process model are correct and align with the actual activities that take place in the process. In other words, the process model provides an accurate representation in terms of recognizing specific steps or activities within the process.

3.4 Petri Net Result

After evaluation, the process model in the 100th generation is the best in terms of fitness. The resulting process model can be seen in the Petri net in Figure 6.

The resulting Petri net can depict the process model of SMB Telkom University as follows: 1. SMB process starts. 2. Participants create an account, or if they already have an account provided by the university, they can proceed to print the participant form or pay and verify the PIN. Activities that do not involve creating an account are likely for scholarship recipients or participants from the Academic Achievement Path. 3. Participants order the PIN. 4. Participants pay for the ordered PIN. If the participant is a scholarship recipient or from the Academic Achievement Path and has

DOI: https://doi.org/10.29207/resti.v7i5.5341

already paid for the PIN, they can proceed to select a study program or print their participant card.

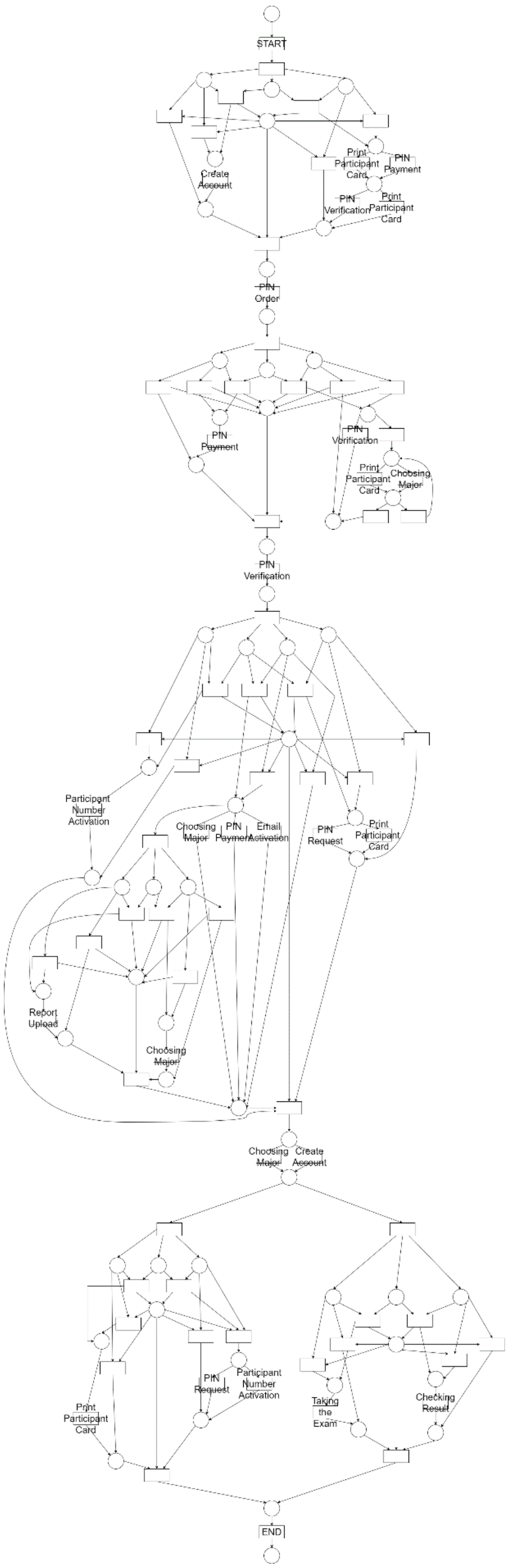

Figure 6 Petri Net Result

5. After PIN verification, several subsequent activities can be performed. Participants can activate their participant number, upload their report, print their participant card, and choose a study program if they haven't done so in the previous steps. 6. Participants choose a study program if they haven't done so in the previous steps. 7. If the participant is from the Written Exam Track, the final step is to take the exam and view the results. Participants from other admission tracks who have been accepted can directly print their participant cards. 8. SMB Telkom University process is completed.

4. Conclusion

Based on the result of the research that has been done, the best process model was obtained in the 100th generation with a fitness replay of 0.755910819 and the 130th generation with a precision of 0.898876404. By examining the resulting Petri net, it can be observed that there are many repeated processes or a non-linear process model. This can be attributed to various factors, and the most likely reason is that the dataset may not be suitable for process mining.

Because these studies combine all participants' activities regardless of their registration pathways, this means that some pathways that have different activities such as taking exams or uploading report cards in advance get mixed up, although genetic mining is supposed to be unaffected by this problem, the second problem that comes to mind is the dataset is not built specifically for process mining, this means that for example there is no timestamp for when participant start their activities and when they stop.

Nevertheless, by comparing the description of the SMB process generated by process mining with the existing SMB Telkom University guidelines, it can be seen that the SMB process derived from genetic process mining, based on the collected data, closely approximates the existing steps.

