2021 International Conference on Frontiers of Information Technology (FIT) | 978-1-6654-0830-1/21/$31.00 ©2021 IEEE | DOI: 10.1109/FIT53504.2021.00064

Optimal Scenario Mining for Business Strategy Decision-making through Process Mining

Yuri van Midden
Faculty of Electrical Engineering, Mathematics and Computer Science
University of Twente
Enschede, The Netherlands
j.h.vanmidden@student.utwente.nl

Jeewanie Jayasinghe Arachchige
Department of Computer Science
University of Ruhuna
Matara, Sri Lanka
jeewanie@dcs.ruh.ac.lk

Faiza Allah Bukhsh
Faculty of Electrical Engineering, Mathematics and Computer Science
University of Twente
Enschede, The Netherlands
f.a.bukhsh@utwente.nl

***Abstract*—Process mining is often used for process discovery and conformance checking for big data analytics.**

Yet, after identifying problems, seeking solutions and deploying them is often the end of the process, whilst event logs analysis can also be used to yield business performance. This paper proposes an approach for utilizing process mining and event logs analysis to aid business decision-makers in narrowing down the considered business options and yielding business performance. After defining how event log analysis can be used to execute performance checking and thus narrowing down business options, a case study is conducted wherein the proposed approach is demonstrated. Evaluation has been performed through three experiments in a logistic case study, where possible business scenarios are considered, and after a given interval, the worst-performing scenarios are eliminated. The experiments show that the proposed approach of narrowing down business scenarios assists business decision-makers in finding an optimal scenario.

***Index Terms*—process mining, event log analysis, scenario mining, decision-making, logistics**

I. Introduction

It is not always that straightforward to effectively make tactical or strategical decisions that improve the business performance. Often there are time and money constraints that restrict testing and evaluating possible business scenarios. As resources, time, and money are typically limited, there is a need to make effective decisions that promise to result in the desired business performance.

Process mining is an emerging data science discipline [1], which is used to carry out performance and conformance checking in a business to answer questions like: “Why does it take so long to get a product from A to B?” or “What part of a process takes a lot of time?”

Even though process mining provides answers to these types of questions, process mining tend to be used as a goal in itself, instead of as a means to improve business decision-making and strategy. Furthermore, process mining and event logs analysis are scarcely used as tools for business decision-makers to evaluate different scenarios whilst the business is operating. Among the limited literature relates with decision mining with more detailed analysis of event logs, the research done by [2] and [3] propose a decision requirement diagram and decision table from event logs. However, this algorithm focuses on decision point analysis and it depends on expert knowledge.

This research aims to propose an approach to help business decision-makers in determining a strategy for effectively narrowing down the number of considered business scenarios. The narrowing down of business scenarios should be time efficient and substantiated by statistical arguments. Furthermore, to the proposed scenario mining approach steers in selecting scenarios that are expected to yield the most promising business performance. This is achieved by evaluating event logs of different scenarios over given periods of time, after which - at given intervals - a scenario selection is made to continue for evaluation. We have considered the CRISP-DM [4] methodology to structure this research. The motivation behind selecting CRISP-DM methodology is twofold. Firstly, this methodology is well-structured iterative approach which is both technology and problem-neutral. Therefore, it can be adopted to any kind of data mining research. Secondly, CRISP-DM encourages data miners to focus on business goals throughout the project. As our research goal is focusing on finding better business scenarios, the mentioned methodology is more suitable. CRISP-DM typically starts with Business Understanding, where we aim to figure out what business intelligence we need to solve the problem. This phase is discussed in section III. The second step of the methodology is Data Understanding in which we process the data to get an usable data set from the raw data. After data preparation phase, we start the Modeling and Design phase, which comprises tweaking the data sets and calibrating them to optimal values to ensure we can tackle the data/process mining project. This phase is explained in section IV. Finally, the Evaluation phase validates the work, meaning ensures that the proposed approach is able to tackle the data/process mining problem (Section V).

II. Background

A. Process Mining and Event Logs

Events in the context of event logs and process mining are actions as recorded in a log. They typically consist of data like the start time, completion time, activity description, allocated resources, cost, and a case ID [5]. Typically process mining

978-1-6654-0830-1/21/$31.00 ©2021 IEEE
DOI 10.1109/FIT53504.2021.00064

is intended to answer two categories of questions:performance and conformance [1]. Examples of performance questions are "Why are these products always late?" while conformance questions can be "Which part in the process is often skipped?". Using the event logs and the discovered Petri net model, an event log replay on the Petri net can be executed to monitor the performance of the events on the given traces. This also enables for evaluating what strategies result in business rules being violated.

