Decision Mining in the Rail Industry: a Case Study in the Context of an Industrial Wheelset Revision Process

Bart Hoornstra, Joris Mens, Iris Beerepoot and Inge van de Weerd

March 27, 2024

Decision Mining in the Rail Industry: A Case Study in the Context of an Industrial Wheelset Revision Process

Abstract Process mining has led to new avenues of analysis and better process understanding. However, the role of decisions within the modeling and analysis of processes is underexplored. Following design science, a methodology for integrated process and decision mining was developed, based on the synthesis of an established process mining project methodology and a systematic literature review of existing decision mining approaches. The methodology was applied and evaluated in a case study at the Dutch national railway company. The results demonstrated that the addition of a decision perspective to process models allows for better process understanding. In addition, the evaluation identified a new form of conformance checking that can be used to validate whether the process was executed correctly in accordance with the decisions taken.

Keywords:
decision mining,
process mining,
process model enhancement,
conformance checking,
rail industry.

1 Introduction

Top-performing organizations typically employ agile decision-making based on rigorous analysis and use these insights to improve their day-to-day operations as well as to guide future strategies (LaValle et al., 2010). However, the upfront understanding of organizational decision-making is paramount for successful business analytics implementations (Sharma et al., 2014). Therefore, these do not inherently create value, especially since the technologies should merely be seen as tools — not drivers — that aid in dealing with information overload (Edmunds & Morris, 2000). The rapid advances in information technology have led to the paradoxical condition that, even though available information is abundant, it is more difficult to extract relevant and useful information when needed (Edmunds & Morris, 2000). Nevertheless, the *potential* value of improved decision-making enabled by inclusion of contextual process information justifies investments in new forms of data-driven analytics (Sharma et al., 2014).

A promising research area in data-driven analytics is *process mining* (van der Aalst & Weijters, 2004. Process mining allows not only for the investigation of causal relations between activities but also additional data attributes that enable the investigation of performance (timestamps) and workload (resources) (van der Aalst & Weijters, 2004). With the abundance of data available, it becomes increasingly relevant to critically assess and evaluate event log quality (Kherbouche et al., 2016). While research has been carried out to address these latter aspects for event logs (Fischer et al., 2020; Suriadi et al., 2017; van Wensveen, 2020), limited attempts have been made to enhance event logs with data from the context of the process execution (Banham et al., 2022). In that respect, the field of *decision mining* recently gained more widespread attention within the scientific community (De Smedt, vanden Broucke, et al., 2017). This development is grounded in the idea that at least some separation of concerns between business logic (rules, decisions) and processes should be achieved for the appropriate balance between flexibility, compliance, efficiency, and effectiveness of supporting information systems (Vanthienen et al., 2013).

While processes and decisions are intertwined by nature, there are several addressable issues observed at their intersection. Firstly, when a process model incorporates too detailed decision paths, it becomes more or less a decision tree represented as a cluttered process model. These unnecessarily convoluted process models are difficult to reuse and maintain (De Smedt, vanden Broucke, et al., 2017). Secondly, in process models where business rules imperatively constrain the control-flow, the flexibility required for the high volatility of such rules might be impaired. Thirdly, decisions might be the driver behind the activities and workflows of all process stakeholders, and as such they should be modeled separately to accurately document the related knowledge and to allow for reuse beyond a single process. Fourthly, a process might be the execution of a complex decision in itself, where the relationships between decisions should be explicitly modeled such that decision-making can be facilitated by an optimal process. Finally, processes that are highly dynamic, human-centric, and non-standardized could benefit from declarative process modeling where the principles are the same, but each case is genuinely distinct (Vanthienen et al., 2013). The aforementioned issues indicate that there does not exist a one-size-fits-all solution to integrate business logic with process knowledge and that knowledge on extending process mining with decision mining is lacking. Therefore, the research question for this paper is as follows: *How to extend process mining with decision mining?*

An existing process mining project methodology is followed in the form of PM² (Van Eck et al., 2015). The extended framework PM²xDM is developed using the DSRM (Peffers et al., 2007), and subsequently applied and empirically validated in an embedded, single-case study (Yin, 2018). The remainder of this paper is structured as follows. First, the background is sketched in terms of fundamental concept definitions related to process and decision mining, before the context of the case study is further elaborated. Then, the research method is explained concerning the phases of the DSRM and the results of the case study are presented. Finally, the implications, contributions, challenges, and limitations of this research are discussed, and an overall conclusion is drawn, complemented by an outlook on future research directions.

