MASTER

Discovering Explicit Scale-Up Criteria in Crisis Response with Decision Mining

Lukassen, Britt G.H.

Award date:
2022

Eindhoven University of Technology

Department of Industrial Engineering and Innovation Sciences
Operations Management and Logistics

Discovering Explicit Scale-Up Criteria in Crisis Response with Decision Mining

Master Thesis

B.G.H. (Britt) Lukassen
0952766

Supervisors:

Dr. Y. Zhang	First supervisor at TU/e
Dr. L. Genga	Second supervisor at TU/e
Arian van Donselaar	Veiligheidsregio Utrecht
Bram Jacobs	Veiligheidsregio Utrecht
Michiel Rhoen	Veiligheidsregio Utrecht

Final version

Eindhoven, Monday 9th May, 2022

Abstract

Crisis management is necessary when a daily incident evolves into a crisis, requiring more coordination and/or causing a large impact. Adequate response prevents crisis situations or minimizes the impact. The overall aim of crisis management is to provide the right resources to control the situation and return to a normal situation as soon as possible. However, no objective criteria are set for when and what scale-up is applicable. Creating a dependency on the experience of the operational commanders to observe and take initiative to implement the appropriate multidisciplinary scale-up. In this research, the decision mining approach is applied to discover explicit criteria that can support decision-makers in crisis response. Decision mining can be used to make implicit knowledge explicit and to discover business rules. The approach has input from historical data and a questionnaire. After which, process mining and data mining techniques are applied. With the questionnaire, insight is gained into which criteria are considered by human decision-makers. These criteria serve as input for the features that are created with the historical data for machine learning. Process discovery is used to identify important decision moments in crisis response. After which, machine learning is used to discover deviations in data patterns that can support decision-making for these different decision moments. For each decision moment criteria are found that positively or negatively impact the decision for scale-up with help of the SHAP explainer. With the found criteria it is possible to accurately predict the human decisions in the past.

Executive Summary

On a daily-base, incidents are reported at the emergency dispatch center. These incidents can be road accidents, kitchen fires, or many more. Based on the emergency of the incident on hand, disciplines are sent there to help solve disturbances of daily life. These disciplines include the fire brigade, the police, the medical support, and the municipality, management by the emergency dispatch center which is a part of the safety regions. These safety regions support the collaboration between disciplines to efficiently coordinate crises.

This thesis focuses on crisis management for the Safety Region Utrecht (VRU). Crisis management is necessary when a daily incident evolves into a crisis, requiring more coordination and/or causing a large impact. A specific procedure is designed to resolve these crises as soon as possible; Coordinated Regional Incident Management Procedure (GRIP). Adequate response prevents crisis situations or minimizes the impact. The overall aim of crisis management is to provide the right resources to control the situation and return to a normal situation as soon as possible. Crisis response processes are characterized by being dynamic, highly knowledgeable, and unstructured (Herrera & Díaz, 2019). Furthermore, there are critical factors such as time and information availability that must be considered when making decisions in crisis situations (Kushnareva et al., 2015). These are also the challenges the VRU has to deal with when resolving crises.

Research motivation

For the VRU several elements complicate the decision regarding scale-up in multidisciplinary incidents. The first element is for operational commanders that are authorized to scale up who all represent other interests and have another situational perspective. The second element that causes complications is the lack of a clear moment in time to decide to scale up or not to scale up. This moment in time is also not realizable since crisis situations ask for a response based on half of all information. The third complicating element is that insight has grown that incidents require customized responses. Not all incidents can simply be classified as routine or GRIP. However, no objective criteria are set for when and what scale-up is applicable. This creates a dependency on the experience of the operational commanders to observe and take initiative to implement the appropriate scale-up. Therefore, the main research question is:

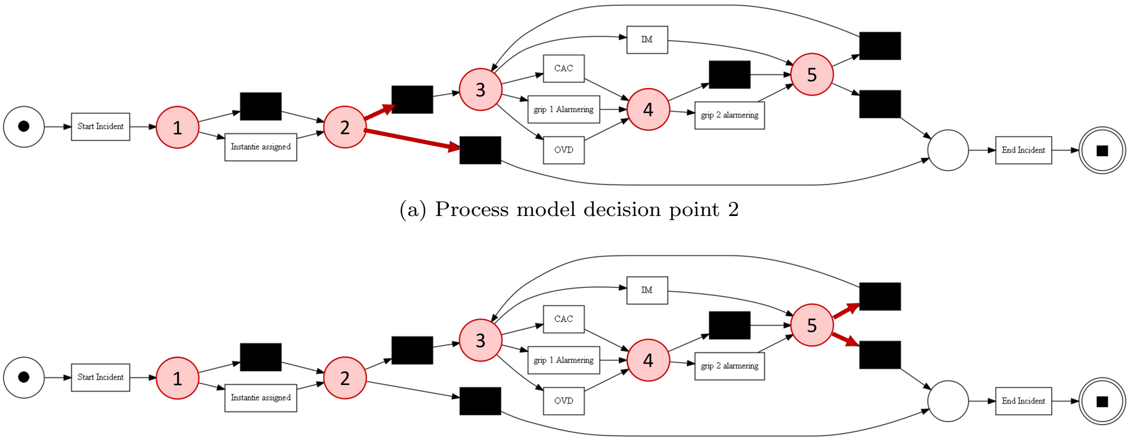
RQ: *Is it possible to define explicit criteria to support decision-makers, leading to the most appropriate multidisciplinary scale-up in crisis management?*

Approach and results

To answer this research question the decision mining approach is applied. Decision mining can be used to make implicit knowledge explicit and to discover business rules. The approach has input from historical data and a questionnaire. After which, process mining and data mining techniques are applied.

With a questionnaire, insight is gained into the implicit knowledge of decision-makers. We found that the criteria considered most by all different decision-makers are: ‘Incident location’, ‘Incident size’, ‘Incident type’, ‘(number of) injured and injury classification’, ‘Sensitivity on social media’, ‘Own disciplines involved’, ‘Expected duration incident’, ‘Possible effects on people/material’, ‘Safe/unsafe area’, ‘Involved partners’ and ‘Duration of incident unknown’. These criteria serve as input to create features in machine learning.

Additionally, the control-flow model is discovered with process mining. In this discovered process model, decision points are identified. The decision points are evaluated with the VRU to decide the most interesting points. This results in the selection of decision point 2 and 5, which are both highlighted in Figure 1. For decision point 2, the decision is between ‘Multi scale up required’ (upper arrow) and ‘Routine incident’ (lower arrow). In decision point 5, the decision is between ‘Additional multi scale up’ (upper arrow) and ‘No additional multi scale up’ (lower arrow).



(a) Process model decision point 2

(b) Process model decision point 5

Figure 1: Process model binary classification problem

For these decision points, machine learning techniques are used to discover deviations in data patterns that can support decision-making. Two modeling techniques are implemented and compared, namely the decision tree model and the random forest model. These models are compared based on performance metrics and explainability. All models have an f1-score above 90 % and have overall good performance. In the explainability of the model the following features are found: *Current duration*, *Deployed vehicles*, *Classification criteria 1*, *Classification criteria 2*, *Municipality*, and *Prefix*.

Finally, the important features found with machine learning are compared with the mentioned criteria in the questionnaire, to evaluate the alignment. Since the input of the model is based on the criteria found in the questionnaire only, the explicit criteria match the implicit knowledge. However, there was some translation required from the found criteria in the questionnaire towards the data features. Therefore, the criteria are not one on one. The features found important by the models are discussed with the VRU for interpretation purposes. From this, it can be concluded that most features do represent the mentioned implicit knowledge, except for the feature *Municipality*. The features found representative are *Current duration*, *Deployed vehicles*, *Classification criteria 1*, *Classification criteria 2*, and *Prefix*. For the different decision moments, different decision rules and important features are extracted. The features *Current duration* and *Deployed vehicles* are identified as the most important features for all data sets and models. However, the interpretation of these features is different for the decision moments.

Conclusion

To answer the main research question, it is possible to define explicit criteria to support decision-makers. The criteria found in this research are general guidelines for how the decisions are made by decision-makers of the VRU right now. With the found explicit criteria it is possible to apply the current judgment to new situations. This provides a general perspective based on the perspectives of all different decision-makers. Therewith, the aim is met to capture criteria independent of the decision-makers' viewpoint.

The found criteria are distinct in three decision moments. For each decision moment, other criteria sets are found, which are summarized here. In the first decision moment, a short current duration, 0 or more than 4 deployed vehicles, and the classification criterion accident, have all a positive impact on the outcome 'Multidisciplinary scale up required'. While a current duration longer than 1 hour, a number of deployed vehicles between 0 and 4, the classification criteria resuscitation and healthcare have a negative impact on the outcome 'Multidisciplinary scale up required'. In the second decision moment, a short current duration, more than 6 deployed vehicles, the classification criteria fire and building have a positive impact on the outcome 'Additional multidisciplinary scale up'. Furthermore, duration longer than 1 hour, less than 4 deployed vehicles, and the classification criteria accident have a negative impact on the outcome 'Additional multidisciplinary scale up'. Finally, for the third decision moment, a short duration, more than 10 deployed vehicles and the prefix 'IM', 'Instantie assigned', and 'GRIP 1' have all a positive impact on the outcome 'Additional multidisciplinary scale up'. Additionally, a current duration longer than 4 hours and less than 10 deployed vehicles have a negative impact on the outcome 'Additional multidisciplinary scale up'.

Future research could be focused on extending the found criteria by discovering specific criteria for each multidisciplinary activity (IM, OvD, CAC, GRIP 1 and GRIP 2). Moreover, this research insight is gained in explicit scale-up criteria but no improvement steps are taken. Future research could search for possible improvements based on the discovered knowledge in this research, to improve the crisis response procedure of the VRU.

Preface

This master thesis is the final deliverable for the master program ‘Operations Management and Logistics’ at Eindhoven University of Technology. The aim of this thesis was to discover explicit scale-up criteria in crisis response. I would like to take the opportunity to thank the people who supported me during my studies and this thesis project.

First of all, I would like to thank Yingqian Zhang. As my mentor and first supervisor of this thesis project, she has guided me through my masters and provided me with valuable insights. Thank you for your time and feedback during the past months. Moreover, I would like to thank Laura Genga as my second supervisor for her support and insightful conversations we had.

Secondly, I would like to thank everyone at Veiligheidsregio Utrecht for their enthusiasm and willingness to help me with this project. In special, I would like to thank Arian van Donselaar, Bram Jacobs, and Michiel Rhoen, who helped me understand the practical side of crisis management and supported me during the past months.

Finally, I would like to thank my friends and family for their unconditional support during my thesis project and entire studies. Thank you for making my student life an amazing time.

Britt Lukassen

List of Abbreviations

AI Artificial intelligence

BPMN Business Process Management and Notation

CAC Communicatie Adviseur (Communication Advisor)

CaCo Calamiteiten Coördinator (Calamities Coordinator)

CoPi Commando Plaats Incident (Commander Site Incident)

CRISP-DM Cross Industry Standard Process for Data Mining

DMN Decision Model and Notation

ERSs Emergency Response Systems

GBT Gemeentelijk Beleidsteam (Municipal Policy Team)

GMS Geïntegreerd Meldkamer Systeem (Integrated Control Room System)

GRIP Gecoördineerde Regionale Incidentbestrijdings Procedure (Coordinated Regional Incident Management Procedure)

IM Informatie Manager (Information Manager)

KiP knowledge-intensive processes

KPIs Key Performance Indicators

LCMS Landelijk Centraal Meldkamer Systeem (National Control Room System)

MCC Matthews Correlation Coefficient

ML Machine Learning

MPE Multi-perspective Process Explorer

OvD Officier van Dienst (Duty officer)

RBT Regionaal Beleidsteam (Regional Policy Team)

ROL Regionaal Operationeel Leider (Regional Operational Leader)

ROT Regionaal Operationeel Team (Regional Operational Team)

SHAP Shapley Additive Explanations

VIC Veiligheidsinformatiecentrum (Safety Information Center)

VP Veiligheidpaspoort

VRU Veiligheidsregio Utrecht (Safety Region Utrecht)

List of Figures

List of Tables

Contents

Chapter 1

Introduction

On a daily-base, incidents are reported at the emergency dispatch center. These incidents can be road accidents, kitchen fires, or many more. Based on the emergency of the incident on hand disciplines are sent there to help solve disturbances of daily life. These disciplines are including the fire brigade, the police, the medical support, and the municipality, management by the emergency dispatch center, and a part of the safety regions. These safety regions support the collaboration between disciplines to efficiently coordinate crises. Safety regions in general are founded by the fire brigade, thereby this is still a significant part of the safety regions. Each region is its own organization and has its own responsibilities. The ambition of the safety regions is to provide society with safety, control incidents, support victims, and resolve disruptions of daily life.

The Veiligheidsregio Utrecht (Safety Region Utrecht) (VRU) is the safety region of the region Utrecht. A collaboration of 26 municipalities with the highest population of all regions. They have policy plans specific for their region and prepare plans for events in their region. Besides, they were one of the first to introduce the Veiligheidsinformatiecentrum (Safety Information Center) (VIC), where all possible threats are monitored by mainly open-source data. They monitor for a confluence of threats that should be taken action on. Furthermore, when a crisis happens, they collect all relevant information needed to control the crisis.

This research focuses on crisis management by the VRU. Crisis management is necessary when a daily incident evolves into a crisis, requiring more coordination and/or causing a large impact. A specific procedure is designed to resolve these crises as soon as possible; Gecoördineerde Regionale Incidentbestrijdings Procedure (Coordinated Regional Incident Management Procedure) (GRIP). Adequate response prevents crisis situations or minimizes the impact. The overall aim of crisis management is to provide the right resources to control the situation and return to a normal situation as soon as possible. Crisis response processes are characterized by being dynamic, highly knowledgeable, and unstructured (Herrera & Díaz, 2019). Furthermore, there are critical factors such as time and information availability that must be considered when making decisions in crisis situations (Kushnareva et al., 2015). These are the difficulties the VRU has to deal with when resolving crises.

1.1 Research Motivation

Previously, research by Meeuwis (2021) for the VRU showed that on average 42 minutes pass before an incident is identified as a crisis, and therefore scaled up to GRIP. In this research, predictive modeling was applied to research the possibility of predicting GRIP in the first minutes after the start of an incident. The research concluded that it is possible to accurately predict

GRIP within the first 15 minutes after the start of an incident. This insight raised interest in why scale-up to GRIP takes on average 42 minutes while it can be predicted within 15 minutes. The VRU is interested in evaluating the decision process leading to a scale-up to GRIP. Expected by the VRU is that this time difference is due to subjective criteria and human factors involved in the decision for scale-up. Therefore, there is a need for more explicit criteria defining when and what scale-up is appropriate in crisis management.

For the VRU several elements complicate to decide for scale-up in multidisciplinary incidents. The first element is for operational commanders that are authorized to scale up who all represent other interests and have another situational perspective, as will be further explained in subsection 3.1.3. The second element that causes complications is the lack of a clear moment in time to decide to scale up or not to scale up. This moment in time is also not realizable since crisis situations ask for a response based on half of all information. The third complicating element is that insight has grown that incidents require customized responses. Not all incidents can simply be classified as routine or GRIP. More often, some operational processes require scaling up while others do not. For example, when a large fire has a high health impact on the surrounding population but the fire can be controlled. Therefore, flexible scale-up is applicable as customization of the crisis response. However, no objective criteria are set for when and what flexible scale-up is applicable. Creating a dependency on the experience of the operational commanders to observe and take initiative to implement the appropriate flexible scale-up.

Decisions for scale-up are based on subjective criteria and human judgment. In the emergency dispatch center, information is provided without any description or probability of the possible outcomes. According to research in experience-based decisions by Hertwig et al.(2004), there is a difference between decisions from description and decisions from experience. In case of decisions made from the description, a summary or probability is provided. On the other hand, for decisions made from experience, a bulk of information is provided only. Therefore, people can only use their own experience to make such decisions. The research by Hertwig et al. (2004) concluded that in the case of decisions from experience, people tend to underweight rare events. The decision-makers in crisis response are making decisions from experience, while rare events happen daily. Possibly a summary description with relevant criteria could help to make a proper estimation for each incident.

1.2 Research Questions

This research investigates if it is possible to extract guidelines for human decision-makers from data patterns. Therefore, the main challenge is to include the implicit knowledge of humans. To search specifically in the data for the quantitative measure of what humans think is important. But also interesting is the possibility of guidelines not yet considered by humans in decisions. The aim is to find these guidelines and make them explicit to support decision-makers. The focus is on operational scale-up for multidisciplinary incidents. Resulting in the following research question.

RQ: *Is it possible to define explicit criteria to support decision-makers, leading to the most appropriate multidisciplinary scale-up in crisis management?*

Decision mining is previously used to make implicit knowledge explicit (Petrusel, 2010) and to discover business rules (Campos et al., 2017). This is focusing on deviation in data patterns to support decision making. Research by Leewis, Smit et al.(2020) evaluated the current state and determined decision mining as; "the method of extracting and analyzing decision logs with the aim to extract information from such decision logs for the creation of business rules, to check compliance to business rules and regulations, and to present performance information". This method will therefore be further researched in the literature review of chapter 2. Corresponding to this method an event log is required as well as attribute data. Resulting in the first sub-question.

Q1: *Which data is available about the multidisciplinary scale-up of incidents?*

The incorporation of implicit knowledge in decision criteria is central to this research. In order to do so, the first step is to investigate what implicit knowledge is essential. Methods for knowledge acquisition should be explored while keeping in mind the aim to capture them as explicit criteria. Therefore, the following sub-question is set.

Q2: *How can implicit knowledge be acquired and taken into account for explicit criteria?*

Understanding the process is a starting point for understanding the decision that has to be made. The context of the process as well as an actual control-flow model is of interest. The context explains aspects decision-makers deal with while the control-flow model reveals the actually executed multidisciplinary scale-up process. The control-flow model is also of interest to find relevant decision points in this scale-up process. The next sub-question is formulated to research these aspects.

Q3: *How does the current multidisciplinary scale-up process look like and what activities are involved?*

As mentioned, data patterns are of interest to see what leads to a certain decision. In decision mining, these are found by applying machine learning techniques. Especially, decision trees are often used because of their explanatory nature. For example, decision trees are used for the purpose of explaining human decision-makers' decisions in ambulance dispatching by Theeuwes (2019). However, this modeling technique is often not the best-performing model. Therefore, other models are explored for better performance, while keeping the explainability in mind. This results in the following question.

Q4: *Which machine learning techniques can be used to predict multidisciplinary scale up while the model is explainable to distract decision rules?*

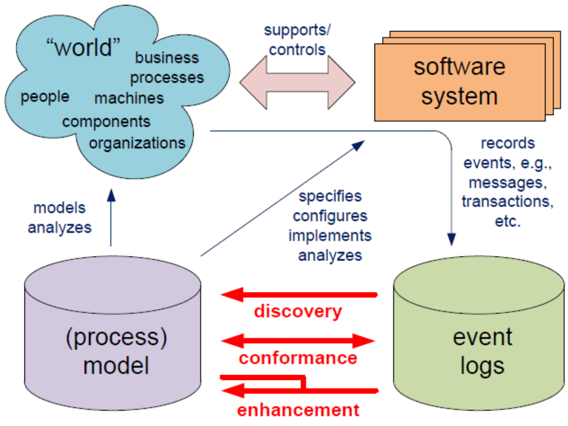
Finally, the insights found with implicit knowledge and explicit criteria are combined. In the final sub-question is researched how these results align.

Q5: *Do the found explicit criteria with machine learning match with implicit knowledge of the decision-makers?*

1.3 Methodology

For decision mining, not a standard framework is provided. However, decision mining uses both process mining and data mining, as will be further explained in chapter 2. Therefore, the frameworks of both these approaches are combined. For process mining an event log is transformed into a process model in three stages; process discovery, conformance checking, and enhancement. These stages and their connections are shown in Figure 1.1. In data mining, the Cross Industry Standard Process for Data Mining (CRISP-DM) is often used. This approach, shown in Figure 1.2, interactively executes the stages; business understanding, data understanding, data preparation, modeling, evaluation, and deployment.

In Figure 1.3, the approach for this research is visualized. Central in this research is the execution of three methods; questionnaire, process mining, and machine learning. In this figure, the input and output for each of these methods is shown. The main approach of the research is based on the CRISP-DM. However, to include the process mining stages as well not the entire CRISP-DM is valid. In this research is started with the business understanding and data understanding. These steps are fundamental for the understanding of the research goal and context. The business understanding includes an explanation of all steps involved in multidisciplinary scale up. In the data understanding incident data is explored and knowledge is acquired with a questionnaire.



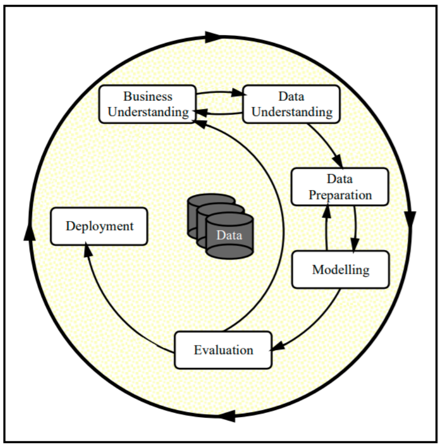


Figure 1.1: Process mining visualization (Van Der Aalst et al., 2011)

Figure 1.2: CRISP-DM (Wirth & Hipp, 2000)

Next, the process model is discovered according to the process mining stages. Therefore, an event log is required that is extracted from the incident data. To discover the process model first different subsets are created from the event log. These subsets are based on the insight gained in the business and data understanding. In the process discovery stage, several miners are fitted on the subsets which are evaluated with conformance checking to select the best-discovered process model.

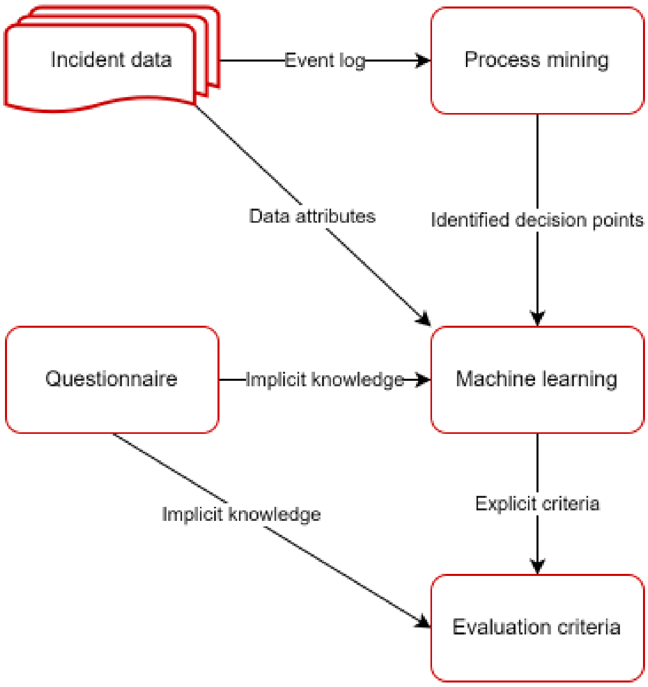


Figure 1.3: Research approach

For the selected process model, the decision points are identified. In these decision points Machine Learning (ML) steps are executed to discover deviation in data patterns that can support decision making. Therefore, first the data is prepared to fit machine learning purposes. New features are added based on the input of the implicit knowledge gathered with the questionnaire. For each decision point models are trained on the prepared data set. These models are compared based on score and explainability.

Finally, the features found with ML are compared with the criteria found during the knowledge gathering. The aim is to evaluate if the model is learning valid features that can be trusted by decision-makers. Therefore, these features are also discussed with the VRU for applicability in multidisciplinary incidents.