We determine the following perspectives for scenario mining: conformance checking perspective, time & case perspective, organizational perspective and the data perspective.

B. Conformance Checking Perspective

Conformance checking is divided into four core principles:: fitness, simplicity, precision, and generalization [1]. Below we describe this in more detail.

We can determine the fitness of an event trace. A process model has a perfect fitness if all traces in the log can be replayed by the model from beginning to end [1]. We can define fitness by looking at the case and determine whether all events in the case exist in the process model from the beginning to the end. Another way of determining fitness is by checking if an event is represented in the process model. If all these analysis turn out to be true, then the fitness is 1. If none of the cases are represented by the model, the fitness is 0. Van der Aalst [1] describes the method of Occam's Razor, which translates to that the simplest representation of the behavior of the event logs is probably the best one. The precision relates to fitness. The model is precise when it does not allow for more types of behavior than can be imposed by the (potential) events. Last, generalization is opposite to precision. When a model is too specific, it may fit for some specific event logs, but not be able to fit other event logs.

C. Case Time Perspective

The case-and-time perspective is related to performance checking in process mining. Executing performance analysis in tools like ProM [6], aids in determining the time for cases to get from the beginning to the end in the model, as well as some basic statistics like the mean, median, min or max time for cases. For businesses in the logistics domain, performance indicators can, for example, be the mean of the clearing times of all cases or the mean plus standard deviation of the clearing times of all cases.

D. Organizational Perspective

Key performance indicators (KPIs) represent a means of the company in order to achieve their goals. For example, if a product at the end of the production line must have a "quality value" of at least 90%, it can be argued that this value should be a threshold for elimination business scenario options.

E. Data Perspective

The data perspective addresses quality measures. The event logs often contain many hidden data records that can be used for performance analysis, one being the case and time indicator. It is possible to infer data from the event logs, and use this data for performance measures. For example, if a cost attribute is included in the event logs, one can classify the costs as high, medium or low, which can be evaluated in the process discovery phase. This classifier method can enrich existing data sets.

This research will mainly focus on processing event logs in a logistics context. There are many KPIs that can be considered in a logistics context [7].

If we would examine event logs of a factory, we can imagine a few data points that are included in the event logs, like product quality, the timestamp, the unique identifier of the ingredient and the resource that is handling the ingredient. Some of these data points are suitable for quality and performance checking. For instance, the time it took for a vehicle to transport the goods can be a quality measure, or the quality decay of the ingredients themselves.

III. Optimal Scenario mining

Using process mining and (continuous) analysis of the event logs, the KPI's success rate can be monitored and evaluated. Even more, when a business scenario does not yield much performance and the KPI cannot be met, event logs analysis can help uncover this and eliminate such scenarios from the set of considered options. Optimal scenario mining can be done by answering the following three questions:

1) What are suitable performance indicators in a business in the logistics domain?
2) How can the event logs be used to derive quality measures?
3) How can event log analysis be used to narrow down possible business scenarios to comply with the requirements of the performance indicators?

The following part will explain how these questions were answered.

A. KPIs and Logistic Domain

Krauth et al. [8] have created a framework that indicates suitable performance indicators for logistics service providers. They divided their framework into internal and external KPIs, where internal means that the KPIs are relevant for the company's management and employees, and external means the customer and society (as a whole). Internal KPIs for managerial use are for example number of deliveries, trips per period, average fuel use per km, percentage of failed orders, and human resource costs.

B. Using Event Logs for Quality Measures

Event logs can hold other data fields than just the case ID, timestamp, and action, such as a temperature measure, or a quality attribute. As Mannhardt et al. [9] state, these attributes are often not used in process discovery, and therefore result in unreliable quality diagnostics for the discovered models. Therefore, methods are sought to perform analysis based on these quality measures.

TABLE I
BASIC CLASSIFIER FOR PERFORMANCE VALUES

value	performance class
$< \mu - 2\sigma$	*very bad*
$< \mu - \sigma$	*bad*
$< \mu$	*insufficient*
$< \mu + \sigma$	*sufficient*
$< \mu + 2\sigma$	*good*
$> \mu + 2\sigma$	*very good*

An approach to classifying these case times is by using the mean case time and the standard deviation to determine a human-readable score of the cases. These performance classes which is depicted in Table I, can be combined with the events in the event logs, which enables for performance checking using process mining (i.e., how often does a trace go from 'good' to 'insufficient'?).

