Turning Logs into Lumber: Preprocessing Tasks in Process Mining

Ying Liu[1], Vinicius Stein Dani[1], Iris Beerepoot[1], and Xixi Lu[1]

Utrecht University, Utrecht, the Netherlands
{v.steindani, i.m.beerepoot, x.lu}@uu.nl

Abstract. Event logs are invaluable for conducting process mining projects, offering insights into process improvement and data-driven decision-making. However, data quality issues affect the correctness and trustworthiness of these insights, making preprocessing tasks a necessity. Despite the recognized importance, the execution of preprocessing tasks remains ad-hoc, lacking support. This paper presents a systematic literature review that establishes a comprehensive repository of preprocessing tasks and their usage in case studies. We identify six high-level and 20 low-level preprocessing tasks in case studies. Log filtering, transformation, and abstraction are commonly used, while log enriching, integration, and reduction are less frequent. These results can be considered a first step in contributing to more structured, transparent event log preprocessing, enhancing process mining reliability.

Keywords: Log preprocessing · Process mining · Event log

1 Introduction

In the landscape of data-driven decision-making, event logs stand as invaluable assets, capturing the execution of activities of processes and their interactions within diverse operational systems. The potential insights that can be obtained from these logs are immense, spanning process improvement, anomaly detection, performance evaluation, and strategic planning [1]. However, the axiom "garbage in, garbage out" holds particularly true in this context [35]. The presence of data quality issues underscores the vital importance of preprocessing techniques. Without proper preprocessing, the very foundation of analysis is compromised.

The importance of data quality and preprocessing in the field of process mining has been acknowledged, as evidenced by the growing attention dedicated to these subjects [35, 39]. Despite the acknowledgment, the execution of log preprocessing seems to remain ad-hoc. Moreover, little support has been provided on which preprocessing tasks are possible and how to select them. Although a few process mining methodologies sketched potential preprocessing tasks, a comprehensive overview of these tasks has been notably absent. Furthermore, the way these preprocessing tasks are used in real-life has remained unclear.

Existing systematic literature reviews (SLRs) have attempted to tackle specific tasks of log preprocessing, such as event abstraction techniques [39] and data extraction [33].

However, a comprehensive review that covers diverse preprocessing tasks and their practical applications in real-world scenarios is lacking.

In this paper, we perform a systematic literature review to establish an initial, comprehensive overview of the preprocessing tasks and their utilization in process mining case studies. By undertaking this endeavor, we aim to create a repository of log preprocessing tasks that may provide guidance and support for researchers and practitioners.

We identified six high-level preprocessing tasks, and for four of these tasks, we observed 20 low-level preprocessing tasks described in the case studies. The results show that log filtering, transformation, and abstraction have been more frequently used in case studies, while log enriching, integration, and reduction (e.g., sampling) are much less frequently performed.

The remainder of this paper is organized as follows. In Section 2, we discuss related work. Next, we explain the methodology followed in Section 3 and present the results in Section 4. Finally, we conclude the paper in Section 5.

2 Related Work

In this section, we discuss the related work, based on which we synthesized an initial set of six high-level preprocessing tasks: (a) *log integration*, (b) *log transformation*, (c) *log reduction*, (d) *log abstraction*, (e) *log filtering*, and (f) *log enriching*, see Fig. 1.

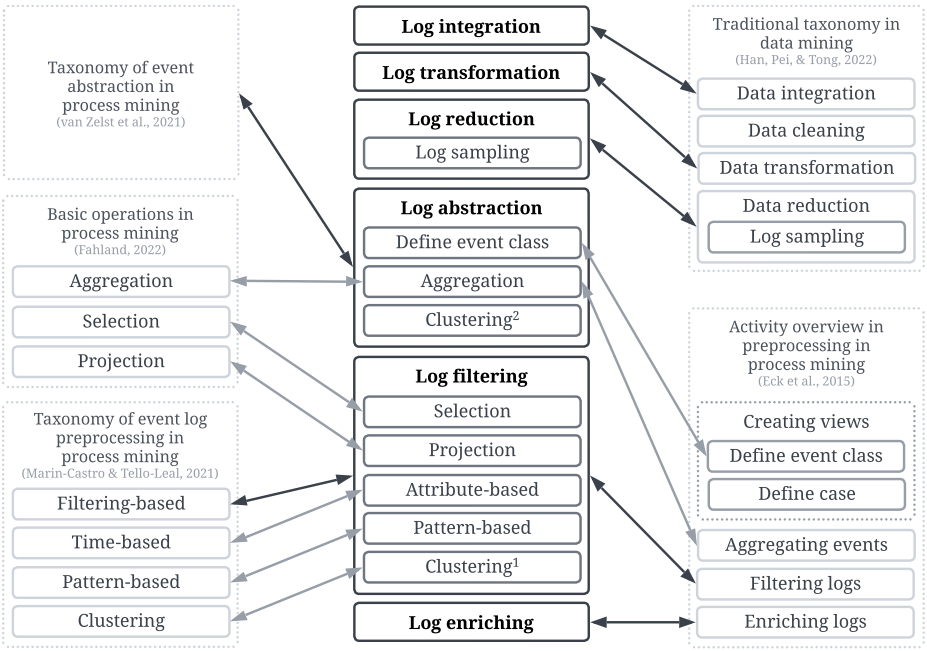

Fig. 1. Initial result of high-level log preprocessing tasks and techniques in the related work.

2.1 Taxonomy of Log Preprocessing Tasks

Han et al. [16] propose four categories of data preprocessing techniques: data cleaning, data integration, data transformation, and data reduction. Data cleaning focuses on handling missing values, identifying noise or outliers, and repairing errors. Since these subtasks are not interesting (e.g., identifying missing variable values) or not directly applicable to process mining (e.g., identifying noise/outliers in a distribution), we omit this task and decide to focus on the latter three tasks. For each, we create a corresponding log preprocessing task: *log integration*, *log transformation*, and *log reduction*.

Van Eck et al. [10] listed four tasks in the preprocessing stage, which are specifically tailored towards event logs: creating views, aggregating events, filtering logs, and enriching logs. We exclude the task "creating views" because this task assumes that there is no event log yet, while we assume we have a raw event log as input. We match the task "aggregating events" to *log abstraction*, also known as *event abstraction* that has been already surveyed [39]. The filtering of event logs ("filtering logs") is also considered a log preprocessing task within our scope, which we refer to as *log filtering*. Finally, the preprocessing task "enriching logs" is mapped to *log enriching*. As for *log enriching* and *log integration*, we consider *log integration* as creating a new event log by integrating one or more external data sources, while *log enriching* focuses on using the information within the event log to derive additional attributes.

Fahland [12] indicated that there are three basic preprocessing operations on event logs, which are: selection, projection, and aggregation. We consider the "selection" and "projection" as a part of the *log filtering* task, while the aggregation operation is considered as part of the *log abstraction*.