1.4 Outline of Thesis

In the following chapters, the sub-questions are answered to gather all the information necessary to answer the main research question as well. First, current methods for decision making in crisis response are evaluated in the literature review of chapter 2. In this chapter also decision mining is evaluated in dept. The definition as well as the possible extensions and applications are reviewed. In chapter 3, the understanding of the research context and the research context of the VRU are discussed. Next, in chapter 4 the available incident data is explored and complemented by gathering implicit knowledge of decision-makers. An event log is constructed to discover a process model in chapter 5. Resulting in decision points that are explored for data patterns with help of ML in chapter 6. For the ML model new features are created based on the implicit knowledge gathered in chapter 4. The important features found with ML are compared with the implicit knowledge as well. Finally, in chapter 7 the conclusion and recommendations for the VRU are summarized. The contribution of this research is to explore the application of the decision mining approach in crisis response. Besides, for the VRU explicit criteria are listed that can be used to distinguish routine incidents from complicated incidents that require multidisciplinary scale up.

Chapter 2

Literature Review

This chapter discusses relevant literature concerning the strategies currently applied in crisis response and the decision mining approach. About crisis response is searched what strategies are relevant to making decisions and the extent of knowledge considered. Furthermore, the decision mining approach is explored for relevance in the crisis management field.

2.1 Crisis Response Methods

Crisis response is recognized in research as the critical phase in crisis management (Shahrah & Al-Mashari, 2017). Mostly because interference directly after the incident occurs is important to protect properties and save lives. For this direct interference to be activated decisions have to be made. The decisions in such situations are found challenging (Shahrah & Al-Mashari, 2017). Therefore, information systems supporting crisis response are extensively explored. However, the complexity of crisis response ensures that most systems are not able to fit the required properties for crisis response. Crisis response processes in general can be very unpredictable and complex as identified in several types of research (Bennet, 2011; Di Ciccio et al., 2015; Kushnareva et al., 2015). The research by Shahrah and Al-Mashari (2017) made an overview of eight different research directions that support Emergency Response Systems (ERSs). These ERSs have to be flexible and scalable, therefore, the explored methods in the research are:

- **Design principles and frameworks**: Due to the complexity and flexibility, ERSs systems require design principles and concepts. Such that the system can support communication and information needs in crisis response.
- **Standardization**: Many researchers have tried to standardize the work in emergency management systems for different aspects. For example, the research by the Incubator Group (2009) concluded an interoperability information framework (Shahrah & Al-Mashari, 2017).
- **Expert systems**: Expert systems are important in the support of decision-making in ERS. Systems that support the needs of experts in the field during crisis response.
- **Agent-based simulation**: Agent-based simulation is a tool well known to help determine the optimal response option in all kinds of situations. Therefore, also applicable for different stages in crisis response.
- **Web technologies**: Enabling the more effective and efficient exchange of information by using web technologies. Also, this direction allows for decision-making support by allowing for communication to make decisions and coordinate actions.
- **Case-based reasoning**: This method focuses on looking for similar incidents in the past to manage current incidents. Starting from the lessons learned from these past incidents to

more efficiently manage new ones.

- **Internet of things**: Recent research discussed how the internet of things has a positive impact on all of the stages in crisis response. The internet of things is based on emergency response systems that supports group decision-making in crisis situations.
- **Business process management**: Combining business process management and workflow techniques has promising results in developing ERSs. These technologies streamline the process and cope with highly dynamic scenarios such as crisis response.

All of these research directions have their own challenges and limitations. Overall, it seems that all of the researches do not manage to deal with the complexity of crisis response (Shahrah & Al-Mashari, 2017). There are some very promising researches such as in business process management to support knowledge work. Expert knowledge in crisis response is still key and not incorporated in these research directions.

2.1.1 Decision Support Systems for Crisis Response

The research by Slam et al. (2015) conducts research in decision support systems for crisis response while emphasizing uncertainty representation, reasoning, learning, and real-time decision-making. Decision support is well known and many articles can be found focusing on applying this system in crisis response. Nevertheless, this research actually focuses on the challenges that have to be addressed for practical application.

The first challenge mentioned by Slam et al. (2015) is 'Knowledge representation and reasoning capabilities'. The uncertainty in crisis response emphasizes the need for representation and reason in crisis knowledge. However, the research identifies the following problem: "Most systems in crisis response either lack an effective knowledge representation scheme or have no reliable inference mechanism that can reason on information with different types and degrees of uncertainties."(Slam et al., 2015). A second challenge mentioned in the same research is the 'Learning capability' of a decision support system. The system should be able to adapt to changes in the environment. However, according to the research, these systems lack learning capabilities to update their knowledge. Furthermore, they identified a challenge in the 'real-time response capability' of decision support systems. In crisis response decisions have to be made immediately to avoid worse. Not much research has been addressed to time-critical characteristics of decision-making. Moreover, the fourth challenge is 'generality'. The decision support system should be applicable to handle different crises with different characteristics. According to Slam et al. (2015), no prior work has achieved a general model that can handle different types of crises and their individual needs.

These challenges are set as requirement by Slam et al. (2015). They constructed a framework in which non-axiomatic logic is applied in representing and reasoning on uncertainty knowledge. The results are promising, but also do not solve all the mentioned challenges. However, there can be concluded that intelligent decision-support in crisis management has some positive contributions.

2.2 Decision Mining

In this section, the meaning and value of decision mining are clarified. Decision mining is an enrichment of the well-known process models used by many organizations (Mannhardt et al., 2016). Process models represent activities and their dependencies in a graph format. According to Rozinat and der Aals (2006), decision mining enriches these models by analyzing how underlying data attributes influence the decision made in the process based on past process executions. The approach they perused will be explained in subsection 2.2.2, together with the implementation tool for decision mining in PROM (Rozinat & van der Aalst, 2006). However, new research enlightens improvements and extensions, which are discussed in subsection 2.2.3. These new insights into decision mining are applied in a diverse application domain as will be elaborated on in subsection 2.2.5.

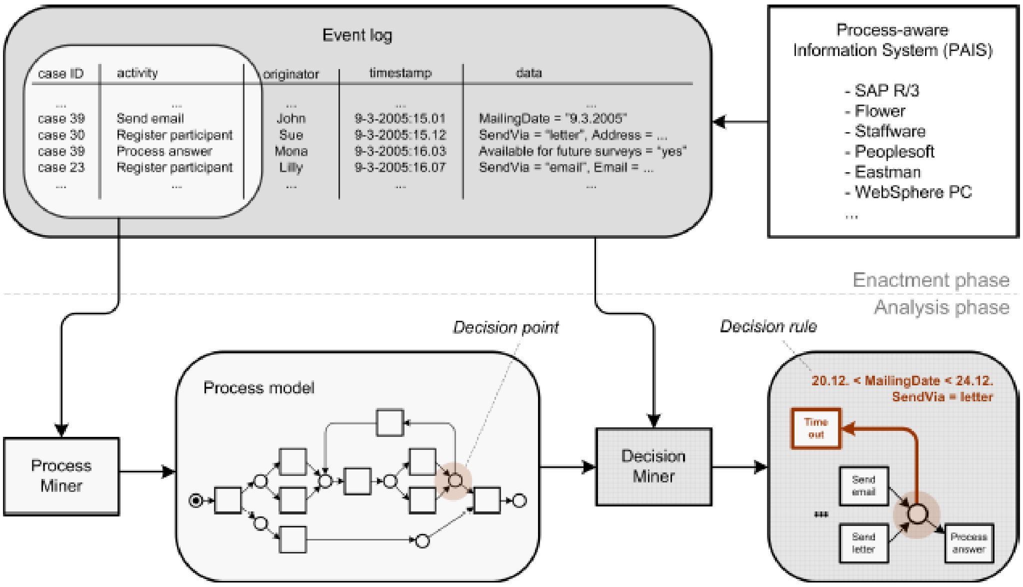
2.2.1 Introduction to Process Mining

Process mining is a technique related to the fields of data science and process management. Event data is used to provide insight into the actually executed process rather than the desired process. The generally applied process mining approach exists of three steps: process discovery, conformance checking, and process enhancement (Van Der Aalst, 2016). The required input to apply this technique is an event log. This event log should at least contain a unique identifier such as a Case ID, an activity that describes the occurring event, and a timestamp (Van Der Aalst, 2016). The sequence of all activities related to one unique identifier is referred to as a trace. In the first step of process mining, process discovery, the main goal is to transform the event log into a process model. For this transformation several techniques are available. The most well-known techniques are the alpha-miner, heuristic miner, and inductive miner. The output of these techniques can differ but a Petri net as output is common. The second step, conformance checking, focuses on evaluating the discovered process model. Therefore, the traces in the event log are compared with the discovered process model. This comparison is often evaluated on four metrics, namely fitness, precision, generalization, and simplicity. The final step of process mining, process enhancement, includes extending the existing process based on the found performance.

2.2.2 Definition of Decision Mining

Research by Rozinat and der Aals (2006) noticed that despite the value of process mining techniques, insufficient attention is paid to how data attributes may affect the routing of a case in the process model. Therefore, they explore the potential of machine learning techniques to gain insight into the data perspective of business processes. Their idea is “to enhance the model by integrating patterns that can be observed from data modifications, i.e., every choice in the model is analyzed and, if possible, linked to properties of individual cases and activities.”. In their approach first, a process model is extracted from an event log. Most process mining techniques construct a process model based on the case ID, activity, and timestamp (Van der Aalst et al., 2003). Such a process model reflects the causal dependencies among activities. However, often resources and additional data are saved in event logs as well (Van der Aalst et al., 2003). Therewith, machine learning can provide insight into distinctive data patterns. Machine learning has become widely adopted to extract knowledge from data attributes as these (Mitchell, 1997). In Figure 2.1 the decision mining approach created by Rozinat and der Aals is visualized.

In the upper left corner of Figure 2.1 an example event log is shown. All the executed activities in a process are stored in an event log (Van der Aalst, 2011). Furthermore, an event log is defined as a collection of unique events (Van der Aalst, 2011). Each row in the event log represents an unique recorded activity execution, together with the corresponding case ID, possible resources



case ID	activity	originator	timestamp	data
...	...	...	...	...
case 39	Send email	John	9-3-2005:15.01	MailingDate = "9.3.2005"
case 30	Register participant	Sue	9-3-2005:15.12	SendVia = "letter", Address = ...
case 39	Process answer	Mona	9-3-2005:16.03	Available for future surveys = "yes"
case 23	Register participant	Lilly	9-3-2005:16.07	SendVia = "email", Email = ...
...	...	...	...	...

Figure 2.1: Decision mining approach according to Rozinat and der Aals(2006)

and related data attributes. In this figure is highlighted that the case ID and activity serve as input for the process miner. In order to use an event log for decision mining purposes, the log needs to contain data attributes as well. These data attributes could related to a specific case or event.

As mentioned the case ID and activity are the least input from the event log for process mining. The process model reflect the behaviour that is collected in the event log, the mining of sequencing patterns (Leewis, Smit et al., 2020). The article by Rozinat and der Aals (2006) chosen is to produce a process model based on the alpha algorithm. This algorithm reconstructs causality from a set of sequences of events (Van der Aalst et al., 2003). The output is a Petri Net as visualization of the process model to identify the decision points.

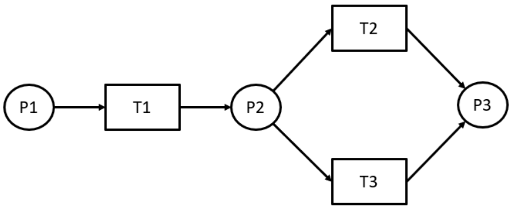
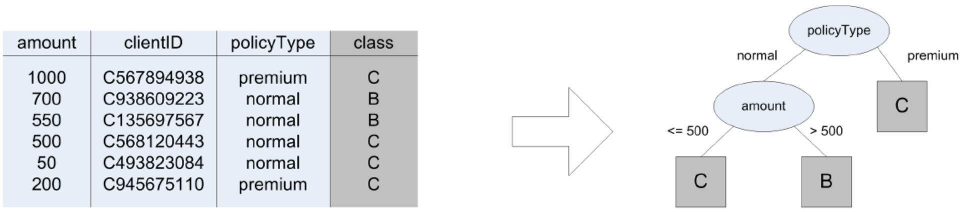


Figure 2.2: Example Petri Net

According to the article by Rozinat and der Aals 2006, in a Petri net a decision point can be recognized as a place with multiple outgoing arcs. Each transition requires a token from all directly connected places. In Figure 2.2 P2 is an example of a decision point. For example, if P2 would contain one token, this token could be consumed by T2 or T3, but not both. In these decision points a choice is made for a process instance. The set of possible choices is provided by the process model that is based on the observed behavior in the event log.

When the decision point is identified in the Petri Net, decision point analysis is conducted to see if decisions might be influenced by case data (Rozinat & der Aals, 2006). More specifically, whether some cases have data attributes that always result in the same routing of a case. Actually this is recognizing structural patterns based on data, in the same way as ML. ML techniques learn structural patterns of data attributes on training instances (Mitchell, 1997). Since the goal of decision mining is to identify underlying decision rules, Rozinat and der Aals concludes that the most obvious algorithm to use are decision trees. Each discovered decision point is a separate classification problem, were we are interested in extracting knowledge about decision rules. Initially, a target class is defined which are the different decisions that can be made. Additionally, the training set is complementd with data attributes that are available at the moment the decision is made. In Figure 2.3 an example is shown provided by Rozinat and van der Aalst (2006), here the data attributes available to learn the decision tree are 'clientID', 'amount' and 'policyType'. The possible target classes are activity B or activity C. In Figure 2.3 b, the retrieved decision tree is visualized.



amount	clientID	policyType	class
1000	C567894938	premium	C
700	C938609223	normal	B
550	C135697567	normal	B
500	C568120443	normal	C
50	C493823084	normal	C
200	C945675110	premium	C

(a) training examples for decision point "p0" (b) decision tree for decision point "p0"

Figure 2.3: Example of classification problem for decision point 'P0' in Rozinat and der Aals (2006)

From this decision tree several decision rules can be formalized. Corresponding to the article of (Rozinat & der Aals, 2006) the rules of Figure 2.3 (b) are formalized with help of the boolean AND and OR operators. "If an instance is located in one of the leaf nodes of a decision tree, it fulfills all the predicates on the way form the foot to the leaf. "(Rozinat & der Aals, 2006). Therefore the following rules can be extracted, an instance will end in class C if ((policyType = 'normal') AND (amount $\leq$ 500)) OR (policyType = 'premium'). An instance will end in class B if (policyType = 'normal') AND (amount > 500). These rules could be applied to make decisions for future cases.

Represented is a general method developed by Rozinat and der Aals providing the steps to analyse these decision mining problems. In addition, they developed a plug-in tool for the ProM framework. This framework includes several tools related to process mining and process analysis. The plug-in by Rozinat and van der Aalst (2006) is called decision miner and is an implementation of the method described above. Of course, there are still a lot of improvements possible and challenges to overcome in order to make decision mining operational for real-life business processes. However, this method provides the initial steps for decision mining.

2.2.3 Directions within Decision Mining

In the recent years, the term decision mining has became prevalent in the business process management field (De Smedt et al., 2016). In the past process mining techniques have proven to be valuable to gain insight into how business processes are executed. In recent years several directions within decision mining are explored. Additionally, several researches focused on improvements and extensions of the current decision mining models. Several important researches in this area are selected and discussed in this section.

Decision Mining Quadrant

The goal of the research by De Smedt et al. (2016) was to define a framework that assesses the definition of decision mining techniques. Therefore, they proposed a framework distinguishing decision mining in two dimensions, namely the control flow and data dimension. Moreover, future additions to this framework are discussed. The proposed framework is designed as a quadrant which is shown in Figure 2.4.

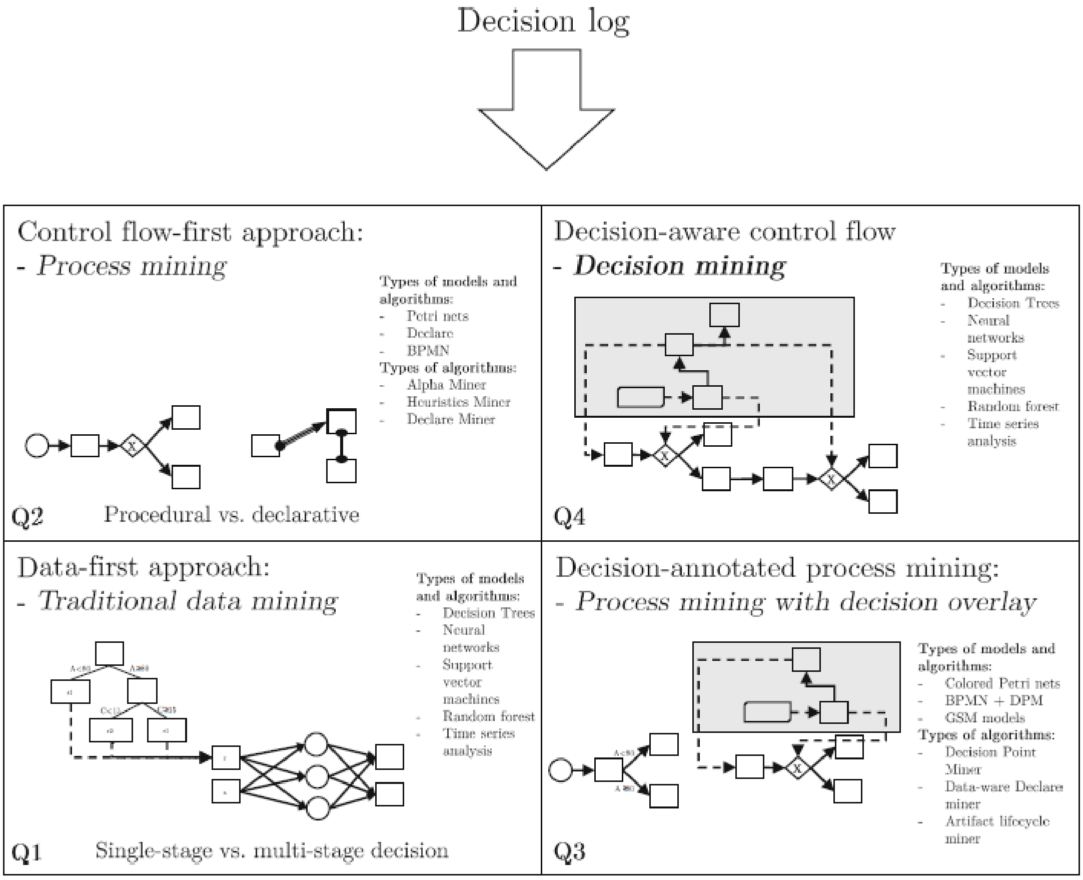

Decision log
Control flow-first approach:
- Process mining
Types of models and algorithms:
- Petri nets
- Declare
- BPMN
Types of algorithms:
- Alpha Miner
- Heuristics Miner
- Declare Miner
X
Q2
Procedural vs. declarative
Decision-aware control flow
- Decision mining
Types of models and algorithms:
- Decision Trees
- Neural networks
- Support vector machines
- Random forest
- Time series analysis
X
X
Q4
Data-first approach:
- Traditional data mining
Types of models and algorithms:
- Decision Trees
- Neural networks
- Support vector machines
- Random forest
- Time series analysis
Q1
Single-stage vs. multi-stage decision
Decision-annotated process mining:
- Process mining with decision overlay
Types of models and algorithms:
- Colored Petri nets
- BPMN + DPM
- GSM models
Types of algorithms:
- Decision Point Miner
- Data-ware Declare miner
- Artifact lifecycle miner
X
Q3

Figure 2.4: Decision mining quadrant (De Smedt et al., 2016)

The framework shown in Figure 2.4 is based on the distinction of two dimensions that will be explained here. On the vertical dimension the decision control flow driving the decision making is displayed. Within this dimension two streams can be separated. On the one hand, the data mining techniques are described as being not aware of any dynamic aspect of the data (De Smedt et al., 2016). On the other hand of the control flow dimension, there is process mining which derives a control flow of activities by fitting process models (De Smedt et al., 2016). According to De Smedt et al. the left hand side of the quadrant decisions within the process are captured implicitly. In contrast with the right hand side, which uses decision models. As in the horizontal dimension referred to as the decision model maturity. These models are beneficial to structure inputs for a decision. The difference between Q3 and Q4 is in the connection between the process model and the decision overlay.

The research by De Smedt et al. (2016) revised the term decision mining and the contrast with process mining as well as data mining. Furthermore, the framework offers an overview of approaches and there differences. For future research this framework allows better targeting for related problems and related research opportunities. In the research three potential opportunities are identified, namely, considerations regarding input data, need for support of available approaches and how new and better techniques can be constructed.

Decision-Annotated Mining and Decision-Aware Mining

According to Leewis, Smit et al. (2020), retrieving and describing the decisions in a process is what is unique about decision mining. In contrast with process discovery which has as main focus the control flow perspective. They conducted a literature review into the current state of the decision mining research field. Exploring related research fields as well as directions within decision mining.

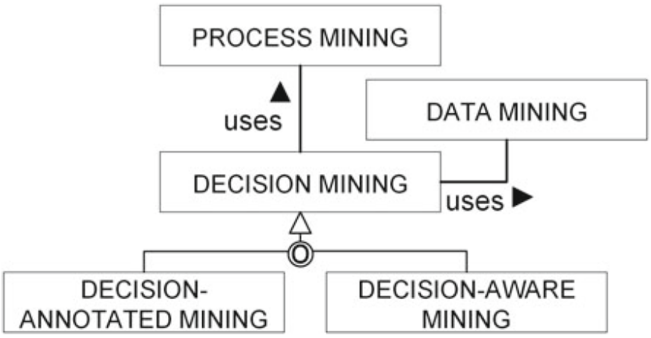


Figure 2.5: Literature relations of decision mining (Leewis, Smit et al., 2020)

Leewis, Smit et al. (2020) retrieved the insight that decision mining is closely related to process mining as well as data mining. In Figure 2.5 these relations are shown, as 'uses', meaning that decision mining uses both techniques from process mining perspective as well as from data mining perspective. Furthermore, the same figure mentions that two directions within decision mining can be distinct, namely decision-annotated mining and decision-aware mining. Decision-annotated mining is focused on mining decision points from business processes (Rozinat & der Aals, 2006) (discussed in subsection 2.2.2) while decision-aware is about taking into account implicit data involved in the decision-making process (De Smedt et al., 2016; Petrusel et al., 2011). Both directions do overlap and utilize process mining as well as data mining techniques.

In data-aware decision mining as presented in Petrusel et al. (2011), aims to create a model of a mental decision making process rather than the physical process. Therefore, other information sources are applicable, since event logs do not retrieve this information. In their approach they use decision-aware software, however, they also imply the option for questionnaires. With this approach it is possible to produce a more objective model which shows what actually happened rather than what users think they have done.

To include both directions within decision mining, Leewis, Smit et al. (2020) provided a new definition for decision mining, namely: "The method of extracting and analyzing decision logs with the aim to extract information from such decision logs for the creation of business rules, to check compliance to business rules and regulations, and to present performance information.".

2.2.4 Extensions

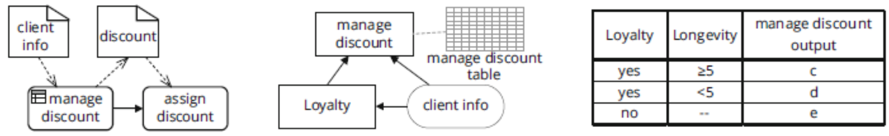
Besides researches into defining directions within decision mining to define the application domain, also several researches are dedicated to extend and improve the decision mining field. These are all extensions on decision mining as it is defined by Rozinat and van der Aalst (2006) as discussed in subsection 2.2.2. Several interesting extensions are highlighted here.

The paper by Leoni et al. (2013) proposes a more general technique to discover branching conditions. The technique combines invariant discovery techniques embodied in the Daikon system with decision tree learning techniques. The foundation of this technique is the decision tree and the Daikon system is an addition. Daikon is a dynamic analysis tool for deriving probable value-based invariants from a collection of execution traces (Ernst et al., 2001). It works by instantiating a set of invariant templates with the variables in the logs and trying to match each instantiated template with the variable assignments recorded in the traces. It yields a set of invariants with sufficient statistical support. To determine which invariants should be combined into branching conditions, they use the notion of information gain from decision tree learning. It can be concluded that this technique allows to detect a wide spectrum of branching conditions from business process executions logs with an increased level of complexity.