2 Background

2.1 Process mining

Process mining aims to *discover, monitor and improve real processes by extracting knowledge from event logs* (van der Aalst, 2011). The smallest unit of examination is an *event*, where each event refers to an *activity* within the process (e.g. a single step that has been completed). Each event belongs to a particular *case*, which is one execution of the process, sometimes referred to as *process instance*. All events must be *ordered* sequentially, either by a numerical property or for example by a *timestamp*. In addition, each event could contain more information such as the *resource* involved with the activity or additional *data attributes* about conditions, the state or execution of the process. All events from a set of process instances combined form an *event log* (van der Aalst, 2012).

Three types of process mining activities are commonly identified: discovery, conformance checking and enhancement (or extension). Process discovery is the creation of a model solely based on the observed events. Conformance checking deals with verifying whether an event log complies with an (existing) process model, and the other way around. Contrary to conformance checking, process enhancement does not compare a model with reality (van der Aalst, 2012). Instead, it tries to change, correct, extend or enrich the already existing model.

This can either be already accomplished by examining timestamps and calculating time differences to demonstrate service times, and to indicate possible bottlenecks. Additionally, one could include the resource attribute to for example identify underutilized resources, frequently execute related activities, or lead to specific or unwanted behavior. These different activities in turn correlate with four dominant analysis perspectives within the process mining paradigm (van der Aalst, 2016): control-flow, time, organizational/resource, and data.

2.2 From decision management and modeling to decision mining

Decision management and modeling are critical components of organizational strategy, that comprises a suite of methodologies and technologies designed to automate and refine decision-making processes (Yates, 2003). Central to this tandem is the use of data analysis, where business rules and business logic are investigated (Morgan, 2002; Von Halle & Goldberg, 2009). Business rules provide granular, formal guidelines for consistent, accurate, and legally compliant operations, while business logic offers a broader set of principles and processes that shape strategic decision-making and organizational operations, integrating goals, strategies, and operating principles with business rules, best practices, and industry standards (Morgan, 2002; Von Halle & Goldberg, 2009; Levina et al., 2010).

Emerging from this complex decision-making landscape is decision mining, a discipline that extends the traditional focus of process mining by exploring the impact of data attributes on decision-making within processes (Beerepoot et al., 2023). Decision mining acknowledges the data perspective of process mining, examining the nuances of how data informs workflow choices and complements process mining analyses (De Smedt et al., 2019; de Jong et al., 2021). It challenges the notion that workflow data and control-flow must be correlated, recognizing that decisions can affect data attributes and activities throughout a workflow without altering the sequence of activity execution (De Smedt et al., 2019). The integration of decision mining techniques with traditional process mining tools offers the potential for a comprehensive approach to process improvement, aiming for an integrated decision and process model representation that can better capture the complexity of organizational decision-making in relation to process execution (De Smedt, vanden Broucke, et al., 2017).

In sum, decision management, modeling, and mining can work in concert to enhance the organizational capacity for informed and strategic decision-making.

By recognizing the distinctive but overlapping roles of these disciplines, organizations can harness a holistic approach to improve their capabilities for process analysis and improvement.

4 Research method

The artifact that is being developed in this study is an extended methodological framework for the application of decision mining within a process mining project. Therefore, this project follows the design science research methodology (DSRM) proposed by Peffers et al. (2007). The steps are illustrated in Figure 1 and further described thereafter.