Regarding *log filtering*, *log abstraction*, and *log reduction*, both *log filtering* and *log abstraction* can reduce the size of the logs, but we consider the following subtle differences in comparison to log reduction here. *Log filtering* tends to focus on the quality issues of the original data. It obtains higher-quality logs by filtering out incorrect, incomplete, inconsistent, and irrelevant data. *Log abstraction* focuses on the complexity and granularity of the original data. It groups the events through aggregation, defining event classes, and clustering to reduce the complexity of logs. *Log reduction* is due to the data volume of the original data. It reduces the amount of data processed in a single analysis by random sampling, dividing, or cutting, but still makes the data representative.

2.2 Literature Review in Event Log Preprocessing

To the best of our knowledge, there is only one literature review focusing on the log preprocessing tasks: Marin-Castro and Tello-Leal [23] reviewed 70 related papers that were published from the years 2005 to 2020 and explicitly mentioned event log preprocessing or cleaning. This literature review grouped preprocessing techniques into two types of *techniques*: transformation techniques and detection-visualization techniques. *Transformation techniques* mark modifications made toward the original structure of the event log, while the events or traces that can lead to issues with data quality are identified, grouped, and isolated using *detection-visualization techniques*. In

this paper, we cover six high-level *preprocessing tasks*, instead of the techniques. We include log enriching, log integration, and log reduction, which have not been discussed.

Van Zelst et al. [39] conducted a review and presented a taxonomy of event abstraction techniques. While valuable and detailed insights are provided into the event abstraction techniques, no insights are provided into their usage in practice, and no overview is provided for other preprocessing tasks. Similarly, Stein Dani et al. [33] report that preprocessing, on a high level represented by filtering-related tasks, is still a manual effort in the event log preparation phase of a process mining project. However, they mainly focus on data extraction tasks and do not provide an overview of the preprocessing tasks, including automated ones, and their usage in real-life case studies.

Currently, there is no clear overview of log preprocessing tasks and how frequently are these preprocessing tasks being used in process mining projects. Using the six high-level tasks as our scope, we conduct an SLR in order to provide insights into the usage of log preprocessing techniques in process mining case studies.

3 Systematic Literature Review

To arrive at an initial selection of relevant papers, and inspired by Kitchenham and Petersen [19, 27], we applied the following search string on Scopus: (“process mining”) AND (“case study” OR “case studies”) within the article title, abstract, and keywords. As of December 20, 2022, we initially found 4565 papers. Fig. 2 shows an overview of the paper screening process we followed. Next, we applied the exclusion and inclusion criteria in order to narrow down the scope of the review. The following exclusion criteria were defined and applied directly via the search engine: (1) the paper is published in 2021 and 2022; (2) the paper is published in conferences or journals under peer-review; (3) the paper explicitly mentions “process mining” in the keywords; and, (4) the paper is written in English. As we are particularly interested in the current trend in case studies that use process mining as the core technique, this is our inclusion criteria. Therefore, only papers meeting these criteria were selected to be further analyzed.

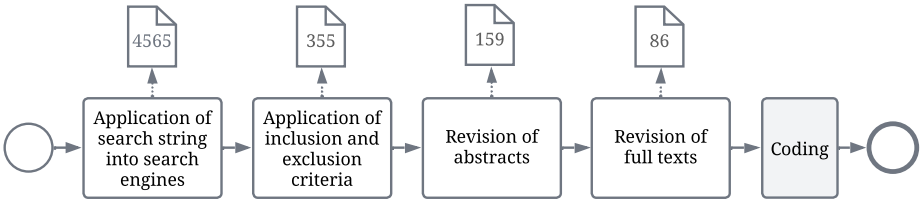

Fig. 2. Paper screening procedure.

After applying our exclusion and inclusion criteria, we obtained 355 papers. Because our focus is on log preprocessing tasks applied in real-world settings, we then read the abstracts of all these papers and filtered out the ones that did not mention collecting data from a real-world scenario. Thereafter, we obtained 159 papers to go through the full paper screening. These papers were downloaded and imported into the

software Nvivo[1] for further analysis. During the full paper analysis stage, the papers that did not mention any data preprocessing steps were discarded and, finally, 86 papers were obtained as relevant papers to go through the coding stage of our work.

The following codes were defined for the analysis: high-level category, low-level category, and data domain. Next, we discuss what each one of them entails. *High-level categories* were defined based on related work and used to deductively categorize the papers. The six high-level categories are (1) log integration, (2) log reduction, (3) log abstraction, (4) log filtering, (5) log enriching, and (6) log transformation. Several *Low-level categories* within the high-level categories were inductively defined from the studied papers. Finally, in addition, we also coded the *data domain* (e.g., healthcare, education, manufacture, etc.), the analysis purpose, the PM task, and the year. Due to space limits, we do not discuss these results in this paper. The initial result of the categorization process is presented in Fig. 1.

4 Results

In this section, we present the results of coding 86 papers. The results are discussed for each high-level category. A complete overview of the results and the detailed coding can be found online, see the Google sheet file. We also include the overview table listed in Appendix A [2] for the purpose of reviewing.

4.1 Log filtering

Log filtering is the most commonly performed preprocessing task, with 55 out of 86 papers performing this preprocessing task. These 55 papers mentioned filtering different objects, such as noise, outliers, redundant, duplicated cases and events, missing values, useless values, blank values, irrelevant values, and so on. Using the objects mentioned in these papers, the category log filtering is subdivided into 9 detailed low-level categories.

Filtering irrelevant data We observed that 29 out of 55 papers mentioned filtering "irrelevant" data. After analyzing these papers, we define *irrelevant data* as *those resources, activities, attributes, events, and traces that are not relevant or not important for the specific analysis to be conducted*.

Whether the data is relevant to the analysis task seems to be mostly determined by experts or analysts based on their domain knowledge and analysis requirements. For example, in [4], the analysis only focused on the students who participated in the class (resource), so the events generated by other resources were defined as irrelevant data and filtered. In [15], the authors intended to analyze the activities of Ph.D. students and improve their journeys. So after a discussion by analysts and stakeholders, a filter is applied to retain the traces of full-time students who completed their Ph.D. and who withdrew (case status). The term *useless data* is also used in some of the papers to describe irrelevant data. For example, in [36], the authors mentioned "*filtering useless information such as links and marker symbols*", since the links and marker symbols

[1] https://lumivero.com/products/nvivo/

[2] The Appendix will be published via Zenodo (https://zenodo.org/)

(attributes) cannot make any contribution to the intended analysis and are regarded as useless data.

Filtering incomplete data In 16 out of 55 papers, the authors mentioned filtering incomplete data. Incomplete data can be divided into *incomplete events* and *incomplete cases*. *Incomplete events* usually refer to events having missing values or missing attributes. Incomplete events include missing case id [20], missing timestamps [9, 15, 26], and missing activities [7, 9], missing other attribute values that are relevant to this analysis [14].

The incompleteness of a case is usually described as cases that are not completed or do not represent the end-to-end process. It means that the cases lack some events, for example, "*remove any record that may create only one event per case as it will not depict the sequence of activities and hinder the performance analysis of the model*" [28] and "*removing cases that did not cover the whole steps*" [7].