The paper by Mannhardt et al. (2016) addresses how existing decision mining methods focus on discovering mutually-exclusive rules. In other words, these rules allow one out of multiple activities to be performed. According to the paper these methods assume fully deterministic decision making and knowledge about all decision influencing factors. However, not all decision situations are framed that way, often have to be worked with incomplete information. "This paper proposes a technique that discovers overlapping rules in those cases that the underlying observations are characterized better by such rules. The technique is able to deliberately trade the precision of mutually-exclusive rules, i.e., only one alternative is possible, against fitness, i.e., the overlapping rules that are less often violated."This method start similar as previous decision mining methods, with an initial decision tree based on observations from the event log. Subsequently, the misclassified instances of each decision tree leaf are used to learn a new decision tree that leads to new rules. These new rules are used in disjunction with the original rules yielding overlapping rules of the form $rule_1 \vee rule_2$. The proposed technique is also evaluated by Mannhardt et al. (2016) on two real-life data sets. Results show that discovering overlapping rules improve the balance in terms of fitness and precision. Furthermore, an implementation of the technique is made available within the MultiPerspectiveExplorer package in the process mining framework ProM (Mannhardt et al., 2015).

In 2013 De Leoni and van der Aalst highlight how discovered control-flow models do not fully conform to the event log. Caused by the fact that infrequent observations are threaded as noise and discarded. Therefore, they wrote a paper to notify this problem while using the recent advances in conformance checking using alignments for a new approach. In this approach, the first step is to discover the process control-flow. Then, the control-flow and event log are aligned, thereby mitigating the effects of non-conformity. After the alignment is computed, they discover the data-flow perspective where transitions guards are required. In this paper the focus is on discovering these guards with help of ML techniques. This technique is quite similar to the decision mining approach discussed earlier. However, an important complement is that this approach includes alignments.

Both Batoulis et al. (2015) and Bazhenova and Weske (2016) have explored the options to extend decision mining by combining Business Process Management and Notation (BPMN) and Decision Model and Notation (DMN). Based on the argumentation that decision logic should be modeled separately from process logic. In this case process logic is captured in BPMN and decision logic is captured in DMN. The aim of the paper by Batoulis et al. (2015) is to extract decision logic form process models for which they introduce a semi-automatic approach. This approach "identifies decision logic in process models, to derive a corresponding DMN model and to adapt the original process model by replacing the decision logic accordingly, and to allow final configurations of this result during post-processing."(Batoulis et al., 2015). An example of this approach is shown in Figure 2.6. The paper by Bazhenova and Weske (2016) starts with the same foundation and the knowledge from Batoulis et al. They propose a four-step approach to derive decision models from process models. First, they identify decision points in a process model. Followed by extracting decision logic to define the data dependencies affecting the decisions in the process model. Next, a decision model is constructed and finally the process model is adapted according to the derived decision logic. Another contribution they make is to measure Key Performance Indicators (KPIs) while deriving the decision logic.



Loyalty	Longevity	manage discount output
yes	≥5	c
yes	<5	d
no	--	e

(a) Decision task without branching behavior.

(b) Adapted DMN model.

(c) Merged decision table.

Figure 2.6: Example of post-processing from Batoulis et al. (2015)

Finally, in 2015 Mannhardt et al. develeped a new tool for ProM, namely Multi-perspective Process Explorer (MPE). This tool is based on multi-perspective process mining techniques which go beyond the techniques that use only event sequences to analyze the control-flow. Multi-perspective includes data attributes attached to these events as another perspective from data encoded in event attributes. The MPE "integrates existing work on multi-perspective process mining with new interactive visualizations and filtering facilities into a scalable and extensible tool."This tool provides the opportunity to integrate existing data-aware discovery, conformance checking and performance analysis. The MPE works with four steps, starting with analysis of the input model. In this step the fitness is calculated with help of alignments. In the second step, the guards are discovered. These can be discovered with for example the use of decision trees. In the third step the performance is evaluated as well as possible bottlenecks in the process. In the final step, a detailed analysis can visualize the results, with a possibility to look for specific traces. This tool is tested on real-life cases and has reached a high level of maturity, which makes it promising to use in new real-life cases.

2.2.5 Application Domain

The decision mining approach is applied in several application fields. However, it is remarkable that most of these application domains have similar characteristics. Decision mining is most often applied to clear and structured processes, because the data quality is quite high in this situations. However, in real-life cases this is not always the case. Table 2.1 compares the application domains of several papers. Furthermore, the approaches used by these papers is also shown. Notice how multiple papers use the same application domain. Especially loan application in the financial sector and liability claims in the insurance company, are examples of cases that have a straight forward process without to complex structures. A process is more complex if it contains invisible activities, duplicate activities or loops. Most approaches are not able to handle these complex structures while they occur often in real-life cases (Mannhardt et al., 2016).

The research by Campos et al. (2017) did include knowledge-intensive processes (KiP) which are characterized by being dynamic and unstructured processes. This paper compares the extraction of rules from structured data and unstructured data analyzed with the MPE. The do this for ICT services that differ for different problems. According to the research, the activity 'AddNote'requires interaction with the customer for context. This results in open text, which is the unstructured data that needs time consuming human interpretation to extract rules. The extracted rules from both the structured and unstructured data are discussed with experts, therefrom can be concluded that the recognize more value in the rules from unstructured data than from the structured data. They understand and are not surprised by the rules from the structured data, but gather more insight from the unstructured data.

Table 2.1: Overview of application domains discussed in papers

References	Loan application bank	Road traffic fine	Pathway of patient in hospital	Liability claim insurance company	ICT services	Emergency response management system	Approach
Rozinat and der Aals (2006)				X			Decision-annotated mining.
Rozinat and van der Aalst (2006)				X			Decision-annotated mining.
Smirnov et al. (2007)						X	Profile modeling and profile-based decision mining.
Petrusel et al. (2011)	X						Decision-aware mining with use of a decision data model (DDM).
De Leoni and van der Aalst (2013)	X						ProM data flow discovery.
Bazhenova et al. (2016)	X						Decision-annotated mining, with identifying data decisions and decision dependencies.
Bazhenova and Weske (2016)	X						Combining BPMN and DMN to seperate process and decision logic.
Mannhardt et al. (2016)		X	X				ProM MPE, with decision trees in decision trees.
Campos et al. (2017)					X		ProM MPE, with discovery data perspective and text mining.
Mertens et al. (2020)			X				Declarative process model with decision mining.

2.2.6 Challenges in Applying Decision Mining

Despite the recent developments in decision mining, there are still a lot of challenges to overcome. Some of these challenges are recurring in each data analysis activity, namely data availability and quality. Other challenges are problem specific like how to handle invisible activities and loops. In this section all of these challenges and there impact on decision mining are discussed.

Good data quality is necessary to actually be able to say something about the results of your model (Leewis, Berkhout et al., 2020; Rozinat & der Aals, 2006). However, good data quality is not granted. It starts with the data gathering. How and what data is stored is key for the data quality. The data can be stored structured or unstructured, variables or open text. Besides, sometimes not everything that should be stored is actually stored, often referred to as noise. Of course this can be do to a lot of reasons, but it is impossible to assume data is always complete. Moreover, the interpretation of data attributes needs human reasoning. Therefore, data preprocessing is a crucial step before applying any process mining or data mining tool. Additionally, parameter tuning can help with to overcome some data quality issues. However, important is to check the data quality before starting to apply modelling since only good data can tell something about the actual process.

A second challenge addressed by Rozinat and der Aals (2006) relates to the correct interpretation of control-flow of a process model to classify the decisions. While in most examples used in papers simple business processes are used, these are not necessarily representative for real-life processes, these real-life processes contain invisible activities, duplicate activities and loops that complicate the process. They also provide guidance how to handle these challenges. Invisible activities should be traced until the next visible activities. A side note is to stop this tracking as soon as a join construct is encountered. In that case alternative paths cannot be specified and are discarded from the analysis. To deal with duplicate activities is similar, that is, to track the succeeding activities until either an unambiguous activity or a join construct is encountered. Dealing with loops is more complicated and no concrete solution is proposed. However, Rozinat and der Aals (2006) show how to distinguish parts of a loop. Figure 2.7 visualizes the loop and the corresponding distinction points. Point (a) is a decision point contained in a loop, “Multiple occurrences of a decision related to this decision point may occur per process instance, and every occurrence of B and C is relevant for an analysis of this particular choice. This means that one process instance can result in more than one training example for the decision tree algorithm.”(Rozinat & der Aals, 2006). For point (b), decision point containing a loop, applies “Although a process instance may contain multiple occurrences of activity B and C, only the first occurrence of either of them indicates a choice related to this decision point.”(Rozinat & der Aals, 2006). Finally, point (c) decision point that are loops, “This choice construct represents a post-test loop (as opposed to a pre-test loop), and therefore each occurrence of either B or C except the first occurrence must be related to this decision point.”(Rozinat & der Aals, 2006).

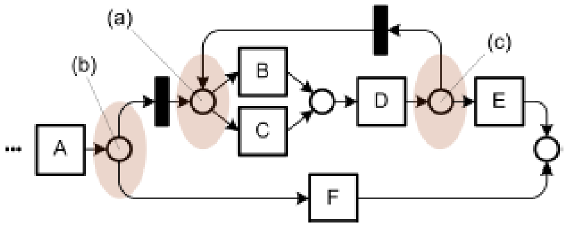


Figure 2.7: Example process model with loop from Rozinat and der Aals (2006)

2.3 Chapter Overview

From literature can be concluded that current crisis response strategies have not accomplished a conclusive strategy. Due to the challenge of the complex nature of emergency response, no strategy is found adequate to efficiently support all the required capabilities of emergency response. The most promising according to Shahrah and Al-Mashari (2017) is BPMN since it may support knowledge work. The literature is divided into researches focus on Artificial intelligence (AI), information technology and the cognitive process of human decision making (Slam et al., 2015). Especially that last category is assumed to be important for decision making in this research.

Decision mining is previously used to make implicit knowledge explicit (Petrusel, 2010) and to discover business rules (Campos et al., 2017). This is focusing on deviation in data patterns to support decision making. The approach combines techniques of process mining and data mining. Besides, several approaches within decision mining are identified by De Smedt et al. (2016). These approaches differ in the integration of process mining and data mining. The basics are the same, a process model is discovered and decisions within this model are discovered with data mining. Most often the decision tree algorithm is used for data mining. Because from this algorithms it is easy to extract decision rules.

In this research the decision mining approach is explored to capture implicit knowledge of expert decision-makers as explicit criteria, that can be applied to decision making in crisis response. The current crisis response strategies have acknowledged the importance of the knowledge of decision-makers, while they still fail to incorporate this proportionally. Decision mining has proven to discover decision rules in several research fields. However, crisis response is not yet one of these application domains. Nevertheless, decision mining is applied in domains with dynamic and unstructured processes by Campos et al. (2017), similar to the crisis management field. Therefore, this research will explore how decision mining can be applied in crisis management to explore explicit criteria for incident handling.

Chapter 3

Business Understanding

Before we can apply decision mining, it is important to fully understand the context in which this research is executed. This chapter will discuss different aspects of the VRU that are significant for this research. Firstly, the emergency dispatch center is explained, which is the place where incidents are first reported and the first disciplines are assigned to the incident. In addition, the VRU uses a scale-up procedure referred to as GRIP, we will elaborate on this procedure in this chapter. The decision to use the GRIP scale-up for a certain incident is based on human judgment. Therefore, the tasks and knowledge of these decision-makers are discussed, as well. Finally, the objectives of the VRU within crisis response are highlighted and put into context for this research.

3.1 Research Context

This section provides a description of the subsequent steps taken after an incident occurs. This process always starts with an incident notification at the emergency dispatch center. Subsequently, disciplines are sent to the incident. At this point, a split is made between incidents that can be handled by the disciplines at the location and the incidents that require further coordination. For the second alternative, several scale-up options are available from assigning more disciplines to the incident till GRIP. These steps are separately elaborated in the following subsections; emergency dispatch center, GRIP, and the decision-makers.

3.1.1 Emergency dispatch center

The emergency dispatch center is the place where all incidents are reported. All disciplines have their separate dispatch center, working from the same room. When an incident is reported, the call is directed to a corresponding dispatcher. The job of the dispatcher is to sketch the incident situation and assign vehicles accordingly. This process is straightforward as long as it is a mono incident without complex characteristics. However, incidents can be complex and involve multiple disciplines, these types of incidents are called multidisciplinary incidents. For this kind of situation, a Calamiteiten Coördinator (Calamities Coordinator) (CaCo) is present at the emergency dispatch center.

In general, the duty of the CaCo is to coordinate multidisciplinary incidents within the emergency dispatch center. Their daily responsibility is to calculate risks and consult senior dispatchers about occurring incidents. That task often starts with monitoring interesting incidents. What is defined as interesting is subjective and based on what the CaCo on duty thinks is worth monitoring. For example, an incident in which little information is known for a long time or an incident that has an incident type that evolves into a larger incident more often. In that case, the CaCo gets in touch with the dispatchers of the involved discipline. The CaCo gathers more information about the situation and assesses if he/she can help to coordinate the situation. The main responsibility of the CaCo is to keep an overview of the incident situation to act upon whenever necessary.

If an incident is of large scale and requires additional coordination, the CaCo has the authority to scale up to GRIP. The dispatchers of each discipline keep in contact with their Officier van Dienst (Duty officer) (OvD) (explained in subsection 3.1.2) at the incident location to exchange information. The CaCo has contact with the different dispatchers to exchange the information they gather about the situation at the incident location and available information from other dispatchers. Generally speaking, the CaCo maintains the overview of all information gathered by the different disciplines.

3.1.2 GRIP

For the multidisciplinary incidents that require additional coordination, GRIP is designed. GRIP is the abbreviation for Coordinated Regional Incident Management Procedure. This procedure involves who should do what, when, and at which location, to collaborate to control an incident as soon as possible.

In the regional crisis plan, the GRIP levels with corresponding operations are defined (VRU, 2020). These levels range from 0 to 4, as shown in Table 3.1. Incidents reported at the emergency dispatch center are routine incidents most of the time (GRIP 0). A routine incident can be mono- or multi-disciplinary depending on the scale of the incident. However, an incident that starts as routine may evolve over time. As mentioned in subsection 3.1.1, the CaCo has the task to monitor these incidents from the emergency dispatch center. On the incident location, OvDs are responsible for coordination. The OvD is the leader of the mono discipline and is responsible for coordinating their units at the location. Each discipline has its own OvD, which is alerted as soon as multiple units of their discipline are involved. Their coordination task involves assuring alignment between disciplines and scale-up to GRIP if necessary. The OvDs stay in contact with their dispatcher. If needed, in doubtful situations, the CaCo has contact with the leaders and coordinators of the incident to share relevant information.

Previously explained coordination assumes a specific incident location. However, an incident does not necessarily have a location. Also without this specific location, an incident can still require scale-up to GRIP, for example, COVID-19. An overview of the different GRIP levels and their general description are shown in Table 3.1. The distinction between different GRIP levels is qualitative and focuses on the desire for coordination at an operational or organizational level. The difference between operational and organizational coordination is important to consider, as it emphasizes respectively the short term and long term coordination.

Table 3.1: Description of GRIP levels (VRU, 2020)

GRIP level	Operational team	Description situation
GRIP 0 (routine)	-	Daily work mono- or multi-disciplinary without additional coordination
Operational level		
GRIP 1	Commando Plaats Incident (Commander Site Incident) (CoPi)	Multidisciplinary coordination is required at the incident location
GRIP 2	Regionaal Operationeel Team (Regional Operational Team) (ROT)	Multidisciplinary coordination is required from an external location or preparation for an incident
Organizational level		
GRIP 3	Gemeentelijk Beleidsteam (Municipal Policy Team) (GBT)	Multidisciplinary coordination is required which influences organizational tasks within the municipality
GRIP 4	Regionaal Beleidsteam (Regional Policy Team) (RBT)	Regional multidisciplinary coordination is required which influences organizational tasks

For each GRIP level, there is a designated operational team appropriate for that specific GRIP level. Figure 3.1 shows which operators are included in the teams at each level. For GRIP 1 this team is known as Commando Plaats Incident (Commander Site Incident) (CoPi). Composing of all OvDs, Informatie Manager (Information Manager) (IM), Communicatie Adviseur (Communication Advisor) (CAC) and possible situation depending support of liaisons. This team, CoPi, keeps operational by further scale-up to GRIP 2. The team is complemented with the Regionaal Operationeel Team (Regional Operational Team) (ROT). These are all operational coordinations. When support from the organizational level is required there is a scale-up to GRIP 3 with a Gemeentelijk Beleidsteam (Municipal Policy Team) (GBT). If the appearance is larger than the municipality then the GBT is replaced by the Regionaal Beleidsteam (Regional Policy Team) (RBT) at GRIP 4.

Different GRIP levels are not sequential but related to the required coordination. For example, it is possible to go immediately to GRIP 2 or immediately to GRIP 3/4, without having GRIP 1 first. The same holds for downscaling, decided by the highest leader in the GRIP level at that moment.

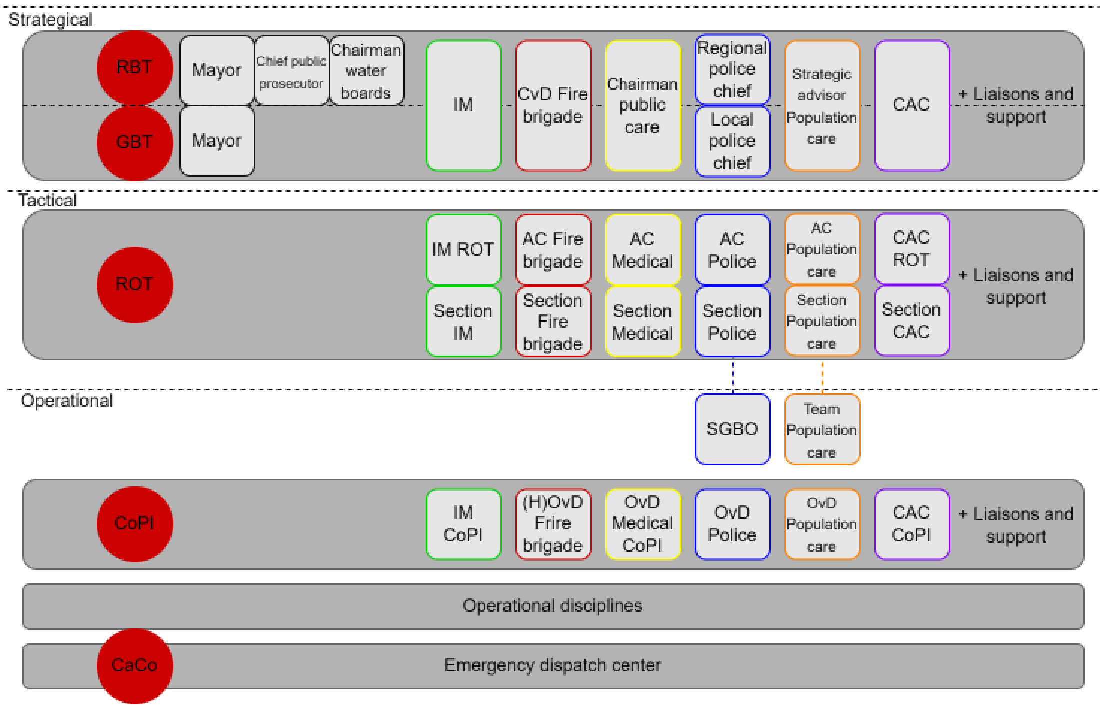


Figure 3.1: Overview operational teams (adjusted from VRU 2020)

Flexible Scale-Up

In chapter 1 flexible scale-up is mentioned which is a concept that the VRU wants to realize. Currently, each GRIP level includes a standard operational team that is alarmed. However, each incident has specific characteristics that determine who could be meaningful to include in the procedure. The scale-up processes used in flexible scale-up depend on the situation, so it can be seen as a customized scale-up to the incident. This alternative scale-up could be especially appropriate for incidents that are classified as routine (GRIP 0) now but are multidisciplinary and some disciplines could use additional processes to resolve the incident.

The associated challenge is that decision-makers have more decision options to properly evaluate. They have to judge every situation with specific characteristics correctly and assign the most appropriate people, that have value to resolve the incident. Another challenge is future incidents that should be classified as GRIP are first scaled up flexibly. The associated risk is that decision-makers decide on flexible scale-up as a safe choice because there is less risk of assigning the wrong people. However, GRIP is drawn up for a good reason and should be used as well.

Nevertheless, implementing flexible scale up promises great improvements to the current system. It results in more suitable incident response. An adequate response would logically decrease the incident duration. There are only around 20 GRIP incidents per year, but many more where there is doubt if scale-up to GRIP is necessary. The doubt is derived from feedback from operators in the past and the insight that not all processes of the scale-up are necessary. For all of these incidents, flexible scale-up would be a solution to obtain effective and efficient incident control.

3.1.3 Decision-Makers

The authority to scale up to GRIP is the responsibility of human decision-makers. For each GRIP level, there are other decision-makers in charge of scale-up. Table 3.2 provides an overview of who is authorized for scale-up at what GRIP level. The distinction between scale-up for operational level and organizational level is recognized here as well, namely in the functions authorized as decision-maker.

For scale-up to operational level (GRIP 1 and GRIP 2) the decision-makers are mostly operational commanders, contributing to the incident for their discipline. Therefore, the decision-maker is colored by the perspective of their discipline, possibly resulting in not adequately considering the needs of other disciplines when making a decision. In addition, operators tend to lose themselves in the heat of the incident, losing the ability to zoom out and look at an overview of the situation. This is a natural human response, however, this clouds their judgment when making decisions about the scale-up. Nevertheless, they are in the position to observe the incident and therefore gain the most information. From that point of view, they are the most informed to make decisions. That is the reason, why they are authorized to scale up. The CaCo is in a position to create a more general and less biased overview of the situation, because fewer emotions are involved. This is a better position for a decision-maker to be in.

For scale-up, to organizational level (GRIP 3 and GRIP 4) the decision-makers are not directly involved at the incident location, but they are accountable for the operational execution and decisions. For scale-up to GRIP 3, this is the mayor of the municipality corresponding to the incident, and for scale-up to GRIP 4 this is the chairman of the VRU. These scale-ups are only relevant when an incident has long-term effects. Therefore, decisions to organizational scale-up are made differently than decisions to operational scale-up. Considered is the impact on the specific municipality or region which is known best by their mayor or the chairman of the VRU. Challenging is that they are not directly involved in the incident, but are informed when needed.

Overall, all decision-makers are in a position where they have the ability to judge the situation from their own perspective. Therefore, it is necessary that they are able to scale up. Nevertheless, for the operational scale-up levels, there are so many decision-makers that it is hard to guide them to the same image for when scale-up is appropriate. Particularly considering every week different people are on duty for these functions. Therefore, guidelines could help to provide a general approach to make these decisions.

Table 3.2: Entitled to scale-up

	GRIP 1	GRIP 2	GRIP 3	GRIP 4
Entitled to scale-up	Mayor ROL of service (H)OvD Fire Brigade OvD Medical OvD Police OvD Population Care CaCo	Mayor ROL of service (H)OvD Fire Brigade OvD Medical OvD Police OvD Population Care Leider Copi CaCo	Mayor	Chairman VRU

3.2 Organization Objectives

Additional to the described process in the research context, the main objectives of the VRU and its crisis management department are important. The research should be in line with the ambitions of the organization. Since the VRU is a government organization, they have other interests than a company would have. The general objectives the VRU is addressing on a daily basis are:

- Promote safety and where possible prevent insecurity.
- Combat incident when they do occur.
- Provide help and support to victims.
- Limit suffering and (health) damage.
- Make an effort to quickly repair disturbances of daily life.

The department where this research is executed, crisis management, is mostly concerned with adequate incident response. For this cause, they are constantly working on procedure development for all different kinds of incidents. For example, they have procedures available for incidents at the train station and incidents combating hazardous substances. These are plans made as preparation for a possible incident, also referred to as the cold phase. Next to the cold phase, there is the operational deployment when an incident occurs, the warm phase. To bridge the gap between the cold phase and the warm phase, the VRU has introduced the VIC. This information center monitors and analyses based on four elements; political, personnel, public, and press, with the aim to act if necessary. In practice, this means they are using many (mainly) open sources to monitor possible risks. They gather information before and during an incident delivering a broader image of the situation.