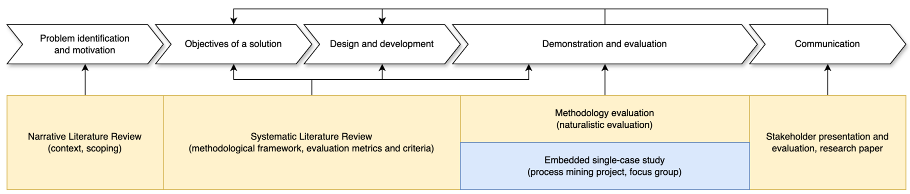

Figure 1: The DSRM and its implementation specific to this research project.
Source: adapted from Peffers et al. (2007).

Problem identification and motivation. As there does not exist a methodology for a decision mining project, an existing process mining project methodology is used as a basis.

Objectives of a solution. The objective is to design an extended methodology that integrates decision mining activities into a process mining project. The subsequent goal is to present an enhanced perspective on the process, where the integration of decision information into the process models allows for a better understanding of the process and relevant analysis activities, such as conformance checking.

Design and development. Based on a systematic review of the state-of-the-art literature, relevant activities and contextualized evaluation strategies are identified. These are subsequently integrated into the proposed methodological framework.

Demonstration and evaluation. The initial framework is applied within a process mining project at the Dutch national railway company in the context of an industrial wheelset revision process. The evaluation of the artifacts and the resulting insights is carried out with the relevant stakeholders and experts through a focus group.

Communication. The results are integrally presented to the stakeholders as part of the evaluation. Furthermore, the publication of this research report is an additional means of dissemination of the findings.

5 The initial methodological framework: PM²xDM

The methodological framework is constructed as an adaptation and extension of the widely-used PM² methodology by Van Eck et al. (2015). Figure 2 shows an overview of the initial framework. For each phase of PM², one or more complementary decision mining-related activities have been identified and assigned to those. The depicted steps are further illustrated and described in the context of the case study in Section 6.

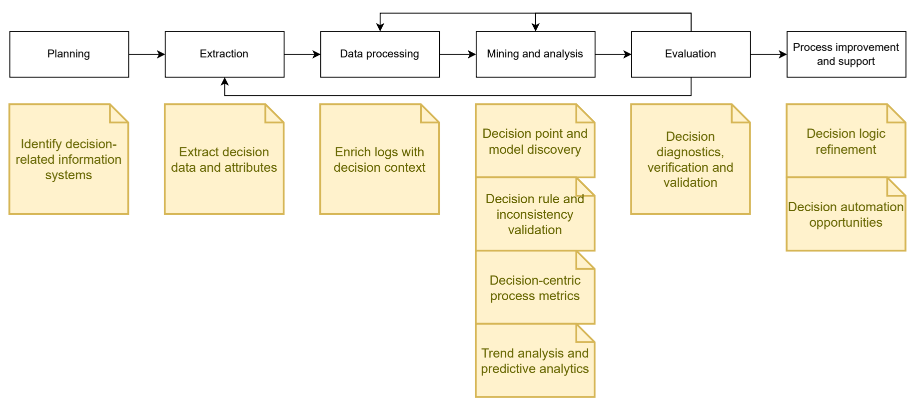

Figure 2: An initial overview of PM²xDM's decision-related activities

Source: based on Van Eck et al. (2015)

6 Case study

The case study is performed within the largest rail operator in the Netherlands. The organization employs around twenty thousand people and is responsible for the operation and maintenance of trains as well as all train stations in The Netherlands.

6.1 Context

Due to the size of the organization a diverse array of process domains is present. The current case study regarding *wheelset revision* (RLW) is a subdomain of the maintenance organization. A previous study by Smit & Mens (2019) identified this process as having a high data and event log quality due to its automated production line. It also scored highly on process mining success factors identified by Mans et al. (2013), when compared to other processes in the organization. The quality and availability of necessary data, as well as stakeholder commitment, contributed to the suitability of this process for the case study.