Filtering infrequent data We use *infrequent data* to refer to the infrequent case variant. In 13 papers, the authors mentioned that they performed the infrequent case variants filtering as a preprocessing task. Filtering infrequent data is done to "*prevent the PM tool from returning incomprehensible or inaccurate results*" [31], and "*to improve the quality of results, and to avoid low precision and highly complex results*" [38].

Filtering inconsistent data A simple example of inconsistent data is that the values are recorded in different formats, e.g., "2023-01-01" and "2023/01/01" as the attribute timestamps. This inconsistency in data format may be due to recording errors or caused by manual input. It may also be that different data sources have used different data formats. Inconsistent event labels make it difficult to assign clear semantics to the activities of a discovered process model [1], and may also bring about a dimensional explosion of the process model.

Filtering incorrect data Incorrect data is erroneous or unreliable data that violates the logic of reality. For example, in the real process, activity A should be executed earlier than activity B, but in the log, the timestamp of A in a specific case is later than activity B [26].

Filtering duplicates Duplicates refer to repeated data. In process mining, the case ID needs to be a unique identifier, and the traces represented by different case IDs must be different, so as to ensure the accuracy of the data. However, in real life, duplicate data is usually generated due to system bugs or other reasons. For example, in [8], repeated events with the same Call-ID were excluded.

Filtering redundant data Only two papers mentioned redundant data [6, 7]. In [7], redundant events were included in data error: "*we conducted some data preprocessing, including handling data error (e.g., removing redundant events and eliminating multiple yield values)*", while there was no further definition and explanation in [6]. We consider redundant events to be events that semantically are the same but not necessarily identical, while duplicates are identical events.

Filtering outliers In [2, 3, 22], the authors only mentioned "*removing outliers*" without any further explanation or definition. In [5], the authors mention "*we noticed the existence of outliers, i.e., cases that take too long, or incomplete*"; so, too long trace

and incomplete data are considered outliers. In [30], "*if lecture activities in the short semester are included, it will be an outlier because it has activities that are far more than short than activities in the semester in general*"; thus, traces that are too short are also considered outliers. It seems that process analysts use the distributions of a case or event-attribute to define outliers, e.g., the number of events per case, the case duration per case, etc.

Filtering noise Noise is an overused word. Data that is not conducive to the analysis task is often defined as noise. An interesting point is that among the 86 papers, more than one paper mentioned noise, but only one paper described what noise is and how to filter it, "*In the original log the noisy activities were conveniently named 'Noise', so they were removed using a filter on the activity name*" [21].

4.2 Log transformation

In 38 of the 86 papers, the authors described that they performed a *log transformation* task. The coding resulted in four data objects that are being transformed, which we use to further divide the high-level category.

Transforming format Among the format transformations, the transformation of the log format from CSV to XES was mentioned the most (14 out of 25 papers), such that the event logs can be used in the PM tool. This is because the log format after extraction is usually CSV, and PM tools require the log format to be imported as XES. The remaining format transformation is related to determining which columns are the key columns (such as case ID, activity name, and timestamps) after importing the log into PM tools.

Transforming values The difference between transforming values and transforming format is that transforming values means the change of one or more specific values in an event. For example, replacing infrequent values with the value 'other' to avoid dimension explosion, replacing missing values, replacing NaN values with 'zero', capturing data, and encrypting data.

Reordering Reordering is the process of sorting the log by a particular timestamp. When the original log is out of order, it is essential to reorder it so that the process model displays the activities' proper execution sequence.

Transition matrices and encoding In particular, transition matrices and trace encoding are used as a preprocessing for predictive process monitoring. Given that the trace encoding is a subfield itself and was not included in the search, we consider this category outside of our scope. We found two case studies mentioning this preprocessing task and coded them without further analysis.

4.3 Log enriching

In 16 out of 86 papers, the log-enriching techniques were applied. Log enriching is split into four categories. Three of them are shown using an example in Fig. 3.

Adding calculation metrics In this low-level category, the calculation metrics are computed from existing attributes in the log. For example, in [8], call center processes of

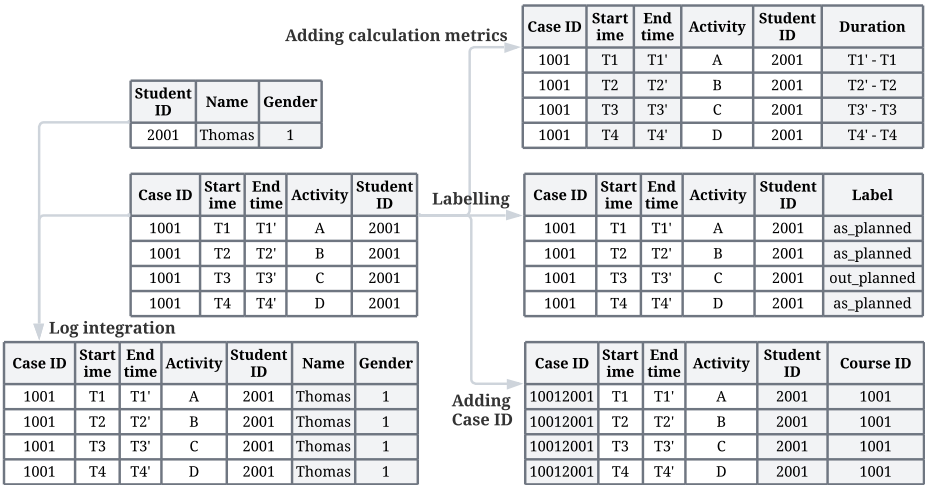

Fig. 3. Examples of three log integration tasks versus log enriching.

a company were examined. In the original event log, each call only had attributes Start and Call Duration, but process analysis required the end time of the call. Therefore, the attribute End was obtained by adding Call Duration to Start.

Labelling Labeling is the task of assigning a tag or a class to an event or a trace. In [37], "*the cases are labeled as either successful or failed, depending on how they have been executed and their outcome*", to further divide the log into two logs. In [25], for recording differences over time between the intended operation and the actual execution, a label was assigned to each event to indicate if the event was carried out on time or not.

Adding case id Case id is a unique attribute in event logs. The data collected in some case studies did not have the attribute of case id, then the case id was created artificially in the data preprocessing stage. For example, in [34], "*the caseid is created by combining the three-digit client number (MANDT) with a ten-digit document number and a five-digit item number*".

Adding noise Adding noise is not a typical preprocessing task, as just one publication described it. [32] evaluated privacy assurance of healthcare metadata. Noise-adding plugins in the tool ProM were used to make the original event logs more privacy-preserving [24].

4.4 Log reduction

In 11 out of 86 papers, the authors used log reduction to do log preprocessing. Examples of the three log reduction tasks are shown in Fig. 4.