C. Narrowing Down the Number of Business Scenarios

This sub-section introduces an approach to narrow down business scenarios for decision-makers, using event logs analysis. Possible scenarios one can think of are combinations of resource allocation (personnel and vehicles), investments and costs, KPIs, machinery, and so forth.

Our approach narrows down business scenarios, so that the promising business scenarios remain. For example, let us consider a set of event logs of a factory of different periods in time concerning hundreds of altering business scenarios. One can use these event logs simultaneously and evaluate the conformance and performance after every day, or every hour, or even at any given moment in time. The conformance and performance can be determined using process mining and basic statistics, and thus more promising business scenarios can be derived.

For this research, various event log replay strategies have been defined. For example: "..Play through the event logs simultaneously and after one hour".

- ...discard half of all the scenarios that performed worst. Then repeat."
- ...discard the single worst-performing scenario. Then, repeat."
- "...calculate the mean score of all scenarios and discard every scenario that performed worse than the mean value. Then repeat ."
- "..discard any scenario that has a lower score than the mean standard deviation. Then repeat."

These findings will be put into practice in the case study which is described in the following section.

IV. CASE STUDY

The case study is building upon previous research of [10] where a conceptual agent-based simulation framework is proposed to analyze and learn from emergent behavior in complex business situations. The simulations that were run involved a factory that processes certain products, which were transported by various types of vehicles: human-driven forklifts (HDF), automated guided vehicles (AGV) and un- manned aerial vehicles (UAV). A visual representation is shown in [10]. We can see the factory is partitioned into region 1, region 2 and region 3. For more information, we refer the reader to the work of [10].

The data sets, obtained include event logs of all different scenarios from [10]. Using the knowledge accounted for in section III-B, various experiments are run to narrow down the possible business scenarios to yield the best-performing scenario. By executing the various methods, the experiment converges to the best strategy after a certain period. One can see that a product arrives at the factory in region 1, is picked up and transported to region 2, and finally dropped off in region 3, which is considered the endpoint.

The experiments will focus on the 'currentDecayLevel' as a performance measure. This value is the representation of the quality of the product throughout the process, between 100 and 0, where 100 is high quality and 0 is low quality.

A. Experiments

Every experiment uses a concise algorithm that returns a ranking of the scenarios over a given period of time. The steps to determine the performance of scenarios is summarized as follows:

1) Traverse through event logs from timestamp A to timestamp B
2) Filter on all events called 'droppedOffRegion3'.
3) Of the remaining events, calculate the mean of all values of 'currentDecayLevel' and save it.

From now on, this algorithm will be referred to as the scenarios performance algorithm. The resulting value is considered the performance measure in these experiments. A high result value means that the scenario yielded better performance than a low result. This algorithm 1 (pseducode) is executed for every scenario, after which a ranking can be made of every scenario over the time span from A to B.

When running the experiments, the algorithm starts at the start of the event logs (e.g. timestamp 0) and traverses from time span AB, to time span BC, to time span CD, and so forth.

The different experiments that were executed for this research were summarized as follows:

1) After every iteration, use the top half of the scenarios according to the performance value and iterate once again until one scenario remains.
2) After every iteration, use the scenarios that scored higher than the mean of the performance values of all scenarios and iterate once again.
3) After every iteration, discard the scenario that yielded least performance and re-iterate with the other scenarios.

All experiments were run using Python and the Pandas2 data analysis library for Python. The data sets were loaded and then processed using the described steps. The experiments themselves are explained in more detail below.

Algorithm 1 Function for calculating the mean of the decay values of 'finished' products, in an array of scenarios (event logs)

function DETERMINEMEANDECAYOVERTIME(array $scenariosData$, timeStamp $start$, int $delta$)
  $calulatedDecayMeans$: [*scenarioID*, *calculatedValue*]
  for all $scenario \in scenariosData$ **do**
    $decayLevels$: $[value]$
    for all $event \in scenario.events$ **do**
      if $start < event.timeStamp < (start + delta)$ AND $event =$ *droppedOffRegion3* **then**
        $decayLevels \frown$ [*event.currentDecayValue*]
      end if
    end for
    calculatedDecayMeans $\frown [scenario.id, mean(decayLevels.values)]$
  end for
  return *calculatedDecayMeans*{Array of tuples of { *scenarioID*, *mean decay level of finished product* }}
end function=0

B. Experiment 1

In the first experiment, after every iteration of the scenarios performance algorithm, the resulting performances are sorted from best performance to worst performance. Then, the lower half of the scenarios are discarded, and the scenarios performance algorithm 2 is executed on the top half of the scenarios. This approach is re-iterated until one scenario remains.