The interest of this research is in the objectives of the operational deployment, also known as the warm phase. All general objectives are somehow relevant in incident control. However, the main focus of this research is on supporting adequate decision-making in order to quickly repair disturbances of daily life, focused on multidisciplinary incidents. During the research it should be kept in mind that this support of decision-making is understandable and usable for all decision-makers.

3.3 Research Scope

The research goal is to determine objective criteria that indicate the most appropriate scale-up based on patterns in the data. Scale-up in this research is defined as a decision for operational scale-up focusing on incidents involving multiple disciplines. In other words, incidents involving only one discipline, called mono incidents, are excluded from this research since this scale-up is limited to the number of vehicles. The research is not interested in optimizing the number of vehicles related to an incident. The research is interested in defining criteria for the decisions that have to be made related to flexible scale-up, GRIP 1, and GRIP 2, as defined earlier in section 3.1. Scale-up to GRIP 3 and GRIP 4 are excluded from the scope of this research because these levels involve organizational scale-up and decisions.

In addition, the scope is partially determined by the availability of data. The VRU is fire brigade related, therefore, emergency dispatch center information of only the fire brigade is available. Information about the same incident, involving other disciplines, is filtered out by the system. This information is not accessible for this research. However, there is data available that indicates which disciplines are involved in the incident. A limitation caused by this missing data is that there are no timestamps available related to the police and ambulance. Therefore, the event log will not be complete. It is possible to explore the extent of the data sources if information is needed to create specific features based on information gathered with the questionnaire.

Chapter 4

Data Understanding

The goal to capture implicit knowledge in explicit criteria with decision mining requires gathering data of implicit decision making as well as already existing data. In this data first the historical data used for this research is discussed in Figure 4.1, section 4.2 and section 4.4. Besides, the first general data preparation steps are taken in section 4.3. For gathering implicit data a questionnaire is distributed among all different decision-makers. The focus of this questionnaire was to understand and gather the knowledge decision-makers take into consideration while deciding to scale up or not. This questionnaire is explained in section 4.5.

4.1 Data Sources

The directly available data sets contain data submitted by dispatchers in Geïntegreerd Meldkamer Systeem (Integrated Control Room System) (GMS) and this data is stored in the Veiligheidpaspoort (VP) database. In Figure 4.1 a visualization is represented. This database contains divers information about the incidents. For this research three of these data sets are retrieved. The basic VP data set and general incident data set contain incident-specific data. Where the deployed vehicles data set contains vehicle-specific data.

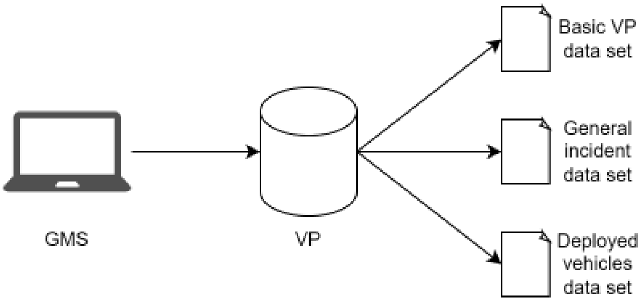


Figure 4.1: Data flow visualization

The incident-specific data is logged in GMS by the dispatchers at the emergency dispatch center after receiving a phone call about an incident. The dispatcher enters a new incident in GMS. The time of logging is automatically added by the system. In the meantime, the dispatchers ask specific questions about the incident on hand. The information from the call is entered in a structured way based on three, single-word, classification criteria. Classification criteria 1 describes the incident type, whereas classification criteria 2 is more specific and classification criteria 3 contains even more detail. For example, classification 1 is 'Accident', followed by classification 2 'Road-transport' and classification 3 'Injury'. These criteria generally describe the incident. Based on these criteria GMS provides a predefined vehicle proposal, which can be adapted for this specific incident by the dispatcher. The vehicle proposal is also depending on the location of the incident. Locations can be automatically tracked from the phone call or specified by the reported person. All the available information about the location is logged in GMS.

When vehicles are assigned to an incident, the corresponding data is stored separately. Important notification is that the available data for vehicles is limited. The VRU only has access to the data in VP that is connected to the fire brigade. This includes monodisciplinary incidents for the fire brigade, but also multidisciplinary incidents to which the fire brigade is assigned. For these multidisciplinary incidents, data of other disciplines only the new data is available that is assigned after a mono incident becomes multidisciplinary.

4.2 Data Description

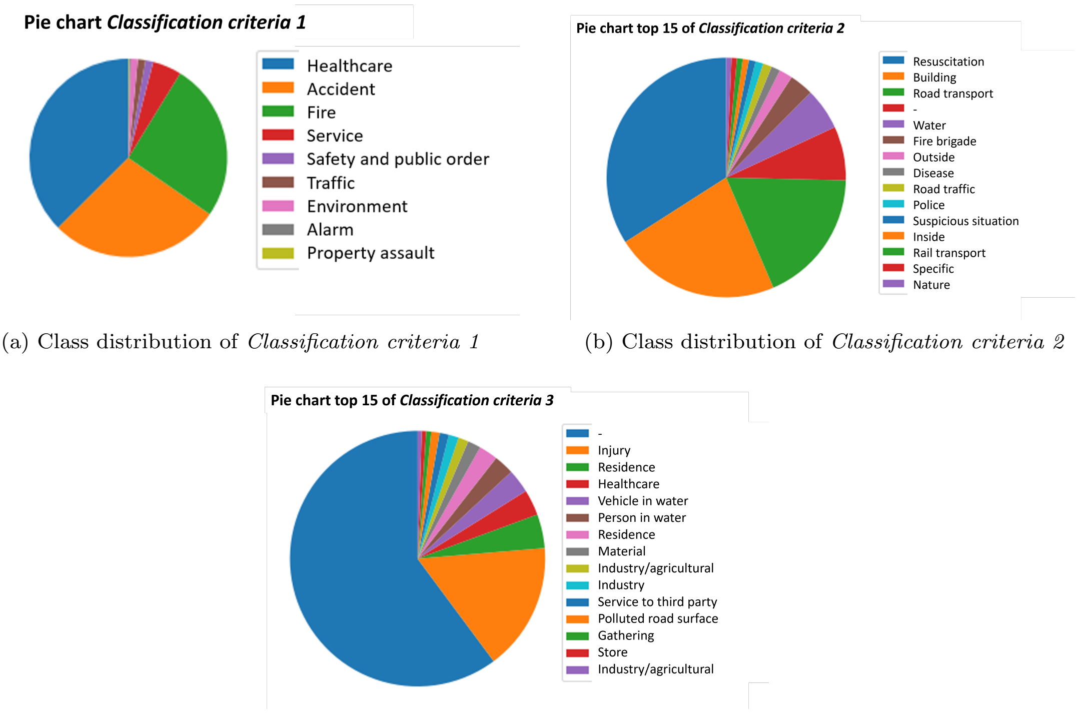
As described in the previous section, three data sets are used. In this section, the valuable information from each of these data sets is described and discussed. Since the basic VP data set and general incident data set are both incident-based, these data sets contain overlapping attributes. Meaning that both data sets have the same attributes such as *Classification criteria 1*, *2* and *3*. On the one hand, the basic VP data set contains 43,905 rows with incidents over a period from January 1 2014 till October 10 2020. On the other hand, the general incident data set contains 47,237 rows with incidents over a period from January 1 2015 till December 31 2019. An overview of the attributes in the basic VP data set is shown in Table 4.1. The attributes of the general incident data set are shown in Table 4.2.

Table 4.1: Description of basic VP data

Attribute Name	**Example**	**Type**
Incident number	12569	Integer
Start time incident	2016-01-13 20:16:00	Temporal
Street name	'Lekdijk-West'	String
Postal code	'2861EV'	String
House number	105	Integer
City	'Bergambacht'	String
Municipality	'Bergambacht'	String
Classification criteria 1	'Brand'	Categorical
Classification criteria 2	'Gebouw'	Categorical
Classification criteria 3	'Woning'	Categorical
Priority	1	Ordinal
OMS	0	Integer
Object incident	'Onbekend'	String
Object report	'Woonfunctie'	String
Type location	'S'	Categorical
Intervention type	'Brandbestrijding'	Categorical

Each row of the basic VP data and general incident data describes a new incident. This incident has a yearly unique *Incident number*, which starts with *1* on January first of each year. As a result, different years contain the same incident numbers in the data. At the same moment, a new incident is logged, the date and time are registered in *Start time incident*. For the location there are multiple attributes, namely *Street name*, *House number*, *Postal code*, *City*, *Municipality*.

This *Street name* can also contain interesting information about type kind of road, for example, when an incident occurs on the highway. In addition, coordinates of the incident are available in the general incident data. As an addition to these location attributes, the attribute *Type location* describes the incident location. The values for this attribute are 'S' (street), 'R' (rails), or 'W' (water). For the description of the incident type the categorical features *Classification criteria 1*, *Classification criteria 2* and *Classification criteria 3* are useful. All criteria are described with a single word and relate to each other. *Classification criteria 1* is a quite general description, whereas *Classification criteria 2* is more specific and *Classification criteria 3* contains even more detail. In Figure 4.2 the pie diagrams show a selection of the fifteen most mentioned values these criteria have. Another feature describing the incident is *Priority*. This feature rates the emergency on a scale from 1 to 5, where 1 is the highest priority and 5 is the lowest. Furthermore, the attributes *Object incident* and *Object report* contain information about the function of the incident location. *Intervention type* contains information about the needed intervention. The possible values for this attribute are 'Combat water accident', 'Combat fire', 'Serves', 'None', 'Assistance' and 'Combat hazardous substances'. The attribute *OMS* represents the code of the automatic alarming system the dispatch center was alarmed by, which is zero if there was no automatic alarm.



(a) Class distribution of *Classification criteria 1*

(b) Class distribution of *Classification criteria 2*

(c) Class distribution of *Classification criteria 3*

Figure 4.2: Exploration of classification criteria

The general incident data set in Table 4.2 contains some additional attributes as well. The primary interest for this set is in the timestamps and the involved disciplines. The timestamps are crucial for the process mining steps in Figure 1.1. *Timestamp function* and *Timestamp partner* both contain information about the moment a specific function or partner organisation is assigned to the incident. The *Function* are multidisciplinary functions from the VRU. For instance, these could be communication advisor or information manager. The attributes *Fire brigade involved*, *Police involved* and *Medical involved* are binary attributes, with values 'Yes' or 'No'. Together these attributes describe if an incident is multidisciplinary, namely as all values are 'Yes'.

The final data set used shows the deployed vehicles at an incident, for which the interesting attributes are shown in Table 4.3. The data set contains 73,904 rows in which vehicles are assigned

Table 4.2: Description of general incident data

Attribute Name	Example	Type
Start time incident	2016-08-31 15:58:27	Temporal
End time incident	2016-08-31 17:51:53	Temoral
Incident number	240202	Integer
Fire brigade involved	'Ja'	Categorical
Police involved	'Ja'	Categorical
Medical involved	'Nee'	Categorical
Priority	1	Integer
Classification criteria 1	'Brand'	Categorical
Classification criteria 2	'Natuur'	Categorical
Classification criteria 3	'Heidebrand'	Categorical
Street name	'Doornseweg - N227'	String
House number		Integer
City	'Leusden'	String
X coordinate	155021	Integer
Y coordinate	448725	Integer
Municipality	'Leusden'	String
Partner organisation	'Boswachters'	String
Timestamp partner	2016-08-31 15:58:27	Temporal
Function	'Operationeel Woordvoerder VRU'	Categorical
Timestamp function	2016-08-31 16:02:45	Temporal

to incidents from the period January 1 2015 till December 31 2019 . The new interesting attributes of this data set are *Vehicle type* and *Alarm time vehicle*. The *Vehicle types* in the data are mainly fire brigade related like 'TS' which stands for water tender, and 'RV' for rescue vehicle. However, also the multidisciplinary functions are submitted in the deployed vehicles data set, namely 'OvD Fire Bridage', 'OvD Medical', 'OvD Police', 'OvD Population Care', 'IM', 'CAC', 'GRIP 1' and 'GRIP 2'. Their alarming time is submitted in *Alarm time vehicle*.

Table 4.3: Description of deployed vehicle data

Attribute Name	Example	Type
Incident number	21546	Integer
Vehicle priority	1	Integer
Classification criteria 1	'Brand'	Categorical
Classification criteria 2	'Gebouw'	Categorical
Classification criteria 3	'Kantoor'	Categorical
Vehicle type	'TS'	Categorical
Start time incident	2016-01-04 16:46:00	Temporal
Alarm time vehicle	2016-01-04 16:46:00	Temporal
End time incident	2016-01-04 19:10:00	Temporal

4.3 Data Integration

In order to apply process mining, an event log is required. This is realized in this section by integrating the different data sets and creating activities based on the timestamps in the data. In an event log, each row represents a new activity with a timestamp. The *Case ID* is the *Incident number*, one incident can have multiple activities. Besides, as many attributes as possible are added for context understanding in the event log.

Before integration is possible the *incident numbers* for all data sets must be updated. In the raw data, these numbers are unique for a single year. However for adequate integration, these should be unique for all years to avoid a data merge from different years of information. Therefore, the current *Incident numbers* are combined with the year. As an example, *Incident number* '2546' in the year 2015, becomes '20152546'. With these new unique numbers, it should be possible to integrate the data sets.

According to the research scope, the aim is to analyze the decisions in multidisciplinary incidents. Therefore a selection is made of which incidents include the assistance of the fire brigade, police, and medical team. In the general incident data, this information is available. The attributes *Fire brigade involved*, *Police involved* and *Medical involved* category should be 'Yes' in all cases that are multidisciplinary. The incident numbers that meet this criterion are labeled as multidisciplinary incidents. Following, the basic VP data and the deployed vehicle data are filtered to contain only these incident numbers that are multidisciplinary, as well. To ensure completeness of the data points, the data sets are checked on incident numbers that are not part of the general incident data. These incidents are excluded from the final data set.

Now the multidisciplinary incidents are selected, the data can be transformed to an event log. Therefore, activities are created based on the timestamps of an incident numbers. There are three options to match activity names with their timestamp. First, the column name is also the activity name. This is the case for *Start incident*, *Partner organisation* and *End incident*. Second, the value of the attribute *Function* is the activity name, which is the case for *OvD Medical*, *GRIP 1* and *GRIP 2*. Third, the value of the attribute *Vehicle type* is the activity name. This results in the activities *IM*, *CAC*, *OvD Fire brigade*, *OvD Police* and additional activities for *OvD Medical*. An example of the event log is shown in Table 4.4.

Table 4.4: Example event log

Incident Number	**Timestamp**	**Activity**	**Attributes**
201513899	2015-01-15 17:10:00	Start Incident	...
201513899	2015-01-15 17:10:00	Partner organization	...
201513899	2015-01-15 17:23:00	OvD Fire Brigade	...
201513899	2015-01-15 17:37:00	GRIP 1	...
201513951	2015-01-15 17:40:00	Start Incident	...
201513899	2015-01-15 17:55:00	IM	...
201513951	2015-01-15 18:54:00	OvD Fire Brigade	...
201513951	2015-01-15 21:27:00	End Incident	...

The following attributes are added to the event log, mostly based on the basic VP data. Therefore, these attributes are incident-based and unique for an incident number. Added to all rows with that incident number. A list of added attributes:

- Classification criteria 1
- Classification criteria 2
- Classification criteria 3
- Priority
- Municipality
- City
- Object incident
- Object report
- Type location
- Intervention type

4.4 Data Quality

The integration results in an event log with 18,349 activities and 6,866 unique multidisciplinary incidents. In this section, the quality of this event log is evaluated by evaluating missing values and other data errors. Therefore, the possible categories of the categorical features are tested for overlapping categories. Furthermore, the traces in the event log are explored for unrealistic traces and completeness of traces. In addition, the activities in the event log are evaluated.

4.4.1 Missing Values

To start, Table 4.5 represents the missing value count of the available attributes in the event log for unique incidents. The classification criteria are represent for all incidents, however, sometimes *Classification criteria 2* and *Classification criteria 3* are filled with '-'. This is not considered a missing value since this sign means that no information of this classification criteria is known or that all knowledge is already summarized in the other classification criteria. The *Municipality* and *Priority* attributes are complete as well. Incomplete is *City*, for approximately 4 % of the incidents this attribute has the value 'Unknown'. The value 'Unknown' is kept in place and considered as a separate category for all incidents where the *City* is unknown. Enough data is available to consider this attribute. For the *Object incident* and *Object report* this is not the case, since they have 98 % and 56 % missing values respectively. However, there seems to be an overlap in the meaning of these attributes. Therefore, the option to merge these attributes is evaluated in section 6.2, there the added value is also tested. *Type location* has 4.5 % incidents without a type since the location type is in 95 % of the cases 'S' of street. It seems logical to fill the missing values with the mode. The *Intervention type* has a higher amount of missing values, namely almost 37 %. For these, the constant value 'Unknown' is imputed.

Table 4.5: Missing values event log

Attribute Name	Missing Values	Percentage of Missing Values
Classification criteria1	0	0 %
Classification criteria 2	0	0 %
Classification criteria 3	0	0 %
Priority	0	0 %
Municipality	0	0 %
City	295	4.3 %
Object incident	6763	98.5 %
Object report	3859	56.2 %
Type location	308	4.5 %
Intervention type	2535	36.9 %

4.4.2 Data Errors

Additionally to missing values, there are categorical features that need data cleaning. The attributes *Classification criteria 2* and *Classification criteria 3* for instance have categories with the same meaning but spelled differently. Besides, during the years some changes are made to the criteria names. All resulting in categories with the same meaning, which should be covered by manually renaming these features into one category. The number of categories of the raw data is shown in the second column of Table 4.6. After reducing there is some reduction in categories as can be seen in the third column.

An example from criteria 2 is that the category 'Autom. Gev. Stof' is spelled in three different ways. These categories are renamed to fit one category. In some years of the categories from criteria 3, the category names started with a number like '02'. In other years this number is not used, while the same category is mentioned. All of these categories are renamed with general category names.

Table 4.6: Categories of categorical attributes

Attribute Name	**Number of Categories before Data Cleaning**	**Number of Categories after Data Cleaning**
Classification criteria1	9	9
Classification criteria 2	49	40
Classification criteria 3	85	65
Priority	5	5
Municipality	55	55
Type location	5	4
Intervention type	6	6

4.4.3 Event Log Exploration

Now the missing data and data errors are assessed, the event log should be evaluated on the quality. The event log is a selection of all activities related to multidisciplinary incidents. The activities and a count are shown in Table 4.7. This table can be used to check if all traces in the data are complete. As mentioned earlier, there are 6,866 unique incidents and 18,349 rows in the data set. The trace of each incident should at least contain the event 'Start incident' and the event 'End incident'. In Table 4.7 the occurrence of these events is equal to the number of unique incidents, so this condition is met. Furthermore, the total number of rows should be and is in agreement with the sum of all activities in the event log.

When looking at the count of activities there are a few interesting insights. The OvDs of the different disciplines differ high in the number of activities. This can partly be declared by the available data that is fire brigade related. Unfortunately, gathering more data is impossible due to privacy regulations. During the process discovery step in chapter 5 this imbalance should be taken into consideration. Also, notice here that there is no data related to the OvD Population Care. Another interesting insight is the number of IM activities. This number is lower than the sum of GRIP 1 and GRIP 2 activities. While it would be expected that there are more incidents where an IM activity is usable than where a scale-up to GRIP is performed.

Table 4.7: Event log activity count

Activity Name	**Count**
Start Incident	6,866
Partner Organization	1,683
CAC (Communication advisor)	401
IM (Information manager)	89
OvD Fire Brigade	2,160
OvD Police	15
OvD Medical	178
GRIP 1	81
GRIP 2	10
End Incident	6,866
Total activity count:	**18,349**

4.5 Implicit Knowledge Gathering

All data discussed in this chapter so far is historical data. In this research the interest in criteria for decision-makers requires not only knowledge of the past decisions, but also an insight on why these decisions were made. In decision-annotated mining as used in Rozinat and van der Aalst (2006), explicit criteria are searched in historical data, only by looking for data patterns. This same method is applied in the thesis of Theeuwes (2019) by searching for human knowledge from a data perspective only. However, another decision mining approach is decision-aware. According to Petrusel et al. (2011), knowledge acquisition is a requirement for decision-aware decision mining. This decision mining technique is considering the knowledge and intuition humans take into account while making a decision. After which, the features are created from the historical data that represent this human knowledge. Possible options to acquire knowledge are decision-aware software, observations and questionnaire. Since decision-aware software is outside the reach of this research and observations are time consuming. Chosen is to use a questionnaire for this research. With a questionnaire the opinion of multiple decision-makers can be considered providing a general view of their considerations from different perspectives. The aim of this questionnaire is to gain insight in the criteria decision-makers mark as important while they make their decision. Additionally, the questionnaire can be used to validate the assumption that decision-makers from different positions make choices based on different criteria.

The group of decision-makers contacted for the questionnaire exists of all people in position to scale-up to GRIP 1 and/or GRIP 2. In total approximately 150 people are contacted to complete the questionnaire. The people contacted operate in the following positions: OvD Fire Brigade, OvD Police, OvD Medical, OvD Population Care, CaCo, leader CoPi/(H)OvD Fire Brigade, Regionaal Operationeel Leider (Regional Operational Leader) (ROL) on duty.

4.5.1 Questionnaire Setup

For the questionnaire we chose to use three real past crisis situations that decision-makers evaluate and make a decision based on this evaluation. All three situations required different handling. One situation requires extensive coordination but no GRIP. Another situation clearly needs scale-up to GRIP. The last situation is doubtful since it is not a regular GRIP but still requires some of the aspect of GRIP. All of these crisis situations are retrieved from 'Lessen uit crises en mini crises'(van Duin & Wijkhuijs, 2017; van Duin et al., 2018; Wijkhuijs & van Duin, 2020). A first selection is made after which the proposed crisis situations are discussed thoroughly with experts of the crisis management department. They confirm that the three situation descriptions match with the intention for no GRIP, certain GRIP, doubt GRIP. The situation descriptions and questions can be found in Appendix B.

In the questionnaire each situation consists of two parts representing different moments in time. The first part consists of incident notification information followed by questions. The first questions asked to the decision-maker is if multidisciplinary scale-up is appropriate, in this situation with the current information. Furthermore, an explanatory statement for this decision and possible game changers, are asked. Since, more information is gathered during the incident that might change the decision, we included a second part as well. The second part starts providing complementary information about the incident. Where the first part focuses only on information available at the emergency dispatch center after the notification, the second part contains also information gathered by the disciplines arrived at the incident location. This might change the decision perspective since these disciplines can make a better assessment of the situation resulting in additional information. The first question of the first part is repeated, to see if the judgement of the decision-makers changes with the additional incident information. In addition, the followup question asks the decision-maker to select points of attention from the situation description. In this question the decision-maker has the possibility to summarize which criteria they think they consider. In the next question criteria are provided based on themes in Landelijk Centraal Meldkamer Systeem (National Control Room System) (LCMS). Eight themes are covered divided into several criteria each. In Table 4.8, the themes and corresponding criteria are shown. The decision-makers have the possibility to check as many of the criteria they consider for this specific incident situation.