Dividing into sub-logs In the example presented in Fig. 4, the original log is divided into two logs by the date in timestamp. In [21, 11], IoT logs were collected in a smart house and the aim was to explore human habits. They firstly divided logs into smaller

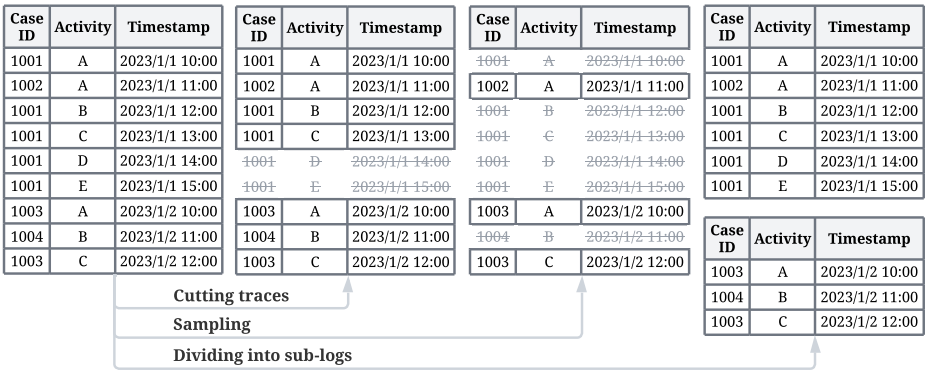

Fig. 4. A simple example of log reduction.

pieces by timestamps to analyse the time distribution of the activities (user habits) within a day [21].

Resource could also be a common attribute for division. The authors of [29] divided the traces into subsets to model different profiles of users. Dividing original logs according to specific attributes is usually for more in-depth analysis [15].

In addition, in order to test the proposed algorithm or approach, the log was divided into training data and test data according to a certain proportion [7, 17].

Sampling The most notable characteristic of sampling is randomness. The reduction here is to reduce the trace; that is to say, the data processing needs to be in the unit of a trace. In the example shown in Fig. 4, there are four traces $[\langle A, B, C, D, E\rangle, \langle A\rangle, \langle A, C\rangle, \langle B\rangle]$. After randomly sampling 50% of the traces, the log $[\langle A\rangle, \langle A, C\rangle]$ in the lower right corner is obtained.

Cutting traces In the example in Fig. 4, compared to other traces, the trace $\langle A, B, C, D, E\rangle$ is obviously longer and contains more events. Cutting off the event at the end of the trace will get the processed log in the lower left corner. The purpose of this technique is to avoid bias from very long traces [13].

4.5 Log integration

Among the 86 papers, 14 papers used log integration to combine multiple data tables. No objects of interest are repetitively mentioned, nor have we observed obvious low-level tasks. Therefore, the log integration task has not been further divided.

Fig. 3 shows an example where a new event log is created by matching two data tables using the shared attribute "*student_id*". It is worth mentioning that some papers mention that additional data was added to the original event data without indicating the source, but we believe that the combination of these data is realized by log integration. According to [18], "*Besides the attributes shown in Table 4, we included the educational level of the nurses executing the activity, as well as their nursing experience/organisational role, the hospital shift and weekday on which the activities were performed, and the ward in which the shift took place*". It is reasonable

to speculate that this additional information actually comes from a separate data table that stores information about all nurses.

4.6 Log abstraction

In 37 out of the 86 papers, the authors used preprocessing techniques in log abstraction, which is the most widely performed task after log filtering and log transformation among the six preprocessing tasks. In [39], a review and taxonomy of event abstraction were presented. Therefore, we will not focus on this category here.

4.7 Discussion

The *log filtering* task emerges as the most commonly performed preprocessing task, with over 63% of the case studies mentioning that some filtering is performed. However, it's worth noting that the specifics of the log filtering tasks appear to heavily rely on domain knowledge. Moreover, more than 30 papers use somewhat ambiguous terminologies such as 'irrelevant' or 'noise'. The *log transformation* task ranks as the second most frequently employed, accounting for 44%. Currently, the majority of subtasks in the log transformation focus on fixing format-related and data-quality issues. This highlights the importance of data quality in process mining and suggests that efforts to enhance data quality should continue to be a focal point in log preprocessing.

In contrast, log enriching (18%), log integration (16%), and log reduction (12%) tasks are notably less commonly performed. One plausible explanation is the limited support for these tasks in both academic and commercial tools. Furthermore, the relatively uncommon use of log reduction can be attributed to the fact that many filtering techniques inherently reduce the log size.

5 Conclusion

In this paper, we conducted a systematic literature review, examining the use of log preprocessing tasks in process mining case studies and presented the results. We identified six high-level tasks that were synthesized from the related work discussion and 20 low-level tasks inducted from the reported case studies. The log filtering task emerges as the most frequently used preprocessing task, featured in over 63% of the case studies reviewed. The log transformation task follows closely behind, accounting for 44% of the cases. Conversely, log enriching, integration, and reduction tasks are less commonly performed, possibly due to limited tool support. Future research can delve into these preprocessing tasks, providing operational guidance. Standardization in reporting practices and greater support for less common preprocessing tasks are valuable for improving traceability and advancing the reliability of process mining results.

References

1. van der Aalst, W.M.P.: Process Mining. Springer-Verlag Berlin Heidelberg (2011)
2. Benevento, E., Aloini, D., van der Aalst, W.M.: How can interactive process discovery address data quality issues in real business settings? evidence from a case study in healthcare. Journal of Biomedical Informatics **130**, 104083 (jun 2022)
3. Birk, A., Wilhelm, Y., Dreher, S., Flack, C., Reimann, P., Gröger, C.: A real-world application of process mining for data-driven analysis of multi-level interlinked manufacturing processes. Procedia CIRP **104**, 417–422 (2021)
4. Cenka, B.A.N., Santoso, H.B., Junus, K.: Analysing student behaviour in a learning management system using a process mining approach. Knowledge Management & E-Learning: An International Journal pp. 62–80 (mar 2022)
5. Chen, L., Klasky, H.B.: Six machine-learning methods for predicting hospital-stay duration for patients with sepsis: A comparative study. In: SoutheastCon 2022. IEEE (mar 2022)
6. Chen, Q., Lu, Y., Tam, C.S., Poon, S.K.: A multi-view framework to detect redundant activity labels for more representative event logs in process mining. Future Internet **14**(6), 181 (2022)
7. Cho, M., Park, G., Song, M., Lee, J., Lee, B., Kum, E.: Discovery of resource-oriented transition systems for yield enhancement in semiconductor manufacturing. IEEE Transactions on Semiconductor Manufacturing **34**(1), 17–24 (2020)
8. Dogan, O.: A process-centric performance management in a call center. Applied Intelligence **53**(3), 3304–3317 (may 2022)
9. Du, L., Cheng, L., Liu, C.: Process mining for wind turbine maintenance process analysis: A case study. In: 2021 IEEE 5th Conference on Energy Internet and Energy System Integration (EI2). IEEE (oct 2021)
10. van Eck, M.L., Lu, X., Leemans, S.J.J., Van Der Aalst, W.M.P.: PM2: A Process Mining Project Methodology. In: International conference on advanced information systems engineering. pp. 297–313. Springer (2015)
11. Esposito, L., Leotta, F., Mecella, M., Veneruso, S.: Unsupervised segmentation of smart home logs for human habit discovery. In: 2022 18th International Conference on Intelligent Environments (IE). IEEE (jun 2022)
12. Fahland, D.: Extracting and Pre-Processing Event Logs (2022)
13. Fahrenkrog-Petersen, S.A., Tax, N., Teinemaa, I., Dumas, M., de Leoni, M., Maggi, F.M., Weidlich, M.: Fire now, fire later: Alarm-based systems for prescriptive process monitoring. Knowledge and Information Systems **64**(2), 559–587 (dec 2021)
14. Gao, W., Wu, C., Huang, W., Lin, B., Su, X.: A data structure for studying 3D modeling design behavior based on event logs. Automation in Construction **132**, 103967 (dec 2021)
15. Goel, K., Leemans, S., Wynn, M.T., ter Hofstede, A., Barnes, J.: Improving PhD student journeys with process mining: Insights from a higher education institution. In: Proceedings of the Industry Forum (BPM IF 2021) co-located with 19th International Conference on Business Process Management (BPM 2021). pp. 39–49 (2021)
16. Han, J., Pei, J., Tong, H.: Data Mining: Concepts and Techniques. Morgan kaufmann (2022)
17. Huda, S., Aripin, Naufal, M.F., Yudianingtias, V.M.: Identification of fraud attributes for detecting fraud based online sales transaction. Indian Journal of Computer Science and Engineering **12**(5), 1409–1424 (oct 2021)
18. van Hulzen, G.A., Li, C.Y., Martin, N., van Zelst, S.J., Depaire, B.: Mining context-aware resource profiles in the presence of multitasking. Artificial Intelligence in Medicine **134**, 102434 (dec 2022)
19. Kitchenham, B., Brereton, O.P., Budgen, D., Turner, M., Bailey, J., Linkman, S.: Systematic Literature Reviews in Software Engineering – A Systematic Literature Review. Information and Software Technology **51**(1), 7–15 (jan 2009)