The wheelset revision process starts with preparation steps that involve cleaning, bearing removal, and gearbox inspection. Furthermore, a material plan is developed from pre-screening results to direct the treatment and routing of wheelsets and components. The actual wheelset revision follows, encompassing disassembly, axle decoating and inspection using non-destructive techniques, conservation with dual-layer coating, reassembly at the on-press station, and final measurements and adjustments. Non-gearbox axles undergo additional balance testing before final assembly and quality checks. The facility accommodates 24 wheelset types, each with a numerical identifier and specific to train models. Wheelsets are categorized into motor types, equipped with gearboxes and brake plates, and running types, which lack a direct drive connection. The treatment path for each wheelset type is predefined in a material plan based on its components, guiding the process flow upon factory entry.

6.2 Stage 1: Planning

The revision process is managed by a Manufacturing Execution System (MES), ranging from measurement assessment, routing decisions, and control of equipment and machines. We identified the related information systems architecture supporting the process through document analysis and meetings with the MES system's product owner. MES as orchestrator interfaces with a system for logistic tracking and financial reporting, a system for asset maintenance tracking, while an ERP system manages inventory. A configuration management system stores unstructured text documents related to work procedures, which is not interfaced with MES. MES has its own internal repository for routing logic and measurement criteria.

6.3 Stage 2: Extraction

Event data for 2022 was provided as a CSV file with over 10 million rows and six columns. The data, in long format, required preprocessing for conversion into a usable event log, using Python with Pandas in a Jupyter Notebook for reshaping. Decision data extraction focused on the MES's descriptive attributes without seeking external sources. This phase aimed to understand routing decisions based on internal criteria, acknowledging the challenges in extracting comprehensive decision data at this stage. Knowledge transfer involved data reshaping and mapping to process mining concepts with domain expert involvement, streamlined into several interactive sessions and communications to minimize the expert burden.

6.4 Stage 3: Data processing

This stage utilized three tools for data exploration, event log manipulation, and model generation: Fluxicon Disco 3.6.7 for exploration of the data sets and creation/manipulation of event logs, ProM 6.13 (Verbeek et al., 2011) for process model generation beyond Directly Follows Graphs (DFGs) and PM4Py 2.7.4 with Scikit-learn (Berti et al., 2023; Pedregosa et al., 2011) for Petri net generation and decision mining. Initial log analysis revealed a highly complex spaghetti-like process model. Further investigation and expert discussions identified discrepancies due to premature equipment start events. To address this, additional activities were added to the event log, ensuring a comprehensive analysis while maintaining data integrity and clarity. This process refinement led to a streamlined dataset that preserves all data attributes, conducive to identifying process variances and generating a readable model despite inherent complexity.

6.5 Stage 4: Mining and analysis

6.5.1 Decision point and model discovery

An initial Directly-Follows Graph (DFG) for the wheelset revision process was generated using Disco, based on a Fuzzy miner approach (Gunther & van der Aalst, 2007). Despite technical challenges, such as Java errors in ProM due to the large feature space, adjustments to noise thresholds and filtering strategies enabled the creation of more interpretable models. Analysis in a Jupyter Notebook with Pandas and PM4Py facilitated the discovery of decision points and the examination of process variants and exceptions. By focusing on complete events and applying filters, issues related to loops were mitigated although this incurred some information loss. This highlighted the importance of considering both low-frequency paths for compliance and more frequent exceptions for pattern analysis.

6.5.2 Decision rule validation and model enhancement

Conformance checking is aimed at aligning real-world behavior with the process model, focusing on fitness and appropriateness. To investigate the different rules and path associations, the paths should be at least present in the model.

Therefore, the emphasis was on accommodating all traces and variants in the log and investigating exceptions through decision mining, even if this meant accepting certain exceptional cases to maintain a fitness level of 100% for an accurate decision mining analysis. The enhancement phase involved refining the process model with additional decision-related information, using a decision tree classifier for attribute analysis. This phase underscored the relevance of feature selection and the need to exclude non-explanatory attributes. Annotated decision points with guard expressions illustrated how specific conditions could direct process flow, enhancing model accuracy and interpretability. Figure 3 presents an example of such an annotation.