In conclusion, although the dataset used for this study can be utilized for process mining, it can still be further improved for future research. For example, adding timestamps indicating when activities start and end, incorporating additional timestamps for repeated activities, and creating a dedicated database to store data used for process mining. These improvements would enhance the quality of future process mining analyses and potentially contribute to improving the SMB process at Telkom University in the future.

Acknowledgment

The authors would like to thank Telkom University, especially the PADMI division, for providing the dataset and support in this research.

References

[1] Telkom University, “Jalur Seleksi.” https://smb.telkomuniversity.ac.id/jalur-seleksi/.
[2] W. M. P. Van Der Aalst, “A practitioner’s guide to process

mining: Limitations of the directly-follows graph," *Procedia Comput. Sci.*, vol. 164, pp. 321–328, 2019, doi: 10.1016/J.PROCS.2019.12.189.

[3] L. Juhaňák, J. Zounek, and L. Rohlíková, "Using process mining to analyze students' quiz-taking behavior patterns in a learning management system," *Comput. Human Behav.*, vol. 92, pp. 496–506, 2019, doi: 10.1016/J.CHB.2017.12.015.

[4] S. Mirjalili, "Genetic Algorithm," in *Evolutionary Algorithms and Neural Networks: Theory and Applications*, Springer International Publishing, 2019, pp. 43–55. doi: 10.1007/978-3-319-93025-1_4.

[5] A. K. A. de Medeiros, A. J. M. M. Weijters, and W. M. P. van der Aalst, "Genetic process mining: an experimental evaluation," *Data Min. Knowl. Discov.*, vol. 14, no. 2, pp. 245–304, 2007, doi: 10.1007/s10618-006-0061-7.

[6] C. dos Santos Garcia *et al.*, "Process mining techniques and applications – A systematic mapping study," *Expert Syst. Appl.*, vol. 133, pp. 260–295, 2019, doi: 10.1016/J.ESWA.2019.05.003.

[7] F. Mannhardt, "Multi-perspective process mining.", 2018

[8] J. Carmona, B. van Dongen, and M. Weidlich, "Conformance Checking: Foundations, Milestones and Challenges," *Lect. Notes Bus. Inf. Process.*, vol. 448, pp. 155–190, 2022, doi: 10.1007/978-3-031-08848-3_5.

[9] M. de Leoni, "Foundations of Process Enhancement," *Lect. Notes Bus. Inf. Process.*, vol. 448, pp. 243–273, 2022, doi: 10.1007/978-3-031-08848-3_8.

[10] S. M. Hejazi, M. Zandieh, and M. Mirmozaffari, "A multi-objective medical process mining model using event log and causal matrix," *Healthc. Anal.*, vol. 3, 2023, doi: 10.1016/J.HEALTH.2023.100188.

[11] W. M. P. van der Aalst, "Foundations of Process Discovery," *Lect. Notes Bus. Inf. Process.*, vol. 448, pp. 37–75, 2022, doi: 10.1007/978-3-031-08848-3_2.

[12] A. Augusto, J. Carmona, and E. Verbeek, "Advanced Process Discovery Techniques," *Lect. Notes Bus. Inf. Process.*, vol. 448, pp. 76–107, 2022, doi: 10.1007/978-3-031-08848-3_3.

[13] S. J. J. Leemans, D. Fahland, and W. M. P. van der Aalst, "Scalable process discovery and conformance checking," *Softw. Syst. Model.*, vol. 17, no. 2, pp. 599–631, 2018, doi: 10.1007/S10270-016-0545-X.

[14] J. Carmona, B. van Dongen, A. Solti, and M. Weidlich, "Conformance checking: Relating processes and models, relating processes and models," *Conform. Checking Relat. Process. Model. Relat. Process. Model.*, pp. 1–277, 2018, doi: 10.1007/978-3-319-99414-7.

[15] H. M. Marin-Castro and E. Tello-Leal, "Event Log Preprocessing for Process Mining: A Review," *Appl. Sci.*, vol. 11, no. 22, 2021, doi: 10.3390/app112210556.

[16] van der Aalst, Wil, "Data Science in Action," *Process Mining.*, 2016, doi: 10.1007/978-3-662-49851-4_1

[17] Benatallah, B., Sakr, S., Grigori, D., Motahari-Nezhad, H. R., Barukh, M. C., Gater, A., and Ryu, S. H., "Process analytics: concepts and techniques for querying and analyzing process data.", 2016

[18] M. Montali, A. Senderovich, M. Weidlich, "Process Mining Workshops", *ICPM 2022 International Workshops.*, 2023,

[19] W. M. P. van der Aalst, and J. Carmona, "Process Mining Handbook", 2022.

[20] M. Dumas, and M. Fantinato, "Business Process Management Workshops", *BPM 2016 International Workshops*, 2017.

DOI: https://doi.org/10.29207/resti.v7i5.5341