Table 4.8: Overview of LCMS themes with corresponding criteria

Theme	**Corresponding Criteria**
Incident	Incident type Incident size Expected duration incident Incident location Timestamp incident
Risks and Safety	Possible Involvement of Hazardous Substances/Smoke Possible perpetrators/fleeing suspects Possible effects on people/material Safe/unsafe area Safe/unsafe approach route Applicable (safety) procedures
Meteo	Wind direction / wind speed Temperature Rainfall / clouds
Victims/population	(Number of) injured and injured classification Population specifications (for example: in a certain neighbourhood) Reception locations/relatives
Environmental analysis	Vulnerable objects in the environment (for example: nursery/town hall) Care objects Events/demonstrations and/or other activities Municipal or regional border crossing incident Busy location/flow-through location
Communication	Sensitivity incident on social media Use NL-alert and/or other means of communication Action perspective communication
Services involved	Own disciplines involved Involved partners (also nationally)
Missing information	Cause of incident Duration of incident unknown Planning and scenarios unknown

4.5.2 Questionnaire Results

The questionnaire is sent by email to approximately 150 people of which 71 responded and completed the questions. For each group of decision-makers, approximately half of this group responded. The correspondence is visualized in Figure 4.3. The larges response came from the OvD Fire Brigade with 21 responses. Followed by the OvD Police and OvD Population Care with 16 and 12 respectively. From the OvD Medical, CaCo and HOvD/Leader CoPi there where 6 people that completed the questionnaire. Finally, 3 decision-makers in as ROL and 2 people closely related to this procedure responded as well.

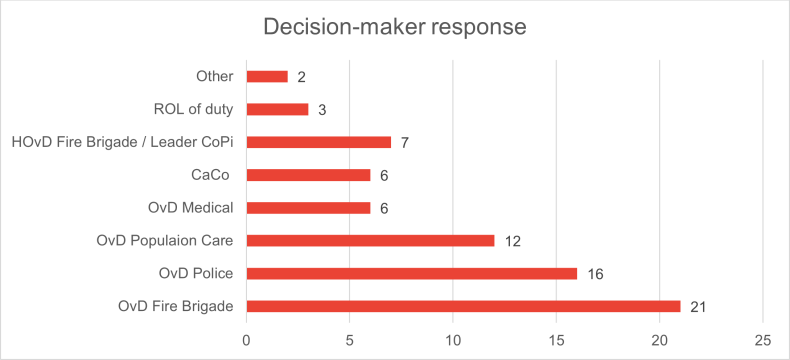
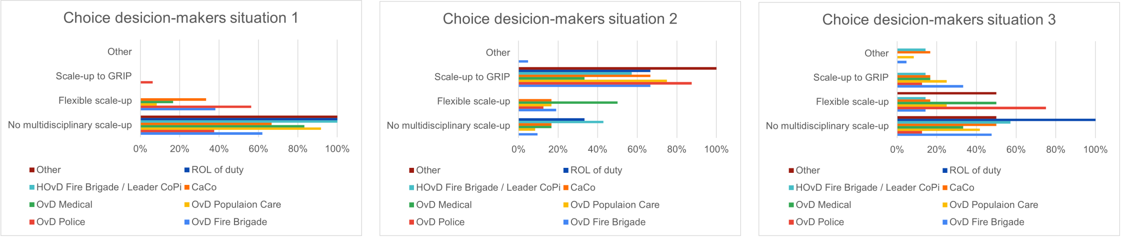


Figure 4.3: Response of decision makers

Problem Validation

In subsection 3.1.3 is stated that different decision-makers make different choices based on their perspective of the situation. The questionnaire asks these different decision-makers to evaluate three situations and make the choice for scale-up or not. Only in the questionnaire all of these decision-makers evaluate the same situations. In normal incidents a selection of one person of each group is involved in this decision. An important difference between scale-up in reality and in the questionnaire, is that decision-makers are specifically asked to make this decision where in reality this choice moment is less obvious since they have other tasks as well.



(a) Crisis situation 1

(b) Crisis situation 2

(c) Crisis situation 3

Figure 4.4: Decision for multidisciplinary scale up

In Figure 4.4 the responses of all decision-makers based on the notification description are visualized. In all of the crisis situations there is variation in the decisions of the different decision-makers. For situation 1, most decision-makers agreed with the actual incident decision that no multidisciplinary scale up was required. However, more than 50% of the OvDs Police would have chosen for flexible scale-up for this incident. This is remarkable more than than the other decision makers. In the second crisis situation, Again the majority of the decision-makers makes the same decision as the actual incident to scale up to GRIP 1. However, this time for the OvDs Medial more than 50% preferred flexible scale up above scale up to GRIP 1. Besides, it should be noticed that more than 40% of the HOvD Fire brigade/Leader CoPi chooses for no scale up at all. In the third crisis situation, more deviation in decisions is visible. This result was intended and expected, since the described situation was doubtful and not a regular GRIP incident. Notice that the majority of the ROLs, leader CoPi, CaCo, OvD Population care and OvD Fire brigade chose for no multidisciplinary scale up. While the majority of the OvD Medical and OvD Police chose for flexible scale up.

The spread in answers confirms the earlier stated assumption by the VRU that the decision for multidisciplinary scale-up depends on the decision-maker. Some of these variations is also possible due to the varied interpretation of ‘Flexible scale-up’. Despite the effort to explain the interpretation at the start of the questionnaire. These results cannot confirm the influence of implicit knowledge. However, when taking the context into account this seems appropriate.

Insights in Decision-Making Criteria

The main reason for using a questionnaire is to define criteria that decision-makers use. For each situation the decision-makers were asked to mention the criteria they were taking into account. In Figure 4.5 the count of the mentioned criteria is shown. From this figure, it seems that all criteria are considered, positively or negatively. From the figure can be concluded that the criteria are situation depending. Therefore the three different situations were included in the first place.

The criteria of Figure 4.5 considered in all three crisis situations are the most general. Since we are interested in criteria that are suitable for other situations as well, we focus on these general criteria. The criteria found important by decision-makers in all three situations are: ‘Incident location’, ‘Incident size’, ‘Incident type’, ‘(number of) injured and injury classification’, ‘Sensitivity incident on social media’ and ‘Own disciplines involved’. These all have more than half of the votes in all three situations. Some other criteria that have an higher than average score are: ‘Expected duration incident’, ‘Possible effects on people/material’, ‘Safe/unsafe area’, ‘Involved partners’ and ‘Duration of incident unknown’. These criteria are discussed with the VRU. They recognize these insights from practice.

The previous mentioned criteria already provide insight into the consideration of decision-makers, but their own proposed criteria are evaluated as well. Selected are the criteria that provide additional insight outside the already listed criteria. One of the comments mentioned in the open text questions is that the incidents requires more structure, which could be accomplished by multidisciplinary scale up. Additionally, the comment is made that there are a lot of different tasks to fulfil and therefore scale up can help to coordinate these tasks. Overall, the comments made seem in line with the listed criteria. Several other comments are made, but these seem incident specific rather than general applicable.

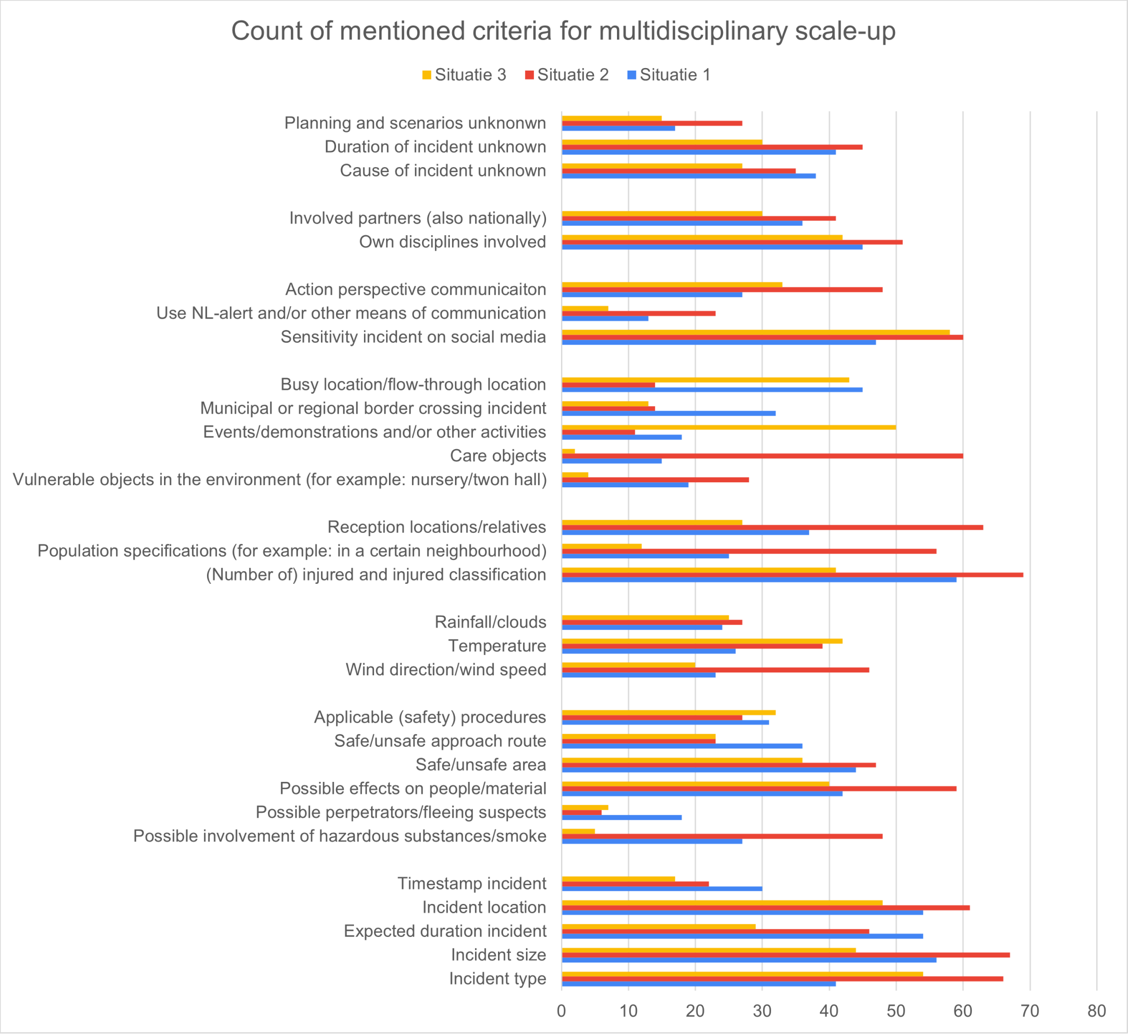


Figure 4.5: Considered criteria by decision-makers

4.6 Chapter Overview

In this chapter an overview is provided of the collected data. This data exists of knowledge acquisition, and historical data which is both incident-based and vehicle-based. The attributes of the historical data are explored and described. For this research only the multidisciplinary incidents are selected, meaning that the fire brigade, police and medical team are involved by the incident. The selected historical data is transformed into an event log. This event log is required for the process mining steps in chapter 5. Furthermore, the data quality is explored after which several categories where merged by renaming, because the registration names changed over the years. In the questionnaire, the knowledge of decision-makers is gathered, regarding decisions in multidisciplinary crisis response. The criteria considered most by all different decision-makers are: ‘Incident location’, ‘Incident size’, ‘Incident type’, ‘(number of) injured and injury classification’, ‘Sensitivity on social media’, ‘Own disciplines involved’, ‘Expected duration incident’, ‘Possible effects on people/material’, ‘Safe/unsafe area’, ‘Involved partners’ and ‘Duration of incident unknown’.

For the upcoming chapter, chapter 5, the created event log serves as input. The retrieved attributes in the historical data and the found implicit criteria are input for chapter 6.

Chapter 5

Process Discovery with Process Mining

Process discovery is the first stage of decision mining. The goal is to discover the actually executed multidisciplinary scale-up process. Such a process model is not yet available by the VRU. Several process mining techniques are possible for process discovery from a provided event log. The visual representation of the process gains insight into the sequence of multidisciplinary scale-up activities. In chapter 4 the event log is already created. In this chapter subsets are created, several models are created and these models are evaluated. Then one of these models is selected to continue the research and identify decision points.

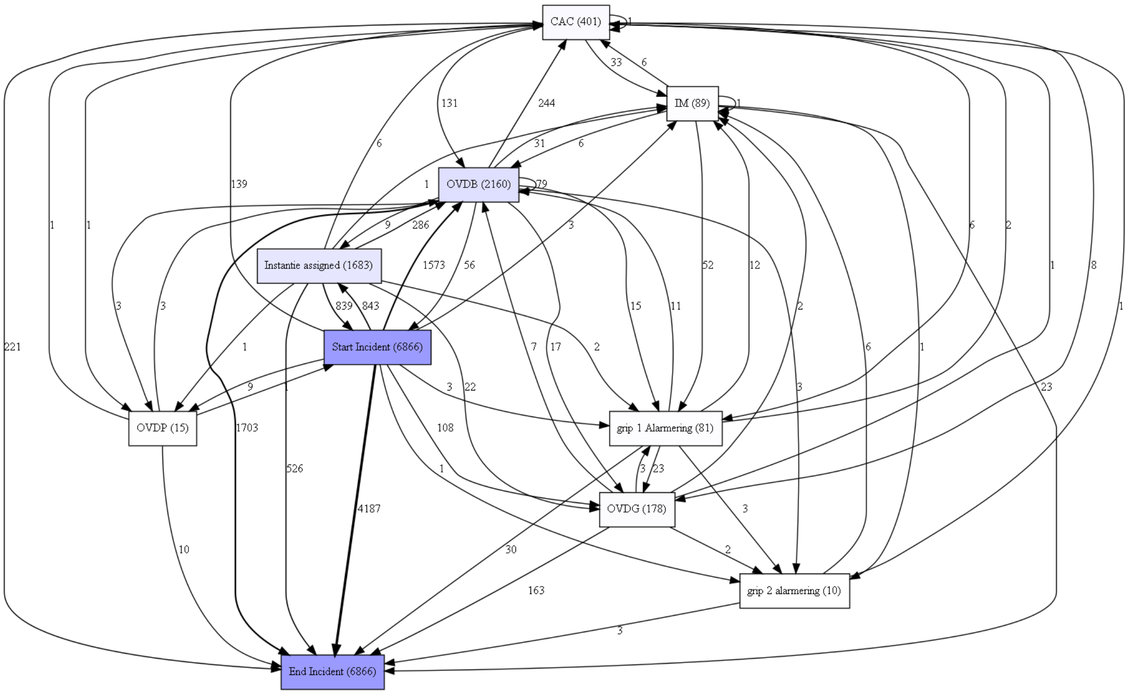


Figure 5.1: Directly follow graph multidisciplinary scale-up activities

Before all of that, in Figure 5.1 the event log is visualized in a directly follow graph to understand the first relations between activities. A directly follow graph is the simplest representation of the process model. Each node is an activity and each arc represents the relation between the attached activities. In the figure, all activities in the event log are shown together with the number of times the activity occurs. The arcs connecting these activities show how often the relation occurs as well. Overall, this graph provides insight in the deviation between often occurring sequences and very specific sequences that only occur ones or twice.

5.1 Subsets

The event log created in chapter 4 requires pre-processing to remove irrelevant data for the process discovery aim. There are three basic pre-processing techniques that allow for filtering in fundamentally different ways (Fahland, 2021). The first one is selection, where specific traces in the subset are selected. Each trace contains the same, but the total number of traces is reduced based on a condition. The second filtering technique is projection. In each trace events are removed that do not satisfy a condition. So selection is a reduction of traces and projection is a selection of events in traces. The third filtering operation is aggregation, where subsequent events in a trace are merged into one event. A projection of the multidisciplinary related events is already been made when the event log is created in chapter 4.

The first filter applied is a selection of all traces that start with ‘Start incident’. In Figure 5.1 can be seen that the event ‘Start incident’ has incoming as well as outgoing arcs. The incoming arcs represent preceding events. In theory, it is assumed that an incident always starts with the notification, which is the meaning of ‘Start incident’. In that case, the event would not have any incoming arcs. When analyzing the incoming arcs, two possible causes are identified. Therefore, the directly follow graph with performance metrics is consulted, included in Appendix C. The incoming arc from ‘Instantie assigned’ has a mean duration of 0 seconds, as well as the outgoing arc to ‘Instantie assigned’. Therefore, the explanation for the incoming arc is that these events are executed simultaneously. Meaning that a partner organization is assigned immediately at the start of an incident. Resulting in registration in half of the traces that ‘Instantie assigned’ is executed before ‘Start incident’. The incoming arcs from ‘OvD B’ (OvD Fire Brigade) and ‘OvD P’ (OvD Police) are evaluated with the same graph in Appendix C. There is shown that the time between these activities and ‘Start incident’ is less than 1 minute. The events ‘OvD B’ and ‘OvD P’ are extracted from a different data set as ‘Start incident’, as a result, there might be some slight differences in the rounding of the timestamp. Since the time between these events is less than 1 minute it is assumed that this is the cause for the unexpected order of events in the trace. Selecting only the traces that start with ‘Start incident’ ensures that the indented order remains. This will be the first subset.

In the second filter that is applied some activities are renamed to generalize the process model. In this case, the separate events ‘OvD B’,‘OvD P’, and ‘OvD G’ are renamed as one general event ‘OvD’. The reason is that the kind of OvD is depending on the incident type. For example, with a large fire, there are many fire brigade units at the incident location so an OvD Fire Brigade is there to coordinate these units. However, for an incident with many injured people, an OvD Medical is at the incident location to coordinate their units. Nevertheless, all OvDs are there for the same purpose, to coordinate their units and ensure collaboration with other disciplines. Taking this into account, the general event ‘OvD’ is created. This will be the second subset.

The subsets described above are compared to the initial event log. Besides, a third subset is created where both filters are combined. Resulting in four subsets in total for which models are created and compared. The first subset is the complete event log created in chapter 4. The second subset is the selection of the start event 'Start incident'. The third subset is the general renaming of ‘OvD’. Finally, the fourth subset is the combination of subsets two and three.

5.2 Process Discovery

In this section, process discovery algorithms are fitted to the event log to receive process models. Different miners have different outputs. The most often used miners are used for this research: the alpha miner, heuristic miner, and inductive miner. Each model is fitted on all four subsets.

The alpha miner (Van der Aalst et al., 2004) is one of the first and best-known algorithms to discover processes. The algorithm looks at the ordering relation in all traces of the event log and creates a footprint matrix. From this, a Petri net with unique events is created. Even though this method is well known there are some cons. The algorithm is not able to discover any loops (of length 1 or 2), invisible activities, or duplicate activities. Furthermore, the model is not necessarily sound, meaning that some traces will not end in the final marking. But most importantly the algorithm does not handle noise well. For the event log on hand, this might be a problem, because incident data is not necessarily well structured. However, this miner is still fitted since its popularity and quite simple approach for discovery.

The Heuristic miner (Weijters et al., 2006) acts on the directly follow graph. By doing so the heuristic miner is able to find a common construct and is better at handling noise compared to the alpha miner. The output of the heuristic miner is a heuristic net of which an example is shown in Figure 5.2. A heuristic net contains information about the relationship between activities. This net can be converted to a Petri net. A benefit of the heuristic miner compared to the alpha miner is that it takes frequency into account and detects short loops. Similar to the alpha miner it does not ensure a sound model. For applying the miner in PM4PY it is possible to tune several parameters regarding the dependency and occurrences. However, to start the default of these parameters is applied.

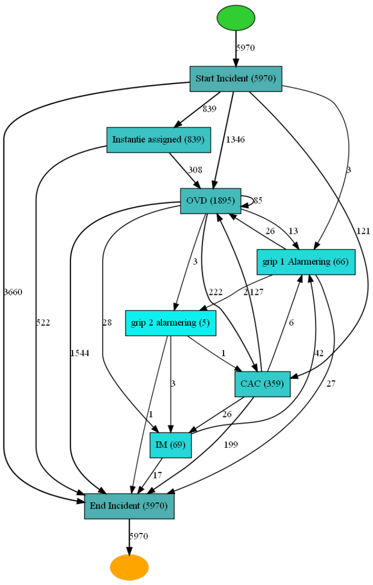


Figure 5.2: Heuristic net

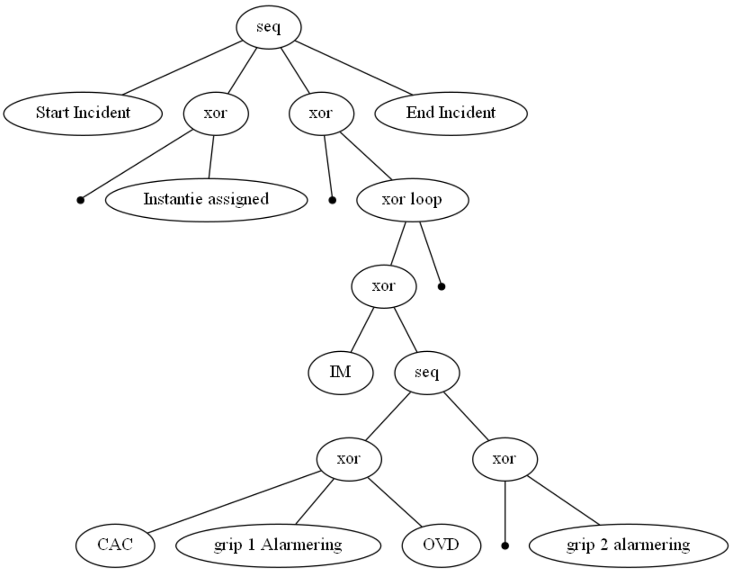
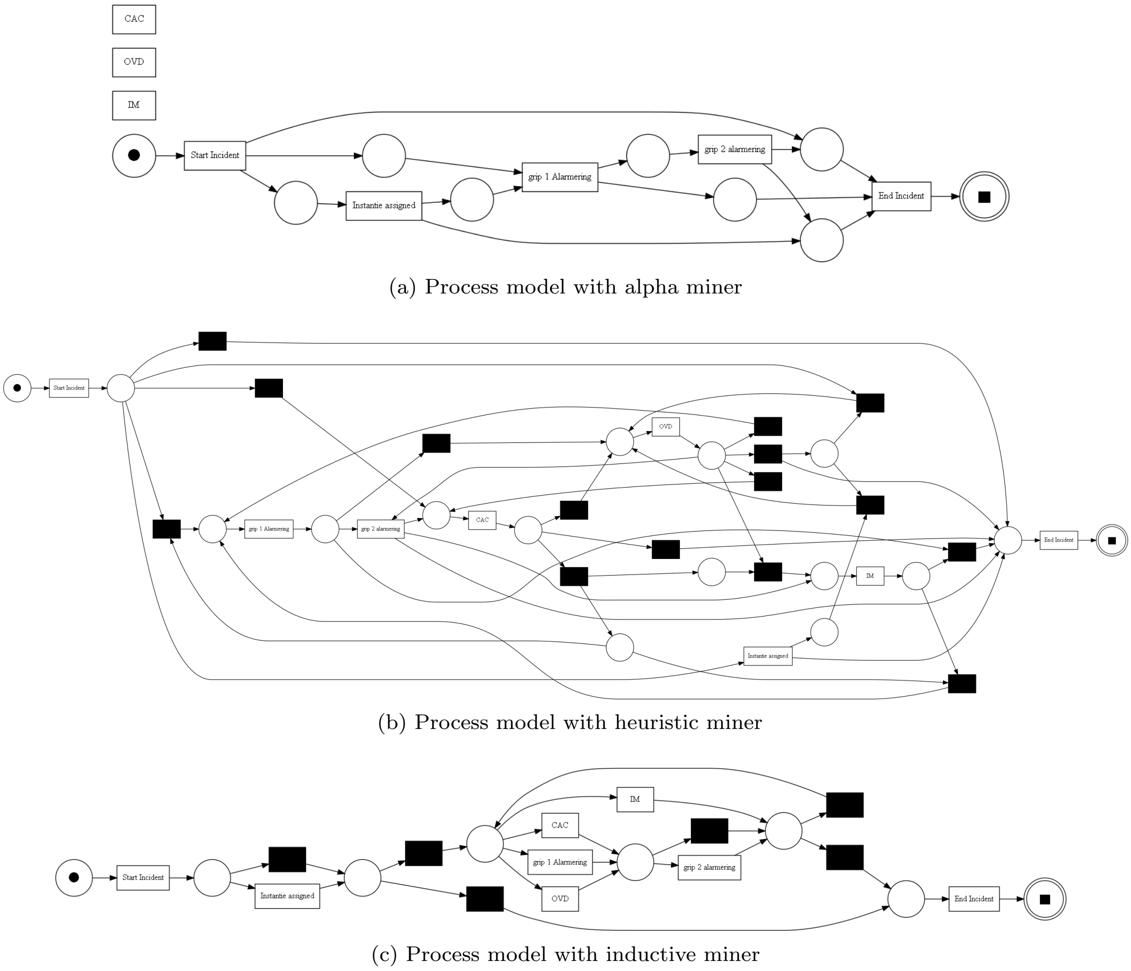


Figure 5.3: Process tree

The inductive miner (Leemans et al., 2013) detects a ‘cut’ in the event log. The detected cuts could be a sequential cut, a parallel cut, a concurrent cut, or a loop cut. These cuts divide the log into sub-logs until a base is found. The cuts can be visualized in a process tree as shown in Figure 5.3. This process tree again can be transformed into a Petri net. The inductive miner contains hidden transitions to create a model, but no duplicate events are possible. The main advantage of this miner is to guarantee a sound model. Therefore, also one of the most popular miners in process mining.