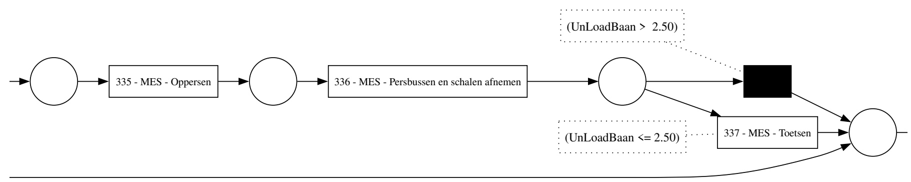

Figure 3: Example of an annotated decision point for an optional examination step in the Petri net for the most common wheelset type.

6.6 Stage 5: Evaluation

A focus group, complemented by intermediate collaborative discussions, evaluated the results of the framework with industry experts, focusing on its application and improvement opportunities. The final focus group evaluation episode, including the researcher and three domain experts, followed a predefined protocol (Krueger & Casey, 2015, Saunders et al., 2009) and lasted slightly more than two hours, discussing the application and the results thematically. The emphasis was on the decision point discovery and validating the respective annotations in the model. More advanced activities from the framework such as decision-based process metric, trend analysis, and predictive analysis were omitted due to feasibility reasons, either incurred by the available data or time constraints.

6.6.1 Process characteristics

The initial part of the evaluation revealed the adaptability of the process and the impact of its physical and logical architecture on the abstraction of event data. It was identified that physical constraints and logical configurability dictate process adaptability. Physically, some activities are time and location bound, due to an ordering constraint or factory layout. Nevertheless, the MES offers infinite logical configurations for extensive customization, influencing routing based on decision thresholds. The decision-making process is embedded in the software, with execution criteria evaluated at each step without making use of forecasting. An interesting notion was that *revision* processes like this reveal needs and information progressively, contrasting with predefined paths within a *production* process. The former trait is also seen in other types of processes, such as patient trajectories in healthcare, where diagnosis outcomes alter needs during execution.

6.6.2 Insights and process improvement potential

The evaluation underscored the importance of refining process models and decision criteria to obtain more accurate, applicable, and useful analysis results. Firstly, incorporating annotated decision points could improve process model accuracy and applicability, as one expert remarked that it is useful in that *"we want to understand the process, not the physical stations."* Although not all validated decision attributes were necessarily correct or explanatory, the expert remarked that *"I am cautiously a bit positive that you are already showing more than what I have seen so far in process mining by adding those decisions [in the model]."* Secondly, future work on this particular case should therefore first focus on refined feature engineering and subsequently on decision criteria representation in other modeling paradigms, such as BPMN. Thirdly, it was also identified that process mining tools and artifacts need better support for handling deliberate loops and rework, as this was represented as an attribute. However, representing a repeated activity separately could lead to a less comprehensible model. Finally, the evaluation concluded with an outlook on future use of the presented concept. The experts indicated that it could be used to validate if the wheelsets have been revised according to the regulations, in what would entail decision-based conformance checking. In other words, the paths in the model should align with the expected attribute values. This is especially relevant if a process exhibits more variation than expected. One expert illustrated that by stating that *"we apparently went through 262 different processes to deliver a wheelset. So, how do we know that all 262 variations are valid and produced a sound product? How can you guarantee that? [...] How can you adequately assess 262 different variations? [...] I think this should be possible if your model is a bit more accurate."* Another expert confirmed: *"Yes, this [concept] could then definitely help with that."*