The time interval chosen in this experiment is 60 minutes, which means that after 60 minutes, all events 'droppenOffRegion3' are evaluated over the time span of the past 60 minutes. Following this time interval, the first iteration will run from timestamp 0:00:00.0 to 0:59:59.9, the second iteration from 1:00:00.0 to 1:59:59.9, and so forth.

C. Experiment 2

This second experiment is similar to the first experiment, but incorporates a key difference. The results of the scenarios performance algorithm are now combined to calculate the mean performance of all scenarios. This repeats until one scenario remains. The time interval remained 60 minutes.

D. Experiment 3

In the third experiment, the results of the scenarios performance algorithm are sorted, and the scenario that yielded the least performance is discarded. The rest of the scenarios are evaluated over the next period of time and this is reiterated until few scenarios remain.

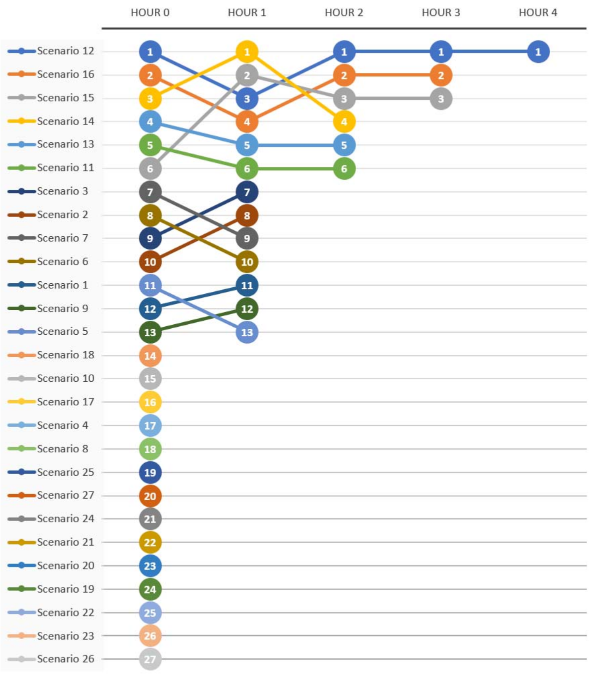

Fig. 1. Visualisation of experiment 1. After every hour, half of the scenarios are discarded. Numbers in the data points represent the current ranking.

V. EXPERIMENTAL EVALUATION

In this section, the results of the experiments will be discussed. It is interesting to look at the results as a business decision maker. They run several scenarios, in this case in a factory, and then eliminate scenarios that yield the least business performance.

A. Experiment 1

The first experiment ran for four hours, and after every hour, the bottom half of the scenarios were eliminated. This results are visualized in Figure 1.

This result shows that scenarios 12, 16, and 15 formed the top three of all the scenarios after 3 hours. As can be seen in Figure1, scenarios 12 and 16 both utilized the product-initiated rule: call 'shortestTravelDistance', where scenario 16 also utilized the vehicle-initiated rule: call the highest quality. According to this experiment, these scenarios yield the most business performance.

B. Experiment 2

The second experiment ran for five hours. After every hour, the mean product quality of all scenarios was calculated, and every scenario that scored under this value was eliminated. The resulting outcome is shown in Figure 2.

It is remarkable that the top three scoring scenarios are different than from the first experiment, while the conditions