20. Lamghari, Z.: Process mining: A new approach for simplifying the process model control flow visualization. Transdisciplinary Journal of Engineering & Science **13** (jul 2022)
21. de Leoni, M., Pellattiero, L.: The benefits of sensor-measurement aggregation in discovering IoT process models: A smart-house case study. In: Business Process Management Workshops, pp. 403–415. Springer International Publishing (2022)
22. Lim, J., Kim, K., Song, M., Yoo, S., Baek, H., Kim, S., Park, S., Jeong, W.J.: Assessment of the feasibility of developing a clinical pathway using a clinical order log. Journal of Biomedical Informatics **128**, 104038 (apr 2022)
23. Marin-Castro, H.M., Tello-Leal, E.: Event Log Preprocessing for Process Mining: A Review. Applied Sciences **11**(22), 10556 (nov 2021)
24. Mivule, K.: Utilizing noise addition for data privacy, an overview (2013)
25. Pan, Y., Zhang, L.: Automated process discovery from event logs in BIM construction projects. Automation in Construction **127**, 103713 (jul 2021)
26. Pang, J., Xu, H., Ren, J., Yang, J., Li, M., Lu, D., Zhao, D.: Process mining framework with time perspective for understanding acute care: A case study of AIS in hospitals. BMC Medical Informatics and Decision Making **21**(1) (dec 2021)
27. Petersen, K., Feldt, R., Mujtaba, S., Mattsson, M.: Systematic mapping studies in software engineering. In: EASE (2008)
28. Pradana, M.I.A., Kurniati, A.P., Wisudiawan, G.A.A.: Inductive miner implementation to improve healthcare efficiency on indonesia national health insurance data. In: 2022 International Conference on Data Science and Its Applications (ICoDSA). IEEE (jul 2022)
29. Ramos-Gutiérrez, B., Varela-Vaca, Á.J., Galindo, J.A., Gómez-López, M.T., Benavides, D.: Discovering configuration workflows from existing logs using process mining. Empirical Software Engineering **26**(1) (jan 2021)
30. Ridwanah, R.D., Andreswari, R., Fauzi, R.: Analysis and implementation of TELKOM university lecture business processes evaluation on heuristic miner algorithm: A process mining approach. In: ISMODE. IEEE (jan 2022)
31. Rismanchian, F., Kassani, S.H., Shavarani, S.M., Lee, Y.H.: A data-driven approach to support the understanding and improvement of patients' journeys: A case study using electronic health records of an emergency department. Value in Health **26**(1), 18–27 (2023)
32. Sohail, S.A., Bukhsh, F.A., van Keulen, M.: Multilevel privacy assurance evaluation of healthcare metadata. Applied Sciences **11**(22), 10686 (nov 2021)
33. Stein Dani, V., Leopold, H., van der Werf, J.M.E.M., Lu, X., Beerepoot, I., Koorn, J.J., Reijers, H.A.: Towards understanding the role of the human in event log extraction. In: BPM Workshops. pp. 86–98. Springer International (2022)
34. Stephan, S., Lahann, J., Fettke, P.: A case study on the application of process mining in combination with journal entry tests for financial auditing (2021)
35. Suriadi, S., Andrews, R., ter Hofstede, A.H.M., Wynn, M.T.: Event log imperfection patterns for process mining: Towards a systematic approach to cleaning event logs. Information Systems **64**, 132–150 (2017)
36. Tang, J., Liu, Y., yi Lin, K., Li, L.: Process Bottlenecks Identification and Its Root Cause Analysis Using Fusion-based Clustering and Knowledge Graph. Advanced Engineering Informatics **55**, 101862 (jan 2023)
37. Tariq, Z., Charles, D., McClean, S., McChesney, I., Taylor, P.: Anomaly detection for service-oriented business processes using conformance analysis. Algorithms **15**(8), 257 (jul 2022)
38. Tavakoli-Zaniani, M., Gholamian, M.R., Hashemi-Golpayegani, S.A.: Improving heuristics miners for healthcare applications by discovering optimal dependency graphs. The Journal of Supercomputing **78**(18), 19628–19661 (jun 2022)
39. van Zelst, S.J., Mannhardt, F., de Leoni, M., Koschmider, A.: Event Abstraction in Process Mining: Literature Review and Taxonomy. Granular Computing **6**(3), 719–736 (2021)

A An overview of preprocessing tasks

Table 1. Category citation details of 86 papers.