7 Discussion

This research has demonstrated the relevance and applicability of decision mining within a process mining project. Enhanced process models were produced using case and activity attribute data, building on limited initial semantic knowledge about the process. An analysis of the decision points within the process aided by such visualizations demonstrated an interesting starting point for further applications, such as richer process documentation that shows under which conditions certain paths are taken (De Smedt, Hasić, et al. , 2017). In addition, an enhanced form of conformance checking could be developed using the enhanced models. Validation of whether the production of assets has been performed in accordance with the required guidelines and regulations could be supported using these artifacts (Levina et al., 2010). This implies that, depending on the project goals, it is worthwhile to assess the suitability for decision-mining analysis. However, improvements should be made to the input data and the decision-mining algorithm. More elaborate feature engineering and reduction of the feature space are areas of optimization. Moreover, the attributes from nonlocal activities should be considered, e.g. by enriching activities with attributes from earlier activities or a symbolic link that states the attributes of which other activities should be considered at a certain decision point.

Furthermore, we investigated what and how activities should be carried out and what they entail in terms of suitable process characteristics and data requirements to pursue a relevant and meaningful decision-mining analysis. A significant observation was that it should be possible to obtain a sufficiently readable process model at fitness levels greater than 80% to be able to perform a meaningful analysis. An argument for this is that if specific deviations are not present in the model, these will also not be annotated with the conditions under which they occur. Therefore, this type of analysis is less applicable to processes that are only loosely structured or exhibit an extreme degree of variation. This is in line with the analysis challenges posed by knowledge-intensive processes (Di Ciccio et al., 2015) or processes that accommodate a wide variety of different needs, such as healthcare processes (Munoz-Gama et al., 2022).

7.1 Contributions

The scientific contributions of this research are twofold. First, this research explored a potential avenue for a more holistic integration between process and decision mining, as suggested by De Smedt, Hasić, et al. (2017). Although it was unfeasible with the present tools and techniques to discover a fully integrated model of control flow and decisions, it supports the notion that the underutilized data perspective of process mining can provide relevant insights (Banham et al., 2022; van der Aalst, 2016). The methodology was implemented within a case study in a real-world context, and the insights were validated and evaluated within a focus group. Second, the foundational PM^2 methodology (Van Eck et al., 2015) has been extended with a decision-mining component. The synthesis of the common activities based on the literature and the practical implementation helps to increase our common understanding of the intersection between process and decision-mining, and helps in shaping future research opportunities for the respective activities that have been defined.

From a practical perspective, the proposed methodology can help practitioners systematically execute decision mining within a process mining project. Furthermore, since it is based on and integrated with a generic process mining project methodology, it can be included in an existing project if it aligns with the project goals. This in turn helps optimize efficient resource usage, as it does not require the creation of a distinct project as is the case with classical data mining projects that serve similar purposes (Osei-Bryson, 2012).

7.3 Future research

Future work could build on this research in several ways. First and foremost, the PM^2xDM framework should be repeatedly applied in different environments and contexts to develop a more robust context-agnostic version. Such follow-up experiments could, in addition, contain a part that also pays special attention to the execution of the methodology itself by process analysts. Second, research could focus on developing a toolkit that integrates several of the decision-mining assessment steps and activities of the framework into a single software package, for a more straightforward application within a process mining project. Furthermore, research could also focus on enabling additional interoperability between visualizations, such as the conversion of Petri nets with data into BPMN diagrams that retain these conditions.

8 Conclusion

The methodological framework PM^2xDM was developed based on the established process mining project methodology PM^2. It allowed us to enrich a Petri net process model with conditions based on the event data attributes, converting it into a Petri net with data (DPN). This research has shown that visualization of decisions in process models can be useful to organizations implementing a process mining project. Additionally, it helps to present a more realistic perspective on the process during discovery, and it allows for enhanced activities, such as decision-based conformance checking.

References

Banham, A., Leemans, S. J., Wynn, M. T., & Andrews, R. (2022). xPM: a Framework for Process Mining with Exogenous Data. *Process Mining Workshops: ICPM 2021 International Workshops, Eindhoven, The Netherlands, October 31–November 4, 2021, Revised Selected Papers*, 85–97.