Algorithm 2 Functions for the three experiments

Require: int $deltaMin = 60$ {Since this is the same value throughout algorithm, this variable is global}
function ELIMINATEBOTTOMHALFANDCONTINUE(array $scenariosData$, timeStamp $previousTimeStamp$)
  $sortedScenariosData \leftarrow sortOn(scenariosData.meanDecayLevels)$
  $meanQualityDecay \leftarrow mean(scenariosData.meanDecayLevels)$
Experiment 1
  $topScenariosData \leftarrow \mathit{sortedScenariosData}[0 \cdots 0.5 \times length]$
Experiment 2
  $topScenariosData \leftarrow \mathit{sortedScenariosData}[0 \cdots \{sortedScenariosData.value > meanQualityDecay\}]$
Experiment 3
  $topScenariosData \leftarrow \mathit{sortedScenariosData}[0 \cdots length - 1]$
  return DETERMINEMEANDECAYOVERTIME($topScenariosData, previousTimeStamp + \mathit{deltaMin}, \mathit{deltaMin}$)
end function
Require: $lastResults \leftarrow [scenarioData]$
Require: $lastStartMinute \leftarrow 60$
Ensure: int $iterations = 5$ {This value is picked by hand}
while $iterations > 0$ **do**
  results $\frown$ ELIMINATEBOTTOMHALFANDCONTINUE($lastResults, lastStartMinute$)
  $lastStartMinute \leftarrow lastStartMinute + deltaMin$
  $iterations \leftarrow iterations - 1$
end while=0

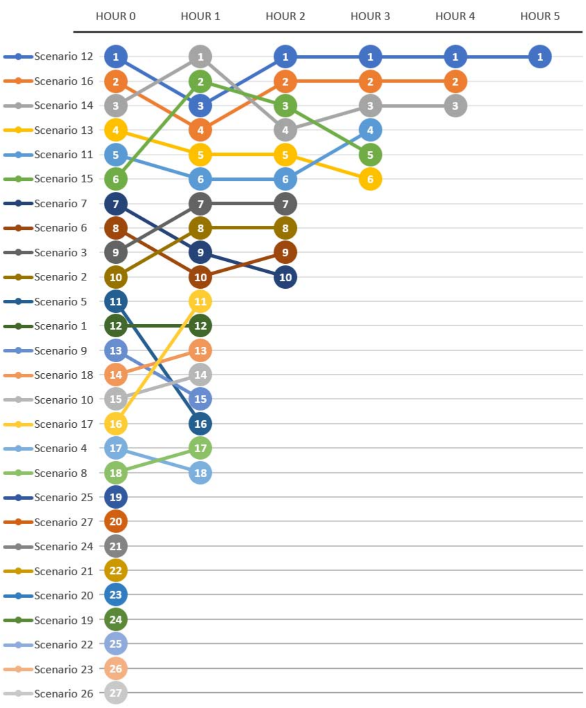

Fig. 2. Visualisation of experiment 2. After ever hour, scenarios performing worse than average are discarded. Numbers in the data points represent the current ranking.

were quite similar. Only the elimination conditions were changed. In this experiment, scenarios 12, 14 and 16 were the highest-scoring scenarios. Scenario 15, which in experiment 1 ended in the top 3 scenarios, was eliminated after the fourth hour into the experiment, as the performance indicator over the past hour scored lower than the average of all the scenarios. In experiment 1, scenario 14 was eliminated as part of the bisecting strategy, which does not necessarily mean that its score was 'bad'. In fact, when looking at the raw results, the difference between the scores of scenario 14 (89.5%) and scenario 15 (89.7%), that was not eliminated, was only 0.2% which is small when compared to other values.

C. Experiment 3

The third experiment was executed with relatively short intervals, namely 5, 10, 20, and 30 minutes. The scenario that yielded the least performance was eliminated and the experiment continued in the next time span. One of the interesting results when the experiment was executed with a time slot interval of 20 minutes (which is considered to be a suitable interval) were: (a)The sample size of finished products is large enough during most time spans. (b) It was possible to finish the algorithm before the event logs ran out, i.e. it was possible to eliminate every scenario but one. A visualization of the last intervals in the experiment (minutes 340-540), filtered on the six top-scoring scenarios, is depicted in Figure3.

Two key events stand out in this graph:

- The scenario that yielded the most performance after 540 minutes (12) was ranked seventh when polled after 360 minutes.
- The scenario that yielded the second most performance after 520 minutes (15), was ranked fifth after 340 minutes into the experiment.