Based on these insights into the miners, it is expected that the heuristic miner and inductive miner will perform best. All the described miners are fitted on the four subsets in python with the package PM4Py. In Figure 5.4 the process models are shown that are discovered by fitting the algorithms on the fourth subset. All other process models can be found in Appendix D. Notice that the alpha miner of Figure 5.4a is not able to include all events in the process model. In the heuristic miner of Figure 5.4b on the other hand many invisible activities are included in the model. The most explanatory model is produced with the inductive miner of Figure 5.4c, in which traces can be clearly tracked.



(a) Process model with alpha miner

(b) Process model with heuristic miner

(c) Process model with inductive miner

Figure 5.4: Process discovery of subset 4

5.3 Conformance Checking

The discovered models of section 5.2 are evaluated for their performance based on four well-known metrics within process mining; fitness, precision, simplicity, and generalization. For each metric is shortly clarified what it means and how it is calculated. Based on these metrics the subsets and miners can be compared.

The performance metrics can be calculated in two ways. Token-replay is a heuristic technique that is easy to understand (Rozinat & Van der Aalst, 2008). It can be calculated by counting for each trace the number of produced tokens, consumed tokens, missing tokens, and remaining tokens.

Although, the calculation for token-replay is easy to understand and implement these are local decisions only, which may lead to misleading results (Rozinat & Van der Aalst, 2008). Another way to calculate the performance is with alignments. With alignments, an exhaustive search is performed to find the optimal alignment between the observed trace and the process model. However, it is guaranteed to return the closest model run compared to the trace. Therefore, this second metric calculation is preferred if this is possible for the models and performance metrics. The difference between these two metrics is nicely explained by Josep Carmona an Associate Professor at Universitat Politècnica de Catalunya: “A nice analogy that tells the difference between token-replay and alignments is searching for a particular place (e.g., a restaurant) in a city: in token-replay, you decide the direction to take just by looking at what you see. With alignments, you take your mobile phone and look at Google Maps, which will tell the optimal route (but pays the price of connecting to a GPS, download the city map, etc.)”(data science group, n.d.). Within the python package PM4PY the alignment can be calculated for the performance metrics fitness and precision.

Fitness quantifies how much of the observed behavior in the event log is captured by the process model (Van Dongen et al., 2016). In simple words, the fitness checks if all traces in the event log can be executed by the process model. If all traces can be executed by the model, the fitness is 1.

Precision quantifies how much behavior exists in the process model that was not observed in the event log (Van Dongen et al., 2016). In other words, what other traces would be possible to be executed by this process model. Especially, with loops, there is an infinite amount of possible traces. A model is found precise if it does not allow for too much behavior, if a model is not precise it is underfitting (Van der Aalst et al., 2012). If the process model only allows for the traces found in the event log then the precision is 1.

Simplicity quantifies the complexity of the model (Van Dongen et al., 2016). The model should not be more complex than necessary to explain the event log (Van der Aalst et al., 2012). This metric compares the simplest process model possible with the current model.

Generalization quantifies how well the model explains unobserved system behaviour (Van Dongen et al., 2016). A model should not be restricted by the behavior observed trace examples in the event log (Van der Aalst et al., 2012). If the model is not general enough it is overfitting.

Subset 1

The results for the first subset are shown in Table 5.1. Subset 1 contains all events without filtering. For the alpha miner, it is not possible to use alignment-based conformance, therefore, token-based replay is used. For all miners, the generalization is calculated with token-based replay as well. The results show that the alpha miner underperforms in each metric. Especially the fitness is low compared to the other miners. The heuristic miner on the other hand perforce well on fitness and precision, but the simplicity and generalization could be better. The inductive miner is outperforming the others in three of the four metrics. Only precision scores are lower than the heuristic miner.

Table 5.1: Conformance checking subset 1

	Fitness	**Precision**	**Simplicity**	**Generalization**
Alpha Miner	0.46	0.68	0.52	0.90
Heuristic Miner	0.89	0.97	0.51	0.66
Inductive Miner	1	0.68	0.58	0.93

Subset 2

The results for the second subset that only includes traces that start with ‘Start incident’ can be found in Table 5.2. The performance of subset 2 is quite similar to the performance of subset 1. The alpha miner only increases slightly on fitness. For the heuristic miner, the fitness and generalization are improving significantly. Also for the inductive miner, there are some improvements, most interesting is the increase in precision with 0.25. All the increases in scores are possibly due to the reduction of trace variants in the event log.

Table 5.2: Conformance checking subset 2

	Fitness	**Precision**	**Simplicity**	**Generalization**
Alpha Miner	0.49	0.67	0.52	0.88
Heuristic Miner	0.97	0.99	0.49	0.76
Inductive Miner	1	0.93	0.60	0.92

Subset 3

In Table 5.3 the results of the third subset are shown, where all different OvDs are renamed as one event OvD. What is most interesting about these results, is that the alpha miner is outperforming the other miners on simplicity and generalization. The increase in performance is not applicable to the other miners. The scores of the heuristic miner are almost similar to those of subset 1. Also for the inductive miner, there are only small changes in performance. The improvement of the alpha miner might be because the renaming of the events is reducing the noise in the traces.

Table 5.3: Conformance checking subset 3

	Fitness	**Precision**	**Simplicity**	**Generalization**
Alpha Miner	0.59	0.55	0.81	0.92
Heuristic Miner	0.89	0.96	0.50	0.79
Inductive Miner	1	0.72	0.65	0.89

Subset 4

The fourth subset is the combination of the trace selection of subset 2 and the renaming of subset 3. Since subsets 2 and 3 are both outperforming subset 1 already, it is expected that subset 4 would have the best overall scores. In Table 5.4 these results are shown and this expectation is confirmed. Regarding the fitness and precision, the alpha miner is clearly less accurate than the other miners. For simplicity it is the other way around, there the alpha miner is clearly better than the heuristic and inductive miner. For the generalization, all heuristics have an acceptable score.

Table 5.4: Conformance checking subset 4

	Fitness	**Precision**	**Simplicity**	**Generalization**
Alpha Miner	0.62	0.56	0.89	0.91
Heuristic Miner	0.91	0.99	0.51	0.80
Inductive Miner	1	0.83	0.65	0.94

The aim is to select the overall best performing miner, therefore all the subsets and miners are compared. The heuristic miner in general has a high precision score, whereas the inductive miner scores are always high on fitness and generalization. Overall, the inductive miner seems the best in balancing all four metrics. Comparing the results of the inductive miner of all subsets the precision, simplicity, and generalization should be considered since the fitness is 1 for all subsets. The precision is highest for subset 2, while the simplicity and generalization are the highest for subset 4. Therefore, the process model found with the inductive miner on subset 4 is chosen as the best performing model.

5.4 Selected Process Model

For decision mining purposes one model is selected to continue with machine learning. The overall best-performing model is chosen for this. An additional benefit of this model is that it is sound. The model is visualized in Figure 5.5. This is the inductive miner fitted on the fourth subset. Besides, this process model is discussed with the VRU, they recognize the elements in the process model of Figure 5.5.

The flow visualized in this process model starts always with ‘Start incident’, in line with the selected traces of the event log. Next, the first decision point is identified with 1. At this point, a decision is made between assigning a partner organization (event: ‘Instantie assigned’) or not (invisible activity). Whether or not a partner is assigned, after these activities the flow merges again in a new decision point, identified in Figure 5.5 with 2. Two choices are possible, with both an invisible activity as the first event. The lower edge goes directly to the event ‘End incident’, in other words, no further multidisciplinary scale-up is necessary for this path. The other choice directly results in a new decision point, namely decision point 3. At this point, there are four outgoing arcs, all defining different choices in multidisciplinary scale up. These four possible events are ‘IM’, ‘CAC’, ‘GRIP 1’ and ‘OvD’. Respectively the second, third and fourth choices lead to the next decision point, number 4. While the event IM leads directly to decision point 5. From decision point 4, the choice for scale-up to GRIP 2 can be made. After which, this flow also enters decision point 5. Decision point 5 is the final decision point in this process model, which includes a loop back to decision point 3. The other option from decision point 5 is to enter the event ‘End incident’. After executing the event ‘End incident’ the process flow is completed.

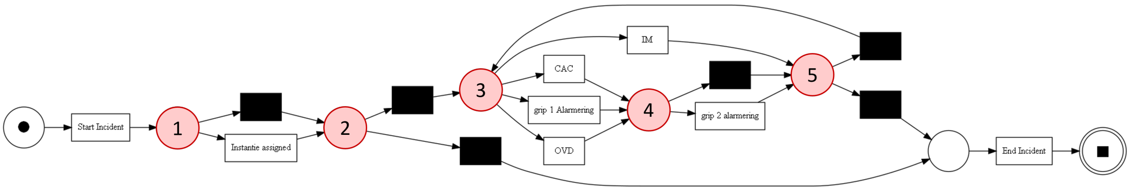


Figure 5.5: Selected process model with decision points

In the literature review of chapter 2, process models with a loop are identified as challenging. This section described that a loop has three decision points involved. For the process model of Figure 5.5, decision point 3 is the decision point that contained a loop. The decision point containing a loop is decision point 2, and the decision point that is a loop is decision point 5. In the article by Rozinat and der Aals (2006) no solution for this challenge is provided, the challenge is notified only. This challenge should be considered during the decision analysis in chapter 6.

5.4.1 Identify Decision Points

In the process model of Figure 5.5 there are five decision points. For decision mining, a ML model should be trained separately on each decision point. But before doing so, first, the context of these decision points is discussed with the VRU. The first decision point is related to involving a partner organization, which is incident type depending, but not really something that slows the process down. Therefore, this decision point is not selected to be further analyzed. For the second decision point, the initial choice has two options, entering multidisciplinary scale up or 'routine' incident handling. However, both have an invisible activity as the first event. Notice that decision point 3 follows immediately after the invisible activity of decision point 2. The second decision point is discussed with the VRU as well. For them, this point is of interest, because they are interested in what the difference is between incidents that require scale-up and those incidents that do not need scale-up. This is what decision point 2 is describing. For decision point 3, there are four outgoing arcs to the activities 'IM', 'CAC', 'GRIP 1', and 'OvD'. This decision point is discussed with the VRU and defined as interesting because the different scale-ups are distinguished in this decision point. However, not enough data is available about these activities so this point is excluded for now. The last three activities mentioned in decision point 3, lead to the next decision point. So the next decision point is 4, where a choice can be made between an invisible activity or 'GRIP 2'. In other words, scale up to GRIP 2 or not. Since there is a limited number of incidents that require scale-up to GRIP 2, it is decided in consultation with the VRU to focus on other decision points. The final decision point of the process model is decision point 5. In this case, there are two outgoing arcs with invisible activities, representing no further scale-up required and a loop back to decision point 3. Also, the fifth decision point is defined as interesting in consult with the VRU. To conclude, decision point 2 and decision point 5 are of most interest to the VRU and this research continues to analyze these decisions.

5.5 Chapter Overview

In this chapter all process mining steps are discussed, leading to the selection of one process model. Therefore, four different subsets are created. The first subset is containing all data, in the second subset the traces with start activity 'Start incident' are selected, in the third subset all different OvD activities are generalized as one activity 'OvD', and in the fourth subset the filtering of both the second and third subset is combined. Next, three different process discovery algorithms are applied on the event log, to explore different process models. Followed by an evaluation of the process models with conformance checking. In which the subsets are compared on the metrics; fitness, precision, simplicity and generalization. Finally, the best performing process model is selected to continue with. This is the inductive miner fitted on the fourth subset. In line with decision mining, the decision points in this process model are identified.

Decision point 2 and 5, are selected as most interesting for the VRU to explore for data patterns. These decision points are the input for chapter 6, in which these decision point have to be transformed into a learning problem.

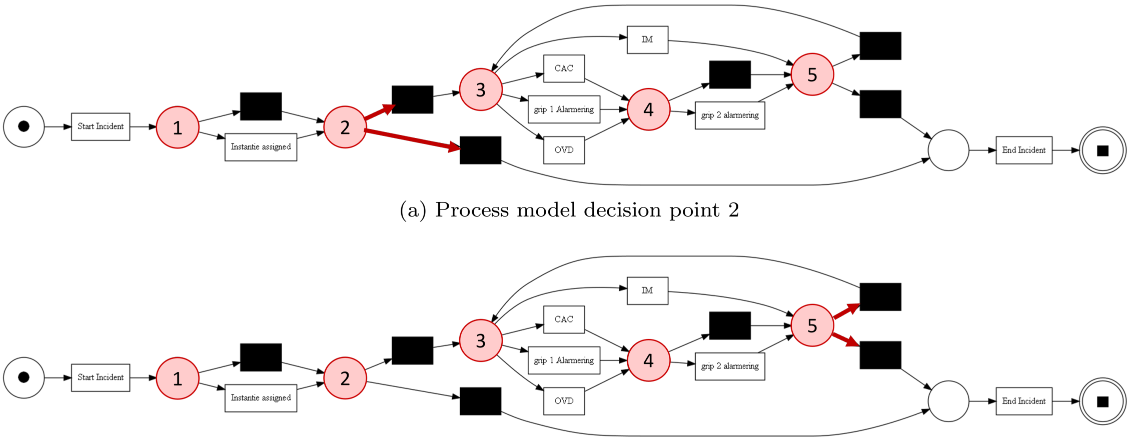
Chapter 6

Analyzing Decisions with Machine Learning

The second stage of decision mining is analyzing the decision paths in the discovered process with ML. The aim is to identify different data patterns for different decisions. The input features are designed based on the criteria highlighted by decision-makers, with help of the available data discussed in chapter 4. Since the final goal is to receive explicit criteria, the main focus is not on training the best model but on balancing model score and explainability. This is considered in both model selection and model evaluation. Finally, the most important features considered by the model are compared to the implicit knowledge insights of decision-makers.

6.1 Identify Learning Problem

In subsection 5.4.1, decision points 2 and decision point 5 are identified as interesting. This section explains how the event log can be transformed into a learning problem for these decision points. Both decision points have two outgoing arcs followed by invisible activities. These decision points are transformed into binary classification problems in this section.



(a) Process model decision point 2

(b) Process model decision point 5

Figure 6.1: Process model binary classification problem

6.1.1 Binary Classification Problem

In Figure 6.1 the discussed process model is represented for decision point 2 in 6.1a and for decision point 5 in 6.1b. In both figures, the outgoing arcs are highlighted in red. These are the directions and therefore the classes of interest.

For the classification problem of decision point 2 can be seen in Figure 6.1a that invisible activities are encountered first. These invisible activities have no labels in the process model. However, if we place this in the context of an incident it is possible to label these activities. The lower arrow of Figure 6.1a leaving decision point 2 does not include any activities until 'End incident'. In the context of an incident, this means that the disciplines on side can handle the incident without further scale-up. Therefore, this invisible activity could be named a 'routine incident' for instance. The upper arrow in Figure 6.1a towards decision point 3, is the connection with all events that are defined as multidisciplinary scale-up. In that perspective, this activity could be named 'Multi scale up required'. These labels are used for the classification of decision point 2.

In Figure 6.1b the process model for decision point 5 is shown, both outgoing arcs encounter an invisible activity at first. The upper arrow is a loop back to decision point 3. In the context of the model, this loop identifies a request for additional multidisciplinary scale-up. Therefore, this invisible activity could be named 'Additional multi scale up'. The lower arc leaving decision point 5 leaves towards 'End incident'. Therefore, the invisible activity can be named 'No additional multi scale up'. These labels are used for the classification of decision point 5.

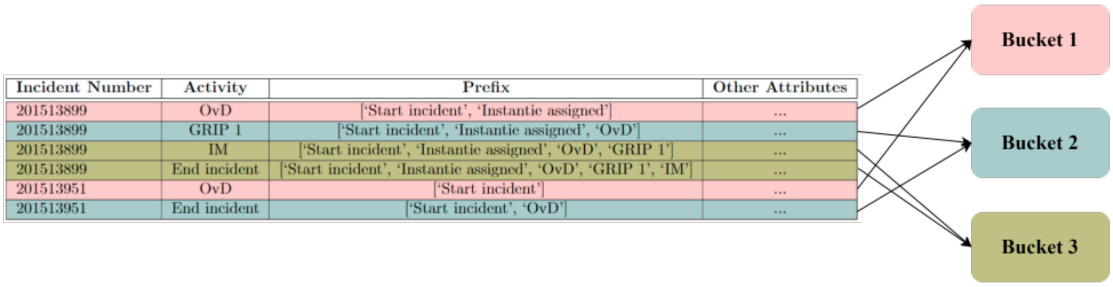
From the event log, a selection is made off all records that correspond to the selected decision points. Therefore, the records with *Activity* names 'Start incident', 'Instantie assigned' and 'GRIP 2' are removed from the data set. The information that these activities were executed for an incidents is remained in the feature *Prefix*.

6.1.2 Create Buckets

At this point, we have the two identified decision point, decision point 2 and 5. However, in Figure 5.5 is noticed that the loop has some challenges. For decision point 2 there is no problem since it contains only the incidents entering the loop for the first time. However, for decision point 5 there are some challenges because this contains all except the first time in the loop. In the context of the VRU it is of interest in the difference between incidents that go through the loop once, and incidents that have multiple rounds in the loop. To investigate this difference, we need to split decision point 5. To do so, the activities prior to the decision point are investigated, referred to as prefixes. This idea is derived from predictive process monitoring (Teinemaa et al., 2019). In which the next activity in a process is predicted based on KPIs.

In order to use the prefixes, first, a new attribute is created. The prefix is a list of all activities prior to the current position in the process model, the decision point. This attribute is added to all records in the data set. The selection for the buckets is based on the length and activities in the prefix. A visual representation is shown in Figure 6.2. Each bucket is described separately below. In this buckets not all records have the same prefix length due to the separation based on the decision points. Eventually, for each bucket, a separate model is trained.

For the first bucket, all records are selected with the prefix <'Start incident'> or <'Start incident', 'Instantie assigned'>. This is equal to all records that should be selected for decision point 2. For the first bucket, 6822 records are selected. The second bucket contains all incidents with the prefix <'Start incident', X> or <'Start incident', 'Instantie assigned', X> for which X is not equal to 'End incident'. In total 1845 records are selected that fulfill this requirement. Finally, the third bucket exists of all records left, which are 816 records. From now on we continue with these three buckets as separate data sets.



Incident Number	Activity	Prefix	Other Attributes
201513899	OvD	[‘Start incident’, ‘Instantie assigned’]	...
201513899	GRIP 1	[‘Start incident’, ‘Instantie assigned’, ‘OvD’]	...
201513899	IM	[‘Start incident’, ‘Instantie assigned’, ‘OvD’, ‘GRIP 1’]	...
201513899	End incident	[‘Start incident’, ‘Instantie assigned’, ‘OvD’, ‘GRIP 1’, ‘IM’]	...
201513951	OvD	[‘Start incident’]	...
201513951	End incident	[‘Start incident’, ‘OvD’]	...

Figure 6.2: Example representation of different buckets

6.2 Data Preparation

The previously described buckets are the three data sets that we continue with from now on. Before the modeling start, the data need to be prepared. The three buckets are selections of the event log that is used for process mining. Now, this event log has to be changed for ML. Therefore, several steps are taken. First, the described names for the invisible activities are added as targets to the data sets. Secondly, new features are created based on the found criteria in the questionnaire. Furthermore, the categorical features are encoded.

6.2.1 Target labeling

For modeling purposes first, the target has to be defined. The target to predict is already discussed in section 6.1. These labels are assigned accordingly. The first bucket is based on decision point 2, therefore the labels ‘Multi scale up required’ and ‘Routine incident’ are desired. Bucket 2 and bucket 3 are based on decision point 5, with the labels ‘Additional multi scale up’ and ‘No additional multi scale up’. The current labels in the data sets are ‘CAC’, ‘OvD’, ‘IM’, ‘GRIP 1’ and ‘End incident’. The split for the new desired labels is based on the difference between scale-up or not. The current labels related to multidisciplinary scale up are ‘CAC’, ‘OvD’, ‘IM’ and ‘GRIP 1’, therefore, these labels are replaced by ‘Multi scale up required’ for bucket 1, and ‘Additional multi scale up’ for bucket 2 and 3. The current label for incidents that do not need (additional) multidisciplinary scale-up is ‘End incident’. Therefore, for bucket 1 this label is replaced by ‘Routine incident’, while for buckets 2 and 3 this label is replaced by ‘No additional multi scale up’.

Table 6.1 shows the distribution of the labels in the different buckets. From now on these buckets are referred to as decision moments, respectively decision moment 1, decision moment 2 and decision moment 3. Notice that for decision moment 1 the label ‘Multi scale up required’ occurs approximately for 33 % of the data in this data set. This distribution is similar for decision moment 3. While for decision moment 2 the imbalance is even larger. Here the label ‘Additional multi scale up’ occurs by approximately 15 % of the data in this bucket. This imbalance should be considered in section 6.3

Table 6.1: Distribution of target labels in the three decision moments

	Multi scale up required // **Additional multi scale up**	**Routine incident//** **No additional multi scale up**
Decision moment 1	2334	4381
Decision moment 2	277	1671
Decision moment 3	222	463

6.2.2 Feature Creation

The implemented features are selected based on the questionnaire results. The data sets are searched for data representing the criteria found most important by decision-makers. These criteria are 'Incident location', 'Incident size', 'Incident type', '(number of) injured and injury classification', 'Sensitivity on social media', 'Own disciplines involved', 'Expected duration incident', 'Possible effects on people/material', 'Safe/unsafe area', 'Involved partners' and 'Duration of incident unknown' as mentioned before in section 4.5.

The attributes already in the data set are explored to match these criteria first. For the 'Incident location' several describing attributes were identified in chapter 4. For the attributes, we include here we have to keep in mind the reason why incident location is important for the decision-maker. Included are *Municipality* and *Type location.* The *Address* and *City* describe the location as well but might be too incident-specific. Therefore, the assessment is made to not include those. The criterion 'Incident type' is represented well in the data sets. The three classification criteria describe the incident type. Therefore, included as attributes are *Classification criteria 1*, *Classification criteria 2*, and *Classification criteria 3.*

Furthermore, new features are created with the data to complement the existing attributes. For the criterion 'Own disciplines involved', a feature is created that counts the number of deployed vehicles at that moment. Other criteria mentioned are 'Expected duration incident' and 'Duration of incident unknown'. Based on these criteria and the information in the open text questions describing the importance of duration, the feature *Current duration* is created. Furthermore, an additional feature is created for the 'Incident location', namely *Highway.* Because this is a specific incident location that might have a large impact according to decision-makers. Finally, one additional feature is created, namely *Event.* This feature is based on the criterion 'Event or demonstration' in the questionnaire. This criterion is not important for all three situations and therefore not selected as the most important criteria. Nevertheless, For one situation this criterion has a high score so this feature is created.

For some of the criteria mentioned in the questionnaire, it is not possible to create relevant features with the available data. For example, the criteria 'Incident size', '(number of) injured and injury classification' and 'Sensitivity on social media'. No structured data is kept or this data is not available due to privacy issues. An overview of all features included and their data type is shown in Table 6.2.

Table 6.2: Overview included features

Feature	**Feature type**
Classification criteria 1	Categorical
Classification criteria 2	Categorical
Classification criteria 3	Categorical
Municipality	Categorical
Type location	Categorical
Deployed vehicles	Numerical
Current duration	Numerical
Highway	Numerical
Event	Numerical
Prefix	List

6.2.3 Feature Encoding

Most models can only handle numerical features. In Table 6.2 is shown that there are several categorical features and one list feature. These need to be transformed into numerical features before modeling can take place.