Beerepoot, I., Barenholz, D., Beekhuis, S., Gulden, J., Lee, S., Lu, X., Overbeek, S., van de Weerd, I., van der Werf, J. M., & Reijers, H. A. (2023). A Window of Opportunity: Active Window Tracking for Mining Work Practices. *2023 5th International Conference on Process Mining (ICPM)*, 57–64.

Berti, A., van Zelst, S., & Schuster, D. (2023). PM4Py: A process mining library for Python. *Software Impacts*, *17*, 100556. https://doi.org/https://doi.org/10.1016/j.simpa.2023.100556

Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS quarterly*, 1165–1188.

De Smedt, J., Hasić, F., vanden Broucke, S. K., & Vanthienen, J. (2017). Towards a holistic discovery of decisions in process-aware information systems. *Business Process Management: 15th International Conference, BPM 2017, Barcelona, Spain, September 10–15, 2017, Proceedings 15*, 183–199.

De Smedt, J., Hasić, F., vanden Broucke, S. K., & Vanthienen, J. (2019). Holistic discovery of decision models from process execution data. *Knowledge-Based Systems*, *183*, 104866.

De Smedt, J., vanden Broucke, S. K., Obregon, J., Kim, A., Jung, J.-Y., & Vanthienen, J. (2017). Decision mining in a broader context: An overview of the current landscape and future directions. *Business Process Management Workshops: BPM 2016 International Workshops, Rio de Janeiro, Brazil, September 19, 2016, Revised Papers 14*, 197–207.

de Jong, R., Leewis, S., & Berkhout, M. (2021). Decision Mining versus Process Mining: a Comparison of Mining Methods. *2021 5th International Conference on Software and e-Business (ICSEB)*, 28–32.

Di Ciccio, C., Marrella, A., & Russo, A. (2015). Knowledge-intensive processes: Characteristics, requirements and analysis of contemporary approaches. *Journal on Data Semantics*, *4*, 29–57.

Edmunds, A., & Morris, A. (2000). The problem of information overload in business organisations: a review of the literature. *International Journal of Information Management*, *20*(1), 17–28.

Fischer, D. A., Goel, K., Andrews, R., van Dun, C. G. J., Wynn, M. T., & Röglinger, M. (2020). Enhancing event log quality: Detecting and quantifying timestamp imperfections. *Business Process Management: 18th International Conference, BPM 2020, Seville, Spain, September 13–18, 2020, Proceedings 18*, 309–326.

Gunther, C. W., & van der Aalst, W. M. (2007). Fuzzy mining–adaptive process simplification based on multiperspective metrics. *Business Process Management: 5th International Conference, BPM 2007, Brisbane, Australia, September 24-28, 2007, Proceedings*, *4714*, 328.

Hammer, M., & Champy, J. (1993). *Reengineering the corporation: A manifesto for business revolution*. Harper Collins, New York.

Kherbouche, M. O., Laga, N., & Masse, P.-A. (2016). Towards a better assessment of event logs quality. *2016 IEEE Symposium Series on Computational Intelligence (SSCI)*, 1–8.

Krueger, R. A., & Casey, M. A. (2015). Focus group interviewing. *Handbook of practical program evaluation*, 506–534.

LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2010). Big data, analytics and the path from insights to value. *MIT Sloan Management Review*.

Levina, O., Holschke, O., & Rake-Revelant, J. (2010). Extracting business logic from business process models.

2010 2nd IEEE International Conference on Information Management and Engineering, 289–293.

Mans, R. R., Reijers, H., Berends, H., Bandara, W., & Prince, R. (2013). Business process mining success. *Proceedings of the 21st European Conference on Information Systems*, 1–13.

Morgan, T. (2002). *Business rules and information systems: aligning IT with business goals*. Addison-Wesley Professional.

Munoz-Gama, J., Martin, N., Fernandez-Llatas, C., Johnson, O. A., Sepúlveda, M., Helm, E., Galvez-Yanjari, V., Rojas, E., Martinez-Millana, A., Aloini, D., et al. (2022). Process mining for healthcare: Characteristics and challenges. *Journal of Biomedical Informatics*, *127*, 103994.