The three conducted experiments show similar results. Most of the experiments return scenarios 12, 14, 15, and 16 as the best scenarios. When the course of the run-time of the

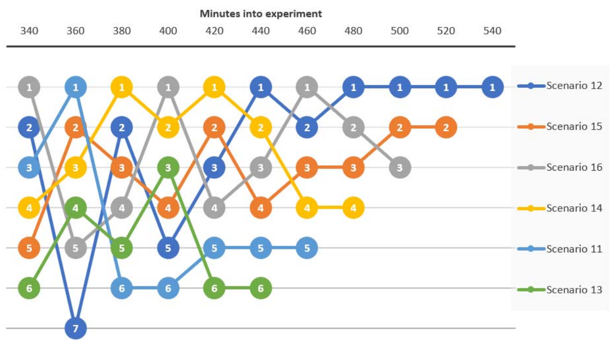

Fig. 3. The rank progression of the 6 best performing scenarios, in the last three hours of the experiment.

experiments is observed, these scenarios often score relatively high. However, an important observation is that the scores of the scenarios fluctuate after every interval, as illustrated clearly in experiment 3. The top-ranked scenarios tend to vary in ranking. However, they stay at the top-scoring best.

Further, a few observations are lead out for an improvement on this particular work. The experiments show that event logs analysis and scenario elimination based on the results can help business-makers narrow down the available options. However, a few ideas are proposed to make the experiments and thus the results more reliable:

- The sample size in the time intervals needs to be considered. When business decisions are made based upon small sample sizes, the decision may turn out to demote business performance.
- Statistical analysis should be taken into account when processing the results. For example, a combination of the mean and standard deviation can be used as performance indicators instead of just the mean values in these experiments

VI. Conclusion

In many business domains, many variables have an influence on the performance of a business. For business decision-makers it can be challenging to make choices that yield the most performance, as time and money constraints limit the ability to test all variations of the variables. Using event logs analysis and performance analysis, this research aims to provide an approach for business decision-makers to narrow down the number of considered business options for optimal decision making.

To achieve this goal, three questions were answered. First, the performance indicators of a business in the logistical sector are defined. Second, an analysis is made on how event logs can aid in providing performance information, and third, definitions are made for how event logs analysis can be used to narrow down business scenario options and comply with the given performance indicators.

These findings are used in a case study, where three experiments are conducted in a simulated factory. First, the performance indicator in the event logs is defined, after which the experiments can be conducted. All possible scenarios (i.e. the variations of the business variables) are evaluated for a given time period, after which the worst-performing scenarios are dropped. The experiments then continue until they converge to the best performing scenarios, according to the performance score. The approach of evaluating the event logs and dropping strategies is altered across the three experiments.

The experiments returned similar results in that the top-performing scenarios were similar across the experiments. Furthermore, these results show that event logs analysis, combined with performance indicators, can aid business decision-makers. More research is required to be done on validating the proposed strategy, for example, by conducting more elaborate case studies on real world datasets. s

Acknowledgment

We would like to thank Rob Bemthuis from University of Twente for his valuable contributions.

References

[1] W. V. der Aalst, "Process mining: Data science in action." Springer, 2016.
[2] E.Bazhenova, "Discovery of decision models complementary to process models," in *PhD Thesis*, 2018.
[3] . M. E. Bazhenova, S. Buelow, "Discovering decision models from event logs," in *Business Information Systems*, pp. 237–251.
[4] R. Wirth and J. Hipp, "Crisp-dm: Towards a standard process model for data mining," in *4th international conference on the practical applications of knowledge discovery and data mining*, vol. 1. Springer-Verlag London, UK.
[5] W. V. D. A. et al, "Process mining manifesto," in *In Lecture Notes in Business Information Processin*, vol. 99. LNBPI, 2012, p. 169–194.
[6] B. F. D. H. M. W. Verbeek, J. C. A. M. Buijs and W. M. P. V. D. Aalst, "Prom 6: The process mining toolkit," in *In CEUR Workshop Proceedings*, vol. 615, pp. 34–39.
[7] V. P. E. Krauth, H. Moonen and M. Schut, "Performance indicators in logistics service provision and warehouse management – a literature review and framework," in *European Operations Management Association International Conference on Operations and Global Competitiveness[EUROMA 2005]*, 2005.
[8] V. P. Krauth, H. Moonen and M. C. Schut, "Performance measurement and control in logistics service providing," in *Proceedings of the 13th European Conference on Information Systems ICEIS*, 2005, pp. 239–247.
[9] M. D. L. F. Mannhardt and H. A. Reijers, "Heuristic mining revamped: An interactive, data-aware, and conformance-aware miner," in *In CEUR Workshop Proceedings*, 2017.
[10] M. E. I. R. Bemthuis, M. Mes and P. Havinga, "Using agent-based simulation for emergent behavior detection in cyber-physical systems," in *2020 Winter Simulation Conference (WSC)*, 2020.