High-level category	Low-level category	References
Log filtering (55)	Filtering irrelevant data (29)	[1, 5, 8, 13, 15, 16, 17, 20, 24, 29, 30, 32, 36, 37, 39, 48, 57, 58, 61, 64, 68, 69, 70, 71, 75, 76, 80, 81, 85]
	Filtering incomplete data (16)	[7, 19, 22, 24, 28, 29, 33, 39, 45, 56, 57, 59, 67, 68, 69, 80]
	Filtering infrequent data (13)	[6, 10, 18, 24, 34, 36, 38, 45, 59, 67, 78, 77, 79]
	Filtering duplicates (8)	[13, 21, 22, 24, 27, 59, 68, 70]
	Filtering outliers (5)	[12, 13, 17, 47, 66]
	Filtering incorrect data (4)	[28, 43, 56, 74]
	Filtering redundant data (2)	[18, 19]
	Filtering inconsistent data (1)	[16]
	Filtering noise (3)	[16, 46, 62]
Log transformation (38)	Transforming format (25)	[5, 9, 11, 13, 15, 16, 60, 22, 24, 25, 31, 32, 35, 36, 39, 41, 51, 55, 56, 59, 61, 66, 82, 83, 84]
	Transforming values (12)	[19, 24, 27, 43, 45, 52, 53, 55, 57, 68, 70, 71]
	Reordering (5)	[8, 22, 24, 48, 56]
	Transition matrices and encoding (2)	[24, 86]
Log abstraction (37)	-	[2, 3, 10, 12, 15, 18, 17, 20, 60, 23, 24, 25, 26, 30, 35, 40, 42, 43, 45, 46, 48, 49, 50, 51, 52, 53, 56, 58, 63, 65, 70, 76, 78, 77, 82, 85, 86]
Log enriching (16)	Adding calculation metrics (9)	[21, 23, 33, 38, 40, 52, 54, 65, 72]
	Labelling (4)	[4, 37, 55, 77]
	Adding case id (2)	[66, 75]
	Adding noise (1)	[73]
Log integration (14)	-	[14, 20, 60, 21, 25, 29, 34, 44, 53, 54, 59, 66, 70, 72]
Log reduction (11)	Dividing into sub-logs (9)	[19, 26, 29, 33, 41, 46, 62, 72, 81]
	Sampling (2)	[27, 74]
	Cutting traces (1)	[27]

Appendix Reference

1. Acacio-Claro, P.J., Estuar, M.R.J., Villamor, D.A., Bautista, M.C., Sugon, Q., Pulmano, C.: A micro-analysis approach in understanding electronic medical record usage in rural communities: Comparison of frequency of use on performance before and during the COVID-19 pandemic. Procedia Computer Science **196**, 572–580 (2022)
2. Adams, J.N., van Zelst, S.J., Rose, T., van der Aalst, W.M.: Explainable concept drift in process mining. Information Systems **114**, 102177 (mar 2023)

3. Ahmad, N.D., Mat, H., Shahuddin, A.Z., Shaffiei, Z.A., Elias, S.J., Hatim, S.M., Ahmad, S.: Process mining of cardiovascular diseases trajectories in malaysia public hospital: A feasibility study. In: 2021 2nd International Conference on Artificial Intelligence and Data Sciences (AiDAS). IEEE (sep 2021)
4. Araghi, S.N., Fontanili, F., Lamine, E., Okongwu, U., Benaben, F.: Stable heuristic miner: Applying statistical stability to discover the common patient pathways from location event logs. Intelligent Systems with Applications **14**, 200071 (may 2022)
5. Ardimento, P., Bernardi, M.L., Cimitile, M.: Using process mining to understand students' and teams' dynamics. In: Higher Education Learning Methodologies and Technologies Online, pp. 63–73. Springer International Publishing (2022)
6. Bahaweres, R.B., Amna, H., Nurnaningsih, D.: Improving purchase to pay process efficiency with RPA using fuzzy miner algorithm in process mining. In: 2022 International Conference on Decision Aid Sciences and Applications (DASA). IEEE (mar 2022)
7. Bahaweres, R.B., Trawally, J., Hermadi, I., Suroso, A.I.: Forensic audit using process mining to detect fraud. Journal of Physics: Conference Series **1779**(1), 012013 (feb 2021)
8. Bajo, O.E., Amarasinghe, I., Gutiérrez-Páez, N.F., Hernández-Leo, D.: Using process mining techniques to discover the collective behaviour of educators in a learning community platform. In: Collaboration Technologies and Social Computing, pp. 175–189. Springer International Publishing (2022)
9. Battineni, G., Chintalapudi, N., Amenta, F.: Model discovery, and replay fitness validation using inductive mining techniques in medical training of CVC surgery. Applied Computing and Informatics **18**(3/4), 245–255 (jul 2020)
10. Battineni, G., Chintalapudi, N., Zacharewicz, G.: Process mining in clinical practice: Model evaluations in the central venous catheter installation training. Algorithms **15**(5), 153 (apr 2022)
11. Beerepoot, I., Lu, X., Van De Weerd, I., Alexander Reijers, H.: Seeing the signs of workarounds: a mixed-methods approach to the detection of nurses' process deviations (2021)
12. Benevento, E., Aloini, D., van der Aalst, W.M.: How can interactive process discovery address data quality issues in real business settings? evidence from a case study in healthcare. Journal of Biomedical Informatics **130**, 104083 (jun 2022)
13. Birk, A., Wilhelm, Y., Dreher, S., Flack, C., Reimann, P., Gröger, C.: A real-world application of process mining for data-driven analysis of multi-level interlinked manufacturing processes. Procedia CIRP **104**, 417–422 (2021)
14. Brockhoff, T., Uysal, M.S., Terrier, I., Göhner, H., van der Aalst, W.M.P.: Analyzing multi-level BOM-structured event data. In: Lecture Notes in Business Information Processing, pp. 47–59. Springer International Publishing (2022)
15. Cenka, B.A.N., Santoso, H.B., Junus, K.: Analysing student behaviour in a learning management system using a process mining approach. Knowledge Management & E-Learning: An International Journal pp. 62–80 (mar 2022)
16. Chanifah, S., Andreswari, R., Fauzi, R.: Analysis of student learning pattern in learning management system (LMS) using heuristic mining a process mining approach. In: 2021 3rd International Conference on Electronics Representation and Algorithm (ICERA). IEEE (jul 2021)
17. Chen, L., Klasky, H.B.: Six machine-learning methods for predicting hospital-stay duration for patients with sepsis: A comparative study. In: SoutheastCon 2022. IEEE (mar 2022)
18. Chen, Q., Lu, Y., Tam, C.S., Poon, S.K.: A multi-view framework to detect redundant activity labels for more representative event logs in process mining. Future Internet **14**(6), 181 (2022)
19. Cho, M., Park, G., Song, M., Lee, J., Lee, B., Kum, E.: Discovery of resource-oriented transition systems for yield enhancement in semiconductor manufacturing. IEEE Transactions on Semiconductor Manufacturing **34**(1), 17–24 (2020)