Osei-Bryson, K.-M. (2012). A context-aware data mining process model based framework for supporting evaluation of data mining results. *Expert Systems with Applications*, *39*(1), 1156–1164.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems*, *24*(3), 45–77.

Saunders, M., Lewis, P., & Thornhill, A. (2009). *Research methods for business students*. Pearson Education.

Sharma, R., Mithas, S., & Kankanhalli, A. (2014). Transforming decision-making processes: a research agenda for understanding the impact of business analytics on organisations. *European Journal of Information Systems*, *23*(4), 433–441.

Smit, K. & Mens, J. (2019). Process mining in the rail industry: a qualitative analysis of success factors and remaining challenges. *Bled 2019 Proceedings, vol. 25*

Suriadi, S., Andrews, R., ter Hofstede, A. H., & Wynn, M. T. (2017). Event log imperfection patterns for process mining: Towards a systematic approach to cleaning event logs. *Information systems*, *64*, 132– 150.

Van der Aalst, W., Weijters, T., & Maruster, L. (2004). Workflow mining: Discovering process models from event logs. *IEEE transactions on knowledge and data engineering*, *16*(9), 1128–1142.

Van der Aalst, W. M. (1998). The application of Petri nets to workflow management. *Journal of Circuits, Systems, and Computers*, *8*(01), 21–66.

Van der Aalst, W. M., & Weijters, A. J. (2004). Process mining: a research agenda. *Computers In Industry*, *53*(3), 231–244. van der Aalst, W. M. P. (2004). Business Process Management: A personal view. *Business Process Management Journal*, *10*(2).

van der Aalst, W. M. P. (2011). *Process Mining: Discovery, Conformance and Enhancement of Business Processes* (Vol. 2). Springer.

van der Aalst, W. M. P. (2012). Process mining: Overview and opportunities. *ACM Transactions on Management Information Systems (TMIS)*, *3*(2), 1–17.

van der Aalst, W. M. P. (2016). *Process Mining: Data Science in Action* (Vol. 2). Springer.

van der Aalst, W. M. P., ter Hofstede, A. H. M., & Weske, M. (2003). Business Process Management: A Survey. *Business Process Management: International Conference, BPM 2003, Eindhoven, The Netherlands, June 26–27, 2003 Proceedings 1*, 1–12.

van der Aalst, W. M. P., Van Dongen, B. F., Herbst, J., Maruster, L., Schimm, G., & Weijters, A. J. M. M. (2003). Workflow mining: A survey of issues and approaches. *Data & Knowledge Engineering*, *47*(2), 237–267.

Van Eck, M. L., Lu, X., Leemans, S. J., & Van Der Aalst, W. M. (2015). PM2: a process mining project methodology. *Advanced Information Systems Engineering: 27th International Conference, CAiSE 2015, Stockholm, Sweden, June 8-12, 2015, Proceedings*, 297–313.

Vanthienen, J., Caron, F., & De Smedt, J. (2013). Business rules, decisions and processes: five reflections upon living apart together. *Proceedings SIGBPS Workshop on Business Processes and Services (BPS'13)*, 76–81.

van Wensveen, B. R. (2020). *Estimation and analysis of the quality of event log samples for process discovery* [Master's thesis].

Verbeek, H., Buijs, J. C., Van Dongen, B. F., & Van Der Aalst, W. M. (2011). XES, XESame, and ProM 6. *Information Systems Evolution: CAiSE Forum 2010, Hammamet, Tunisia, June 7-9, 2010, Selected Extended Papers 22*, 60–75.

Von Halle, B., & Goldberg, L. (2009). *The decision model: a business logic framework linking business and technology*. CRC Press.

Yates, J. F. (2003). *Decision management: How to assure better decisions in your company*. John Wiley & Sons.

Yin, R. K. (2018). *Case Study Research and Applications: Design and Methods* (6th ed.). Sage.