For the categorical features, one hot encoding is the most often used method. With this encoder, all categories of a feature are separated as single features that have the value 1 for True and 0 for False. This method is applied for the features *Classification criteria 1*, *Classification criteria 2*, *Classification criteria 3*, *Municipality*, and *Type location*.

In the feature overview of Table 6.2, the feature *Prefix* is a list. This list contains a subtrace with information on the prior executed activities. This list should be transformed into a numerical feature as well. Since the prefixes in the different decision moments are not all the same length, we chose to use aggregation encoding to transform the *Prefix*. Aggregation encoding means that for each process activity a feature is created. For example, the feature for the activity 'Start incident' is *Prefix_StartIncident*. This feature represents the number of times this activity has appeared in the prefix.

After both encoding steps, there are 161 features in the train data of decision moment 1. In the second decision moment, there are 119 features in the training data and for the third decision moment, there are 92 features.

6.3 Modeling

In the previous sections, all data is prepared for modeling. We have three data sets for different moments in the process model. In this section, we discuss what modeling techniques are applied. Additionally, the evaluation metrics are selected and introduced. After which, the performed experiments are discussed in the experimental setup.

6.3.1 Model Selection

As discussed in chapter 2, the modeling technique used in decision mining is a decision tree. This ML technique is highly explainable because of the white box structure, unlike most other ML techniques. The decision tree algorithm is therefore applied in this section. The simplicity of decision trees which ensures the explainability may also result in lower model performance. Therefore, a comparison is made with random forest algorithms. In random forest multiple decision trees are trained and averaged. Resulting in a more complex model with lower explainability, but generally higher model performance.

Decision tree

The decision tree algorithm classifies instances by sorting them based on feature values. It is a supervised learning algorithm used for classification and regression problems (Myles et al., 2004). A decision tree trains a model that can be used to predict the target class by learning simple decision rules inferred from prior data, the training data. The tree exists of different nodes and branches, as shown in the example of Figure 6.3 (Myles et al., 2004). The root node is the starting point, after which branches lead to the next node. These internal nodes contain decision rules based on the available features. This continues till the leaf node is reached. The decision rules are selected based on the impurity of the nodes. Because of the decision rules, decision trees are easily interpretable (Kotsiantis et al., 2007). Most well-known ML techniques such as neural networks are often black boxes, showing only the in- and out-put of the model. However, for knowledge discovery, the underlying reasons for the mapping should be interpreted (Liu et al., 2017). The possibility to discover the underlying reasoning allows for rule distraction for the multidisciplinary scale-up process.

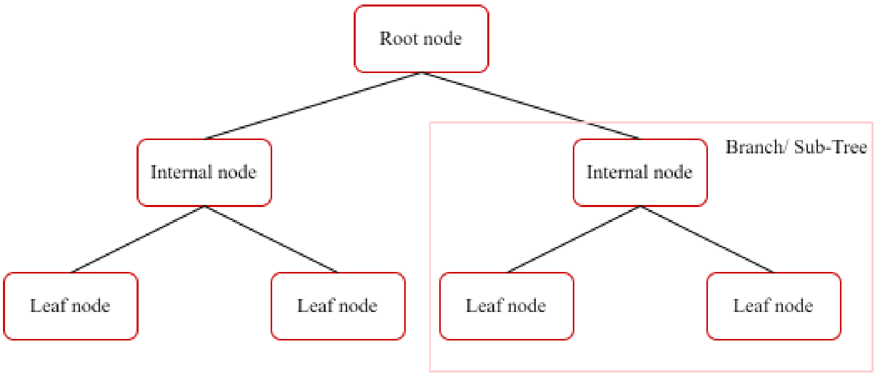


Figure 6.3: Example decision tree.

Random forest

The random forest algorithm is an ensemble method based on using the decision tree algorithm in combination with bagging (Breiman, 2001). It builds a number of decision trees on bootstrapped training samples, but each time a split in a tree is considered, a random sample of m features is chosen as split candidates from the full set of p features. This selection is made so that not all the trees use the same strong feature. Suppose that we have a strong feature in the data set along with a number of moderately strong features, then in the collection of bagged trees, most or all of them will use the very strong feature for the first split. As a result, all bagged trees will look similar. Hence all the predictions from the bagged trees will be highly correlated. Highly correlated trees do not lead to a large variance reduction. Therefore, a random selection of m features is available for each tree, so each individual tree has high variance but low bias, without high correlation to other trees. Averaging these trees reduces the variance, and thus both variance and bias are lowered. However, the bagging of trees may increase the performance, and the easy interpretability of the decision trees is not possible anymore. In the context of knowledge discovery an explainer for the black box is required for insights.

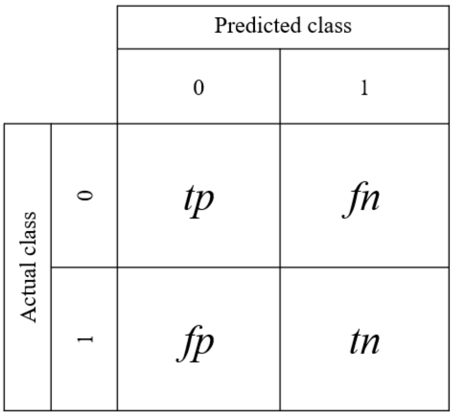
6.3.2 Evaluation Metrics

In line with the aim for good model performance and an explainable model, several performance metrics are selected. For the performance evaluation of the selected models, metrics should be selected to evaluate an imbalanced binary classification problem. The explainability of the decision tree can be explored by extracting the tree. However, we also want to explore the possibility to interpret the random forest model, therefore, model explainability with Shapley Additive Explanations (SHAP) is explored, to gain more insight in the contribution of features than feature importance only.

Model Performance

The most common performance measure for classification problems is accuracy. This metric defines how many observations, both positive and negative, were correctly classified. However, accuracy is misleading for an imbalanced data set (Luque et al., 2019). Because a high accuracy score can be reached by classifying all observations as the majority class. Since we have data sets that are clearly imbalanced, the accuracy is not reliable so other metrics should be used.

The **Confusion matrix** provides insights into the performance of imbalanced data sets. The matrix is a common way to present the true positives (*tp*), true negatives (*tn*), false positives (*fp*) and false negatives (*fn*). In Figure 6.4, the common presentation of these metrics is shown, where the x-axis shows the predicted class and the y-axis shows the actual class.



		Predicted class	
		0	1
Actual class	0	*tp*	*fn*
	1	*fp*	*tn*

Figure 6.4: Example confusion matrix with true positives (*tp*), true negatives (*tn*), false positives (*fp*) and false negatives (*fn*).

Given the confusion matrix, we can calculate the **Recall** and **Precision**. On the one hand, the recall measures how many of all positive observations, are classified as positive (Equation 6.1). On the other hand, the precision measures how many of the predicted positive observations are indeed classified as positive (Equation 6.2). Generally, increasing the precision leads to a decrease in recall and the other way around.

$$Recall = \frac{tp}{tp + fn} \tag{6.1}$$

$$Precision = \frac{tp}{tp + fp} \tag{6.2}$$

However, in case recall and precision are equally important the **f1 score** balances these metrics. The formula to balance these metrics is shown in Equation 6.3. For the learning problem on hand, these metrics are equally important. The f1 score can also be used with imbalanced data. The f1 score is computed for each class. With the macro f1 score the average of these classes is taken without considering the proportion for each class in the data set. For imbalanced data, also the weighted f1 score can be calculated. Now the average over the classes is calculated while taking the proportion of each class into consideration. Both the macro and weighted f1 scores were measured in the experiments of this research.

$$F_1 = 2 * \frac{\text{precision } * \text{ recall}}{\text{precision } + \text{ recall}} \tag{6.3}$$

The final performance metric included is the **Matthews Correlation Coefficient (MCC)**. This metric calculates the correlation between the predicted classes and the ground truth (Chicco & Jurman, 2020), as can be seen in Equation 6.4. According to Chicco and Jurman (2020), the MCC is a more reliable metric than the accuracy and f1 score since it produces only a high score if the prediction obtained good results in all of the four categories in the confusion matrix. Therefore, this metric is included as a performance metric to evaluate and compare the different models.

$$MCC = \frac{tp * tn - fp * fn}{(tp + fp)(tp + fn)(tn + fp)(tn + fn)} \tag{6.4}$$

Model Explainability

Model explainability refers to the understanding of the ML model. The ability to interpret the features the model is learning. This interpretation is especially necessary to gain the trust of people. For this research specifically, the interpretation is required to establish decision rules as explicit criteria.

We can distinguish two types of models for explainability, the white box model and the black-box model. White box models refer to a model like decision trees, that are inherently interpretable. For a decision tree, the learned tree can be distracted, this tree can be interpreted as decision rules. For black-box models, this is more complex since the models are not that easy to interpret. Therefore, post-hoc explanation is necessary for black box models such as random forest.

There are two ways to interpret a model, with local and global interpretation. Local interpretation helps understand how the model makes decisions for a single record, whereas global interpretation helps understand the overall decision structure of the model. Therefore, we are mainly interested in global interpretation in this research. The model can be interpreted globally with SHAP. SHAP is a game-theoretic approach to explaining the output of any machine learning model (Lundberg, 2018). The python package allows to for an easy explanation of the fitted models. For global interpretation, the summary plot and dependency plots are explored. The summary plot provides several important features insight into the relationship between the value of a feature and the impact on the prediction. The features are ordered based on importance and the color indicates the feature value. On the x-axis, the impact on the prediction is plotted. It can be seen in the plot that for some specific feature values the impact on the prediction is positive and for others negative. However, for more insight in the exact form of the relationship the dependency plot can help. There the impact on the prediction is plotted against the feature value. Since dependency plots have less meaning if the feature is strongly related to another feature, the color indicates the relation with the strongest related feature.

6.3.3 Experimental Setup

After identifying the modeling techniques and evaluation metrics, the experimental setup is defined. All steps are illustrated to implement and optimize the models. These steps are repeated for all three data sets.

For training and testing purposes the data set is split. A subset of 30 % is subtracted for testing. Remaining a 70 % subset for training the model and optimizing the hyper parameters. Because the data is imbalanced as shown in Table 6.1 it is chosen to use a stratified split. A stratified split ensures that the train and test set have the same distribution of target classes as the original data set. Besides, the samples are randomly distributed over the train and test set.

In Figure 6.5, the experimental set-up is illustrated. Each decision moment is a separate data set. For each of these data sets, a decision tree model and random forest model are trained on the training data. The parameters of these models are optimized with a grid search, using 5-fold stratified cross-validation. With the grid search, all possible combinations of the provided parameter settings are fitted. For both models the class weight parameter ‘balanced’ is used to compensate for the class imbalance. The parameter ‘balanced’ adjusts the weights inversely proportional to the class frequencies. The other parameters optimized for both models are the maximum depth of the tree(s) and the minimal samples for each leaf. For the decision tree, the max depth is between 2 and 10, to ensure that it is possible to interpret the tree. The scoring used in the grid search is the weighted F1 score. This score computes the f1 score for each class and returns the average while considering the proportion of samples in each class.

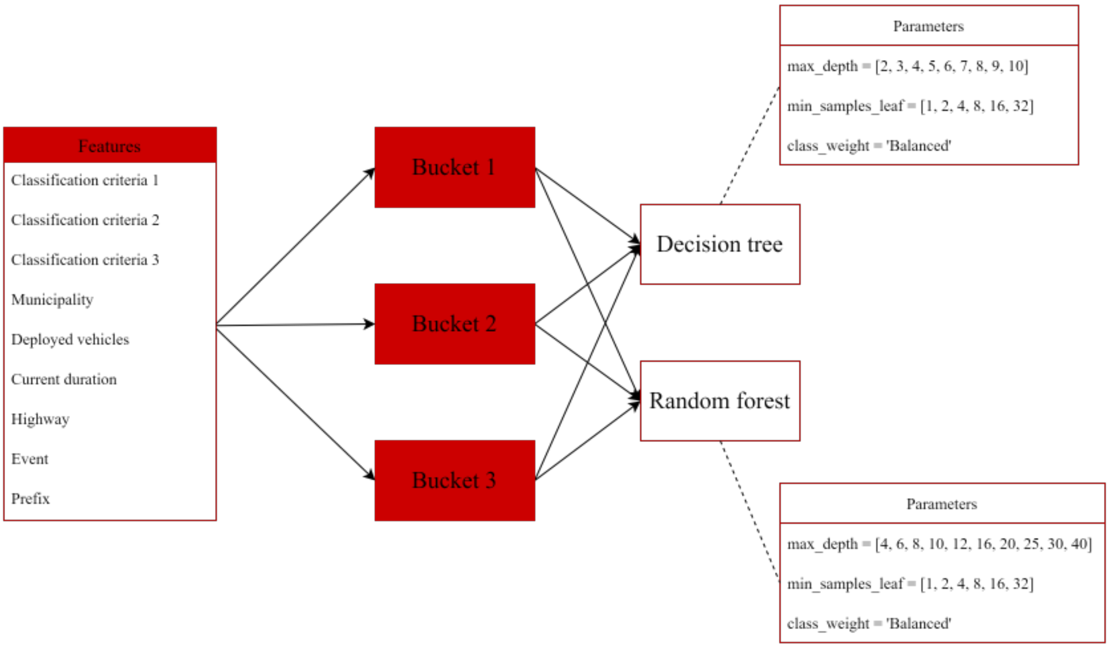

Features
Classification criteria 1
Classification criteria 2
Classification criteria 3
Municipality
Deployed vehicles
Current duration
Highway
Event
Prefix
Bucket 1
Bucket 2
Bucket 3
Decision tree
Random forest
Parameters
max_depth = [2, 3, 4, 5, 6, 7, 8, 9, 10]
min_samples_leaf = [1, 2, 4, 8, 16, 32]
class_weight = 'Balanced'
Parameters
max_depth = [4, 6, 8, 10, 12, 16, 20, 25, 30, 40]
min_samples_leaf = [1, 2, 4, 8, 16, 32]
class_weight = 'Balanced'

Figure 6.5: Experimental set-up

6.4 Results

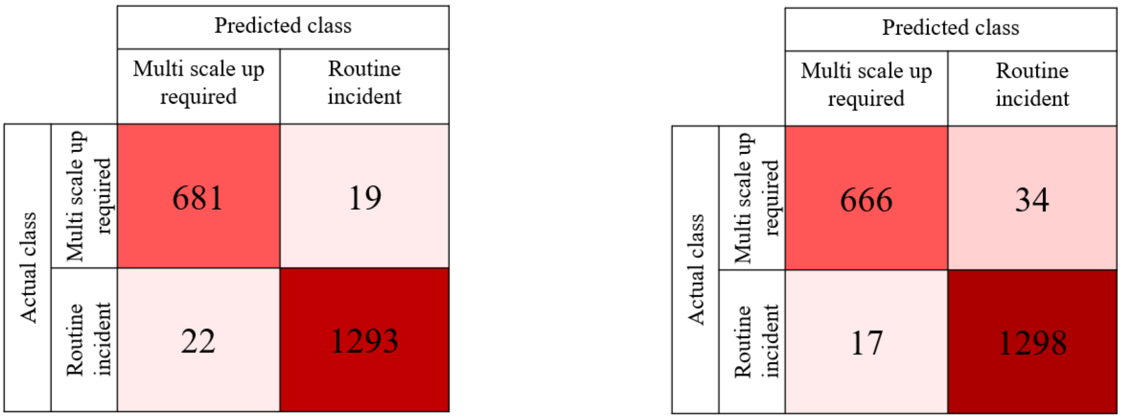
The experiments are executed and in this chapter the results are discussed. For each data set the decision tree and random forest model are evaluated based on performance and explainability. Besides, these models are compared to each other and important criteria are extracted.

6.4.1 Results Decision Moment 1

In this section the results of the first bucket are shown. Remember that decision 1 refers to decision point 2, here distinction is made between incidents that require multi scale up and routine incidents. This decision is predicted with the decision tree model and the random forest model. The parameters for both models are optimized with a grid search. To compare the decision tree model and random forest model, first the confusion matrices are shown for both models. After which the other performance metrics are discussed. Next, the explainability of both models is explored with the SHAP summary plot and some additional dependency plots.

Performance Metrics

In Figure 6.6 the confusion matrices of the decision tree model and the random forest model are shown. Both models have a good performance. There is some difference in the false predicted instances between the models. On the one hand, the decision tree model of Figure 6.6a has an almost equal amount of *fn* and *fp*. On the other hand, the random forest model of Figure 6.6b has double as much *fn* as *fp*. From this can be concluded that the random forest model is predicting the target class 'Routine incident' more often. With as result more *tn* but also more *fn*. The target class 'Routine incident' is also the majority class, however, from these results there is no indication that this is causing any problems with predicting.



		Predicted class	
		Multi scale up required	Routine incident
Actual class	Multi scale up required	681	19
	Routine incident	22	1293

(a) Decision Tree (max_depth = 6, min_samples_leaf = 1)

		Predicted class	
		Multi scale up required	Routine incident
Actual class	Multi scale up required	666	34
	Routine incident	17	1298

(b) Random Forest (max_depth = 25, min_samples_leaf = 1)

Figure 6.6: Confusion matrix decision moment 1 on test data

An overview of all performance metrics calculated for decision moment 1 is shown in Table 6.3. Notice that both models perform really well with a weighted f1 score of 0.980 for the decision tree model and 0.975 for the random forest model. The slightly less performance of the random forest model is a result of the model predicting the majority class more often, as mentioned earlier. Therefore, the f1 score for the target class 'Routine incident' is higher but the f1 score for 'Multi scale up required' is lower. This can be seen by the fact that the macro f1 score of the random forest model is higher than the of the decision tree model. But for the weighted score it is the other way around. Overall, the models are close to each other based on the f1 score. The same can be said about the MCC score. The decision tree has a slightly higher MCC score than the random forest model. Based on this metric can be concluded that both model do explain a lot compared to the ground truth since the value is close to 1.

Table 6.3: Summary of performance metrics for decision moment 1, comparing the decision tree model and the random forest model.

	Precision	**Recall**	**F1 macro**	**F1 weighted**	**MCC**
Decision tree	0.977	0.978	0.968	0.980	0.955
Random forest	0.975	0.969	0.972	0.975	0.944

Overall can be concluded that the decision tree is performing slightly better on all important performance metrics. The decision tree model is better at balancing the target classes, higher weighted f1 score and higher MCC. However, notice that the differences are only small.

Model Explainability

The model explainability of the decision tree model and random forest model are examined next. For the decision tree it is possible to extract the learned tree. This tree is visualized in Appendix E. The impurity of the leaf nodes is of interest. There are 4 leaf nodes with a gini index higher than 0.4. These leafs are found most confused about the predicted class since the almost fifty-fifty for both classes in this leaf. Based on this model decision rules can be extracted. These are listed here:

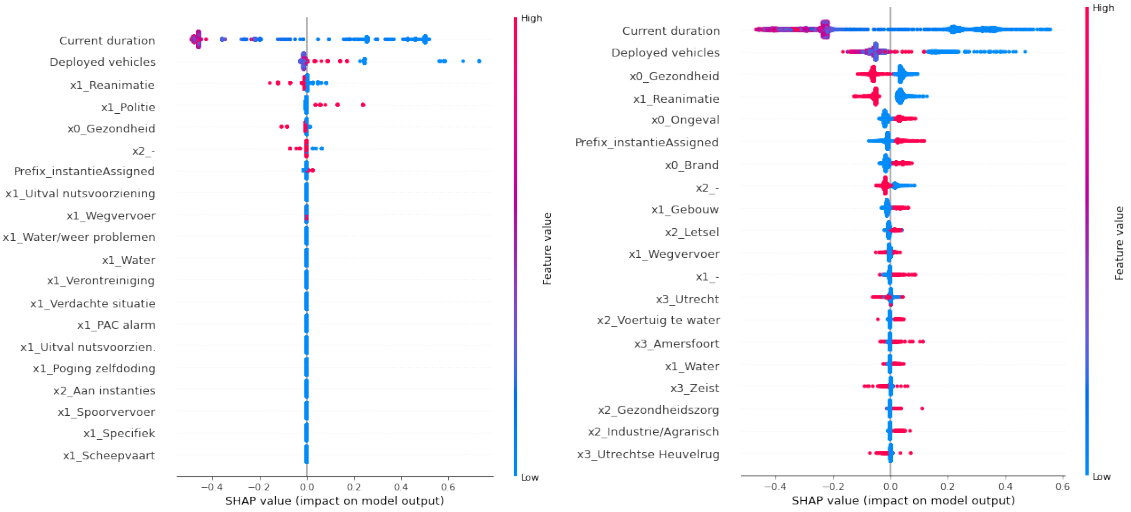
Decision rules leading to class ‘Multi scale up required’:

- *Current duration* $\leq$ 0.495 hour.
- *Current duration* $\geq$ 0.495 hour AND *Deployed vehicles* $\leq$ 0.5.
- 0.495 $\leq$ *Current duration* $\leq$ 0.964 hour AND *Deployed vehicles* $\geq$ 3.5 AND *Resuscitation* = 0.

Decision rules leading to class ‘Routine incident’:

- 0.495 $\leq$ *Current duration* $\leq$ 0.964 hour AND *Deployed vehicles* $\geq$ 0.5 AND *Resuscitation* = 1.
- c0.495 $\leq$ *Current duration* $\leq$ 0.964 hour AND 0.5 $\leq$ *Deployed vehicles* $\leq$ 3.5 AND *Resuscitation* = 0.
- *Current duration* $\geq$ 0.964 hour.

For the random forest model it is not possible to extract just one tree, therefore, the SHAP explainer is used as described before. Because this might result in different interpretation as the decision tree the SHAP explainer is applied on both models for comparison possibilities. The summary plot shown in Figure 6.7 shows the most important features for both models with their impact on multi scale up required. As can be seen the upper two, and therefore most important, features are the same for both models. The decision tree seems to consider around seven features. Where the random forest model considers many more. This is due to the bagging of the models in the random forest were other features are selected to build a tree each time. For all binary features clear interpretation is possible form this plot, since 1 is red and 0 is blue. So if there are only red values on the right side, the value 1 has a positive impact on multi classification and the other way around. However, for the continues values this is harder to interpret from this plot only. Therefore, the dependency plots have to be explored.