20. Cuendet, M.A., Gatta, R., Wicky, A., Gerard, C.L., Dalla-Vale, M., Tavazzi, E., Michielin, G., Delyon, J., Ferahta, N., Cesbron, J., Lofek, S., Huber, A., Jankovic, J., Demicheli, R., Bouchaab, H., Digklia, A., Obeid, M., Peters, S., Eicher, M., Pradervand, S., Michielin, O.: A differential process mining analysis of COVID-19 management for cancer patients. Frontiers in Oncology **12** (dec 2022)
21. Dogan, O.: A process-centric performance management in a call center. Applied Intelligence **53**(3), 3304–3317 (may 2022)
22. Du, L., Cheng, L., Liu, C.: Process mining for wind turbine maintenance process analysis: A case study. In: 2021 IEEE 5th Conference on Energy Internet and Energy System Integration (EI2). IEEE (oct 2021)
23. Duma, D., Aringhieri, R.: Real-time resource allocation in the emergency department: A case study. Omega **117**, 102844 (jun 2023)
24. Dupuis, A., Dadouchi, C., Agard, B.: Predicting crop rotations using process mining techniques and markov principals. Computers and Electronics in Agriculture **194**, 106686 (mar 2022)
25. Erdogan, T.G., Tarhan, A.K.: Multi-perspective process mining for emergency process. Health Informatics Journal **28**(1), 146045822210771 (jan 2022)
26. Esposito, L., Leotta, F., Mecella, M., Veneruso, S.: Unsupervised segmentation of smart home logs for human habit discovery. In: 2022 18th International Conference on Intelligent Environments (IE). IEEE (jun 2022)
27. Fahrenkrog-Petersen, S.A., Tax, N., Teinemaa, I., Dumas, M., de Leoni, M., Maggi, F.M., Weidlich, M.: Fire now, fire later: Alarm-based systems for prescriptive process monitoring. Knowledge and Information Systems **64**(2), 559–587 (dec 2021)
28. Gao, W., Wu, C., Huang, W., Lin, B., Su, X.: A data structure for studying 3D modeling design behavior based on event logs. Automation in Construction **132**, 103967 (dec 2021)
29. Goel, K., Leemans, S., Wynn, M.T., ter Hofstede, A., Barnes, J.: Improving PhD student journeys with process mining: Insights from a higher education institution. In: Proceedings of the Industry Forum (BPM IF 2021) co-located with 19th International Conference on Business Process Management (BPM 2021). pp. 39–49 (2021)
30. Grüger, J., Geyer, T., Kuhn, M., Braun, S., Bergmann, R.: Verifying guideline compliance in clinical treatment using multi-perspective conformance checking: A case study. In: Lecture Notes in Business Information Processing, pp. 301–313. Springer International Publishing (2022)
31. Grüger, J., Kuhn, M., Bergmann, R.: Reconstructing invisible deviating events: A conformance checking approach for recurring events. Mathematical Biosciences and Engineering **19**(11), 11782–11799 (2022)
32. Hachicha, W., Ghorbel, L., Champagnat, R., Zayani, C.A., Amous, I.: Using process mining for learning resource recommendation: A moodle case study. Procedia Computer Science **192**, 853–862 (2021)
33. Huda, S., Aripin, Naufal, M.F., Yudianingtias, V.M.: Identification of fraud attributes for detecting fraud based online sales transaction. Indian Journal of Computer Science and Engineering **12**(5), 1409–1424 (oct 2021)
34. van Hulzen, G.A., Li, C.Y., Martin, N., van Zelst, S.J., Depaire, B.: Mining context-aware resource profiles in the presence of multitasking. Artificial Intelligence in Medicine **134**, 102434 (dec 2022)
35. Husin, H.S., Ismail, S.: Process mining approach to analyze user navigation behavior of a news website. In: 2021 The 4th International Conference on Information Science and Systems. ACM (mar 2021)
36. Ivanka, M.D., Andreswari, R., Fauzi, R.: Bottleneck analysis of lectures grades input process at information system academic using inductive miner. In: 2021 International Seminar on Machine Learning, Optimization, and Data Science (ISMODE). IEEE (jan 2022)

37. Jonk, J., Schaller, M., Netzer, M., Pfeifer, B., Ammenwerth, E., Hackl, W.: Process mining of nursing routine data: Cool, but also useful? In: Studies in Health Technology and Informatics. IOS Press (may 2022)
38. Kecht, C., Egger, A., Kratsch, W., Röglinger, M.: Quantifying chatbots' ability to learn business processes. Information Systems **113**, 102176 (jan 2023)
39. Khaosanoi, L., Limpiyakorn, Y.: Conformance checking and discovery of information service request process. In: 2021 14th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI). IEEE (oct 2021)
40. Koçi, R., Franch, X., Jovanovic, P., Abelló, A.: Web API evolution patterns: A usage-driven approach. Journal of Systems and Software **198**, 111609 (apr 2023)
41. Kołakowska, A., Godlewska, M.: Analysis of factors influencing the prices of tourist offers. Applied Sciences **12**(24), 12938 (dec 2022)
42. Kropp, T., Bombeck, A., Lennerts, K.: An approach to data driven process discovery in the cost estimation process of a construction company. In: Proceedings of the 38th International Symposium on Automation and Robotics in Construction (ISARC). International Association for Automation and Robotics in Construction (IAARC) (nov 2021)
43. Kropp, T., Faeghi, S., Lennerts, K.: Evaluation of patient transport service in hospitals using process mining methods: Patients' perspective. The International Journal of Health Planning and Management **38**(2), 430–456 (nov 2022)
44. Kumbhar, M., Ng, A.H., Bandaru, S.: Bottleneck detection through data integration, process mining and factory physics-based analytics. In: Advances in Transdisciplinary Engineering. IOS Press (apr 2022)
45. Lamghari, Z.: Process mining: A new approach for simplifying the process model control flow visualization. Transdisciplinary Journal of Engineering & Science **13** (jul 2022)
46. de Leoni, M., Pellattiero, L.: The benefits of sensor-measurement aggregation in discovering IoT process models: A smart-house case study. In: Business Process Management Workshops, pp. 403–415. Springer International Publishing (2022)
47. Lim, J., Kim, K., Song, M., Yoo, S., Baek, H., Kim, S., Park, S., Jeong, W.J.: Assessment of the feasibility of developing a clinical pathway using a clinical order log. Journal of Biomedical Informatics **128**, 104038 (apr 2022)
48. López-Pernas, S., Saqr, M., Viberg, O.: Putting it all together: Combining learning analytics methods and data sources to understand students' approaches to learning programming. Sustainability **13**(9), 4825 (apr 2021)
49. Macak, M., Kruzelova, D., Chren, S., Buhnova, B.: Using process mining for git log analysis of projects in a software development course. Education and Information Technologies **26**(5), 5939–5969 (may 2021)
50. Macak, M., Oslejsek, R., Buhnova, B.: Process mining analysis of puzzle-based cybersecurity training. In: Proceedings of the 27th ACM Conference on on Innovation and Technology in Computer Science Education Vol. 1. ACM (jul 2022)
51. Martinez, P., Montañes, O., Serralta, J.M., Tansini, L.: Modelling computer engineering student trajectories with process mining. In: LALA. pp. 48–57 (2021)
52. Mehraby, N., Neysiani, B.S., Nogorani, M.Z., Ataabadi, P.E.: Abnormal behavior detection in health insurance assessment process. In: 2022 8th International Conference on Web Research (ICWR). IEEE (may 2022)
53. Mertens, S., Gailly, F., Sassenbroeck, D.V., Poels, G.: Integrated declarative process and decision discovery of the emergency care process. Information Systems Frontiers **24**(1), 305–327 (oct 2020)
54. Oberdorf, F., Schaschek, M., Weinzierl, S., Stein, N., Matzner, M., Flath, C.M.: Predictive end-to-end enterprise process network monitoring. Business & Information Systems Engineering **65**(1), 49–64 (dec 2022)

55. Pan, Y., Zhang, L.: Automated process discovery from event logs in BIM construction projects. Automation in Construction **127**, 103713 (jul 2021)
56. Pang, J., Xu, H., Ren, J., Yang, J., Li, M., Lu, D., Zhao, D.: Process mining framework with time perspective for understanding acute care: A case study of AIS in hospitals. BMC Medical Informatics and Decision Making **21**(1) (dec 2021)
57. Pieters, A.J., Schlobach, S.: Combining process mining and time series forecasting to predict hospital bed occupancy. In: Health Information Science, pp. 76–87. Springer Nature Switzerland (2022)
58. Porouhan, P.: Optimization of overdraft application process with fluxicon disco. In: 2022 20th International Conference on ICT and Knowledge Engineering (ICT&KE). IEEE (nov 2022)
59. Pradana, M.I.A., Kurniati, A.P., Wisudiawan, G.A.A.: Inductive miner implementation to improve healthcare efficiency on indonesia national health insurance data. In: 2022 International Conference on Data Science and Its Applications (ICoDSA). IEEE (jul 2022)
60. R., D.S., Patil, M.M.: Study of learners behaviour in virtual learning environment using process mining. In: 2021 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT). IEEE (jul 2021)
61. Rahmawati, R., Andreswari, R., Fauzi, R.: Analysis and exploratory of lecture preparation process to improve the conformance using process mining. In: 2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC). IEEE (jan 2022)
62. Ramos-Gutiérrez, B., Varela-Vaca, Á.J., Galindo, J.A., Gómez-López, M.T., Benavides, D.: Discovering configuration workflows from existing logs using process mining. Empirical Software Engineering **26**(1) (jan 2021)
63. Rashid, K.M., Louis, J.: Integrating process mining with discrete-event simulation for dynamic productivity estimation in heavy civil construction operations. Algorithms **15**(5), 173 (may 2022)
64. Real, E.M., Pimentel, E.P., Braga, J.C.: Analysis of learning behavior in a programming course using process mining and sequential pattern mining. In: 2021 IEEE Frontiers in Education Conference (FIE). IEEE (oct 2021)
65. Revina, A., Ünal Aksu: An approach for analyzing business process execution complexity based on textual data and event log. Information Systems **114**, 102184 (mar 2023)
66. Ridwanah, R.D., Andreswari, R., Fauzi, R.: Analysis and implementation of TELKOM university lecture business processes evaluation on heuristic miner algorithm: A process mining approach. In: ISMODE. IEEE (jan 2022)
67. Rismanchian, F., Kassani, S.H., Shavarani, S.M., Lee, Y.H.: A data-driven approach to support the understanding and improvement of patients' journeys: A case study using electronic health records of an emergency department. Value in Health **26**(1), 18–27 (2023)
68. Rojas, G.S.P., Armas-Aguirre, J.: Integration method to protect the privacy and security of information in process mining projects: a case study on surgery block. In: 2021 IEEE Sciences and Humanities International Research Conference (SHIRCON). IEEE (nov 2021)
69. Ruschel, E., de Freitas Rocha Loures, E., Santos, E.A.P.: Performance analysis and time prediction in manufacturing systems. Computers & Industrial Engineering **151**, 106972 (jan 2021)
70. Sanchez-Segura, M.I., González-Cruz, R., Medina-Dominguez, F., Dugarte-Peña, G.L.: Valuable business knowledge asset discovery by processing unstructured data. Sustainability **14**(20), 12971 (oct 2022)
71. Saralaya, V., Saralaya, S., Kotian, L., Miranda, A., Bekal, I., Jyothi, Y.: Application of process mining for tuberculosis testing process. In: 2022 IEEE 7th International conference for Convergence in Technology (I2CT). IEEE (apr 2022)

72. Schuh, G., Gützlaff, A., Schmitz, S., Kuhn, C., Klapper, N.: A methodology to apply process mining in end-to-end order processing of manufacturing companies. In: Lecture Notes in Mechanical Engineering, pp. 127–137. Springer Singapore (oct 2021)
73. Sohail, S.A., Bukhsh, F.A., van Keulen, M.: Multilevel privacy assurance evaluation of healthcare metadata. Applied Sciences **11**(22), 10686 (nov 2021)
74. Song, W., Chang, Z., Jacobsen, H.A., Zhang, P.: Discovering structural errors from business process event logs. IEEE Transactions on Knowledge and Data Engineering **34**(11), 5293–5306 (nov 2022)
75. Stephan, S., Lahann, J., Fettke, P.: A case study on the application of process mining in combination with journal entry tests for financial auditing (2021)
76. Tang, J., Liu, Y., yi Lin, K., Li, L.: Process Bottlenecks Identification and Its Root Cause Analysis Using Fusion-based Clustering and Knowledge Graph. Advanced Engineering Informatics **55**, 101862 (jan 2023)
77. Tariq, Z., Charles, D., McClean, S., McChesney, I., Taylor, P.: Anomaly detection for service-oriented business processes using conformance analysis. Algorithms **15**(8), 257 (jul 2022)
78. Tariq, Z., Charles, D., McClean, S., McChesney, I., Taylor, P.: Time efficient end-state prediction through hybrid trace decomposition using process mining. In: 2022 14th International Conference on Computational Intelligence and Communication Networks (CICN). IEEE (dec 2022)
79. Tavakoli-Zaniani, M., Gholamian, M.R., Hashemi-Golpayegani, S.A.: Improving heuristics miners for healthcare applications by discovering optimal dependency graphs. The Journal of Supercomputing **78**(18), 19628–19661 (jun 2022)
80. Tavazzi, E., Gerard, C.L., Michielin, O., Wicky, A., Gatta, R., Cuendet, M.A.: A process mining approach to statistical analysis: Application to a real-world advanced melanoma dataset. In: Lecture Notes in Business Information Processing, pp. 291–304. Springer International Publishing (2021)
81. Theis, J., Galanter, W.L., Boyd, A.D., Darabi, H.: Improving the in-hospital mortality prediction of diabetes icu patients using a process mining/deep learning architecture. IEEE Journal of Biomedical and Health Informatics **26**(1), 388–399 (2021)
82. Thiyagarajan, G., S, P.: A process mining approach to analyze learning behavior in the flipped classroom. In: 2021 2nd International Conference on Communication, Computing and Industry 4.0 (C2I4). IEEE (dec 2021)
83. Tridalestari, F.A., Warsito, B., Wibowo, A., Prasetyo, H.: Analysis of e-commerce process in the downstream section of supply chain management based on process and data mining. (2022)
84. Valensia, L., Andreswari, R., Fauzi, R.: Implementation of process mining to discover student learning patterns using fuzzy miner algorithm (case study: Learning management system (LMS) telkom university). In: 2021 3rd International Conference on Electronics Representation and Algorithm (ICERA). IEEE (jul 2021)
85. Wisudiawan, G.A.A., Kurniati, A.P.: Process mining on learning activities in a learning management system. In: 2022 24th International Conference on Advanced Communication Technology (ICACT). pp. 476–482. IEEE (2022)
86. Yang, M., Moon, J., Jeong, J., Sin, S., Kim, J.: A novel embedding model based on a transition system for building industry-collaborative digital twin. Applied Sciences **12**(2), 553 (jan 2022)