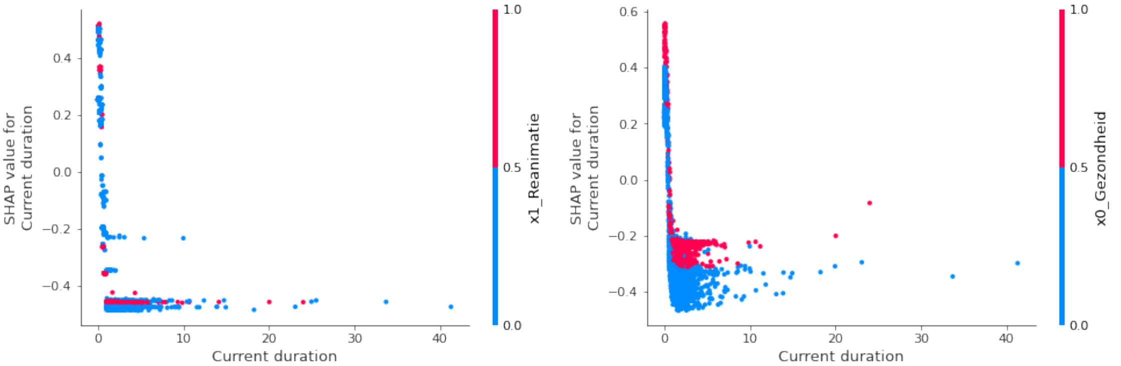


(a) Decision tree model

(b) Random forest model

Figure 6.7: SHAP summary plots for Decision moment 1

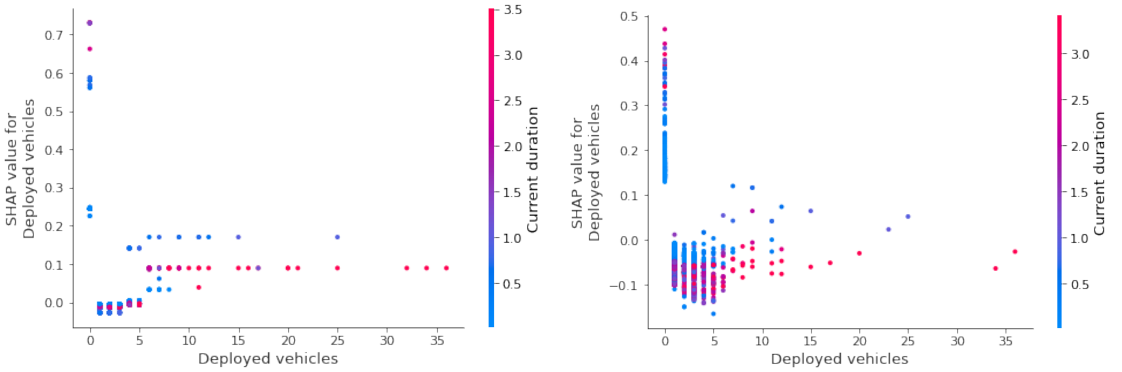
In the Figure 6.8 and Figure 6.9, the dependency plots for the features *Current duration* and *Deployed vehicles* can be found. In these plots the feature value is shown on the x-axis and the impact on the y-axis. For the *Current duration* both models have a similar plot, with positive SHAP values for very short current duration of the incident and negative SHAP values is the incident has an longer duration at this first decision point. For the *Deployed vehicles*, there is a difference in the plots of the models. In the dependency plot for the random forest model, it can be seen that for more than zero vehicles, the impact on the prediction for multi scale up required, is negative or very low. For the decision tree model, the same is true but only between one and four deployed vehicles. As more than four vehicles are deployed the model has a positive impact on predicting multi scale up required.



(a) Decision tree model

(b) Random forest model

Figure 6.8: SHAP dependency plots for the feature Current duration of decision moment 1



(a) Decision tree model

(b) Random forest model

Figure 6.9: SHAP dependency plots for the feature Number of deployed vehicles of decision moment 1

To summarize the top five of explicit criteria that can be retrieved from the SHAP plots. On the one hand for the decision tree model the features *Current duration*, *Deployed vehicles*, *Resuscitation*, *Police* and *Healthcare* seems to have positive impact on predicting the class multi scale up required. More specifically, for the *Deployed vehicles* positive impact is found by 0 or more than 4 vehicles. For the *Resuscitation* there is positive impact if this feature is zero. For the *Police* there is positive impact if this feature is one. For the *Healthcare* there is negative impact if this feature is one. The specific current duration that has positive impact is to short to retrieve from the plot. On the other hand we have the random forest model with as top five features: *Current duration*, *Deployed vehicles*, *Healthcare*, *Resuscitation* and *Accident*. The *Deployed vehicles* equal to zero has an positive impact on the prediction. Also the *Healthcare* and *Resuscitation* is zero has positive impact. Besides, the *Accident* is equal to one has an positive impact as well.

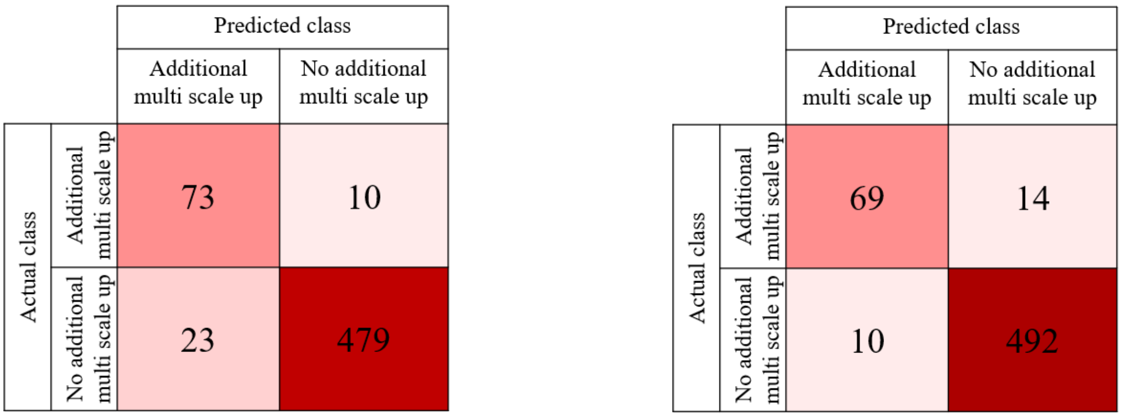
6.4.2 Results Decision Moment 2

Next, the results for decision moment 2 are discussed. This decision moment refers to the first time the loop is passed and a decision has to be made for more multidisciplinary scale up or not. This decision is again predicted with the decision tree model and the random forest model. The parameters for both models are optimized with a grid search. To compare the decision tree model and random forest model, first the confusion matrices are shown for both models. After which the other performance metrics are discussed. Next, the explainability of both models is explored with the SHAP summary plot and some additional dependency plots.

Performance Metrics

In Figure 6.10 the confusion matrices of the decision tree model and random forest model are shown. Both models perform quit well, despite the large class imbalance in the data. The decision tree model seems to predict the minority class a little bit better. However, this also results in a larger number of *fp* compared to the random forest model. From this results it seems that the decision tree model is a little overfitting on the class ‘Additional multi scale up’. Therefore, the random forest model seems to balance this a little better. However, both models seem to represent the data well.

All other performance metrics for decision moment 2 are summarized in Table 6.4. Again both models score well on all performance metrics. The weighted f1 score is 0.945 for the decision tree



(a) Decision Tree (max_depth = 4, min_samples_leaf = 1)

(b) Random Forest (max_depth = 12, min_samples_leaf = 1)

Figure 6.10: Confusion matrix decision moment 2 on test data

model, and slightly higher with 0.959 for the random forest model. The difference between those models on the macro f1 score is a bit more. There the decision tree model scores under the 0.9 with a score of 0.891. While the random forest model scores 0.914. This larger difference in the macro f1 then the weighted f1 is caused by the fact that the decision tree has a lower f1 score for the class ‘Additional multi scale up’. When we compare the models based on the MCC score. The random forest model is performing better with a score of 0.828 than the decision tree model that has a score of 0.785. So the random forest model seems to explain mare compared to the ground truth.

Table 6.4: Summary of performance metrics for decision moment 2, comparing the decision tree model and the random forest model.

	Precision	**Recall**	**F1 macro**	**F1 weighted**	**MCC**
Decision tree	0.870	0.917	0.891	0.945	0.785
Random forest	0.923	0.906	0.914	0.959	0.828

Overall, can be concluded from these metrics that the random forest model the best predicting model for decision moment 2. However, the predictions of the decision tree model are found good as well.

Model Explainability

Besides the model performance, we look at the model explainability of both models. The trained tree by the decision tree model is extracted and visualized in Appendix E. From this tree some decision rules can be distracted. These rules are summarized below. However, leaf nodes with only 1 sample are found to overfitting to extract as decision rules. So these are left out of the lists.

Decision rules leading to class ‘Additional multi scale up’:

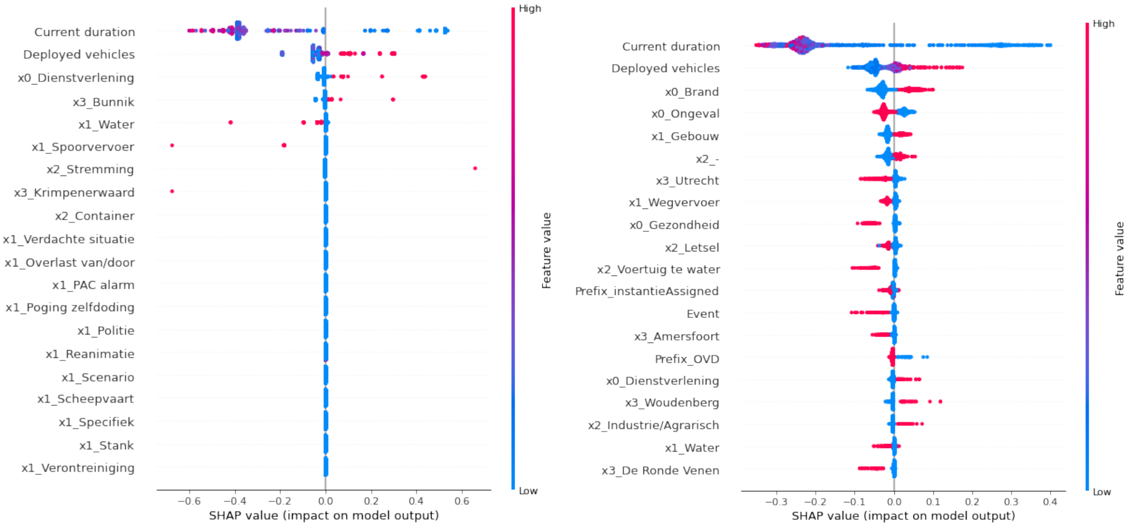
- *Current duration* $\leq$ 0.382 hour AND *rail transport* = 0.
- 0.382 $\leq$ *Current duration* $\leq$ 0.875 hour AND *Deployed vehicles* $\leq$ 4.5 AND *Municipality Bunnik* = 1.
- 0.382 $\leq$ *Current duration* $\leq$ 0.875 hour AND *Deployed vehicles* $\geq$ 4.5 AND *Water* = 0.

- $0.875 \leq$ *Current duration* ≤ 2.477 hour AND *Service* $= 0$ AND *Deployed vehicles* ≥ 6.5.
- $1.818 \leq$ *Current duration* ≤ 2.898 hour AND *Service* $= 1$.

Decision rules leading to class ‘No additional multi scale up’:

- $0.382 \leq$ *Current duration* ≤ 0.875 hour AND *Deployed vehicles* ≤ 4.5 AND *Municipality Bunnik* $= 0$.
- $0.382 \leq$ *Current duration* ≤ 0.875 hour AND *Deployed vehicles* ≥ 4.5 AND *Water* $= 1$.
- *Current duration* ≥ 0.875 AND *Service* $= 0$ AND *Deployed vehicles* ≤ 6.5.
- *Current duration leq* 2.477 AND *Service* $= 0$ AND *Deployed vehicles* ≥ 6.5.
- $0.875 \leq$ *Current duration* ≤ 1.818 hour AND *Service* $= 1$.
- *Current duration* ≥ 2.898 AND *Service* $= 1$.

For the random forest model it is not possible to extract such decision rules. With the SHAP explainer it is still possible to gain some insight in the learned behavior of the model. Both the models are explainer with SHAP such that the results can be compared. The summary plots in Figure 6.11 show the impact of the most important features of both models. As can be seen the upper two features are the same, namely *Current duration* and *Deployed vehicles*. The decision tree model is only considering a few of these features, where the decision tree is again considering more features. For the binary features, it is quiet easy to see which values of the feature have an positive and which have an negative impact on the prediction. For the two most important features, this is not easy to distinguish. Therefore, the dependency plots have to be explored for more insights.

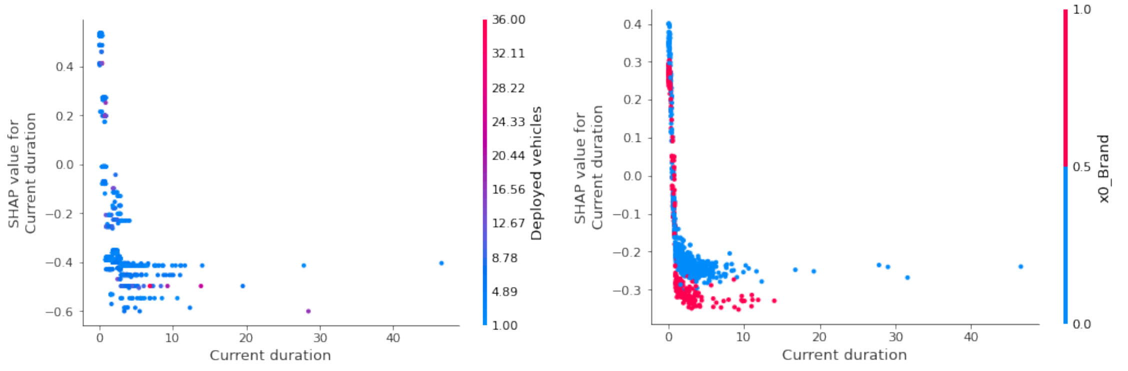


(a) Decision tree model

(b) Random forest model

Figure 6.11: SHAP summary plots for decision moment 2

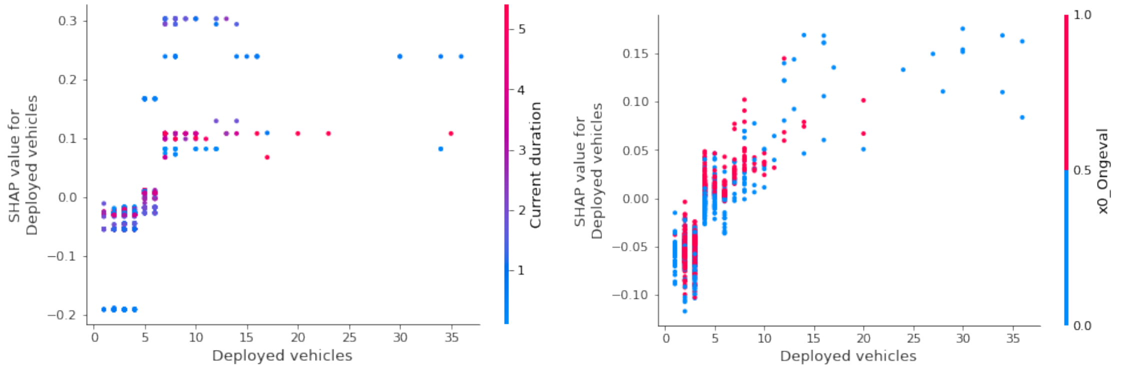
In Figure 6.12 the dependency plot of the feature *Current duration* is visualized. In this plot a similar pattern can be seen for both models. However, the scale of the y-axis differs between the models such that the decision tree model has higher negative impact for incident with a longer current duration. Furthermore, in Figure 6.13 the dependency plot of the *Deployed vehicles* is shown. Notice again the difference in the scale of the y-axis between the plots. For the decision tree model, on the one hand, there is a negative impact for less than four deployed vehicles and for five and six this is around zero. For the random forest model, on the other hand, there is a negative impact for less than tree deployed vehicles only.



(a) Decision tree model

(b) Random forest model

Figure 6.12: SHAP dependency plots for the feature Current duration of decision moment 2



(a) Decision tree model

(b) Random forest model

Figure 6.13: SHAP dependency plots for the feature Number of deployed vehicles of decision moment 2

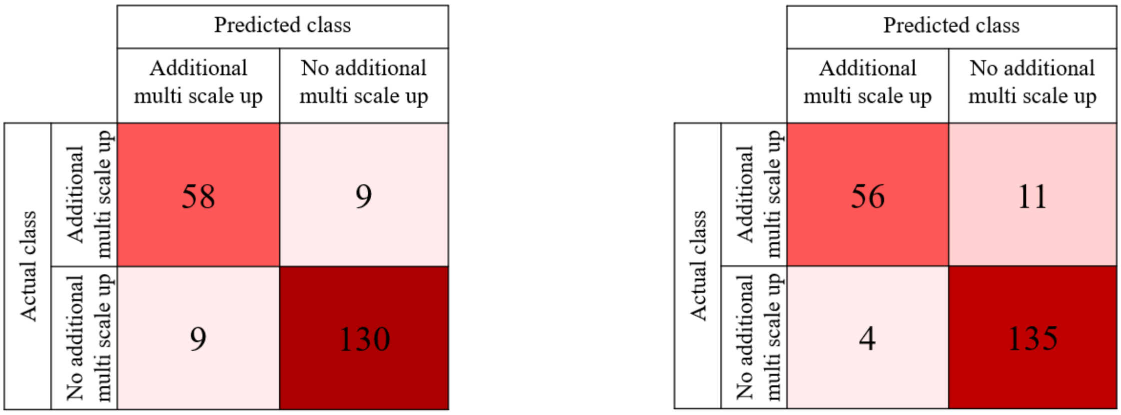
To summarize the results of the top five features of both models as explicit criteria. For the decision tree model the top five features are *Current duration*, *Deployed vehicles*, *Service*, *Municipality Bunnik* and *Water*. For the *Current duration* is hard to see in the dependency graph when there is a positive impact, however, after 1 hours it seems there is always a negative impact on predicting 'Additional multi scale up'. For the number of *Deployed vehicles* the impact is positive as more than 6 vehicles are deployed. It applies for both *Service* and *Municipality Bunnik* that they have positive impact if the feature value is equal to 1. For *Water* this is the other way around, with a negative impact if this feature value is 1. The same list can be made for the random forest model. In this model the features *Current duration*, *Deployed vehicles*, *Fire*, *Accident* and *Building* are most important. For the impact of the *Current duration* it is again hard to make a statement, however, negative impact is clearly in place after less than 1 hour already. The *Deployed vehicles* have a negative impact with less than 3 vehicles and a positive impact with more than 6 vehicles. In between it is not clear. The features *Fire* and *Building* both have an positive impact as the feature value is equal to 1. For the *Accident* it is the other way around.

6.4.3 Results Decision Moment 3

In this section the results for the final decision moment are discussed. This decision moment refers to decision point 5, but only contains the incidents that have already had multiple scale ups. Therefore, the decision to be made is between additional multi scale up or not. This decision is predicted with the decision tree model and the random forest model. The parameters for both models are optimized with a grid search. To compare the decision tree model and random forest model, first the confusion matrices are shown for both models. After which the other performance metrics are discussed. Next, the explainability of both models is explored with the SHAP summary plot and some additional dependency plots.

Performance Metrics

In Figure 6.14 the confusion matrices for the decision tree model and random forest model trained on the third data set are visualized. The decision tree model has an equal amount of *fn* and *fp*. When comparing to the random forest model, it is seen that this second model has more *fp* and less *fn*. However, both models seem to fit the data well based on the confusion matrices.



		Predicted class	
		Additional multi scale up	No additional multi scale up
Actual class	Additional multi scale up	58	9
	No additional multi scale up	9	130

(a) Decision Tree (max_depth = 9, min_samples_leaf = 4)

		Predicted class	
		Additional multi scale up	No additional multi scale up
Actual class	Additional multi scale up	56	11
	No additional multi scale up	4	135

(b) Random Forest (max_depth = 25, min_samples_leaf = 1)

Figure 6.14: Confusion matrix decision moment 3 on test data

The remaining performance metrics are summarized in Table 6.5. Again both models have good performance on all metrics. The random forest model is outperforming the decision tree model on all important metrics. For instance, the weighted f1 score of the decision tree model is 0.913 where the random forest model scores 0.926, and the macro f1 score of the decision tree model is 0.900 where this is 0.926 for the random forest model. Therefore, the random forest model seems to learn and balance everyting slightly better than the decision tree model. For the MCC score the random forest model scores around 3 % higher than the decision tree model, with a score of 0.832. Therefore, the decision tree model explains the most compared to the ground truth.

Table 6.5: Summary of performance metrics for decision moment 3, comparing the decision tree model and the random forest model.

	Precision	**Recall**	**F1 macro**	**F1 weighted**	**MCC**
Decision tree	0.900	0.900	0.900	0.913	0.801
Random forest	0.929	0.904	0.915	0.926	0.832

Overall, both models score well on all performance metrics. However, the random forest model is slightly better on all of the metrics.

Model Explainability

Both model seem to have a good performance, so next we discuss the explainability of these models. For the decision tree model the trained tree is extracted and visualized in Appendix E. The decision rules that can be extracted from this tree are listed below for each target class.

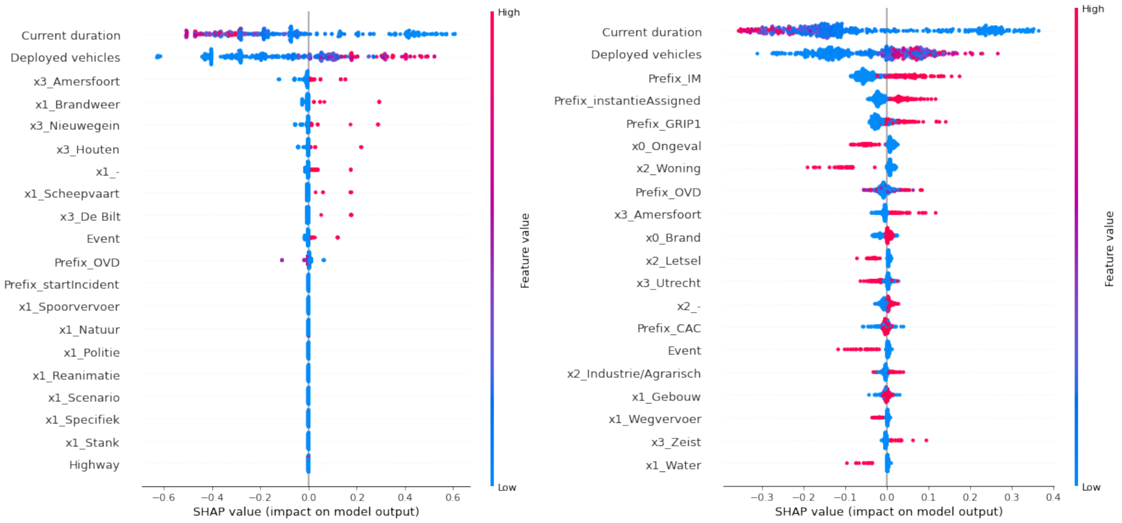
Decision rules leading to class 'Additional multi scale up':

- *Current duration* $\leq$ 0.925 hour AND *Deployed vehicles* $\geq$ 5.5 vehicle.
- *Current duration* $\leq$ 0.401 hour AND *Deployed vehicles* $\leq$ 5.5 vehicle.
- 0.925 $\leq$ *Current duration* $\leq$ 2.027 hour AND *Deployed vehicles* $\geq$ 8.5 vehicle.
- *Current duration* $\geq$ 2.027 hour AND 8.5 $\leq$ *Deployed vehicles* $\leq$ 25.5 AND *Fire brigade* = 1.
- 2.027 $\leq$ *Current duration* $\leq$ 5.858 hour AND *Deployed vehicles* $\geq$ 25.5 vehicle AND *Municipality Amersfoort* = 1
- 2.027 $\leq$ *Current duration* $\leq$ 4.374 hour AND *Deployed vehicles* $\geq$ 25.5 vehicle AND *Municipality Amersfoort* = 0.
- *Current duration* $\geq$ 5.858 AND *Deployed vehicles geq* 25.5 AND *Municipality Niewegein* = 1.
- *Current duration* $\geq$ 5.858 AND *Deployed vehicles geq* 25.5 AND *Municipality Niewegein* = 0 AND *Municipality Houten* = 1.
- *Current duration* $\geq$ 5.858 AND *Deployed vehicles geq* 25.5 AND *Municipality Niewegein* = 0 AND *Municipality Houten* = 0 AND *Municipality Amersfoort* = 1.

Decision rules leading to class 'No additional multi scale up':

- 0.401 $\leq$ *Current duration* $\leq$ 0.925 hour AND *Deployed vehicles* $\leq$ 5.5 vehicle.
- *Current duration* $\geq$ 0.925 hour AND *Deployed vehicles* $\leq$ 8.5 vehicle.
- *Current duration* $\geq$ 2.027 hour AND 8.5 $\leq$ *Deployed vehicles* $\leq$ 25.5 vehicles AND *Fire brigade* = 0.
- 4.374 $\leq$ *Current duration* $\leq$ 5.858 hour AND *Deployed vehicles* $\geq$ 25.5 vehicle AND *Municipality Amersfoort* = 0.
- *Current duration* $\geq$ 5.858 AND *Deployed vehicles geq* 25.5 AND *Municipality Niewegein* = 0 AND *Municipality Houten* = 0 AND *Municipality Amersfoort* = 0.

Furthermore, with the SHAP explainer the important features for both models are extracted and visualized in Figure 6.15. The most important features for both models are *Current duration* and *Deployed vehicles* again. Interesting to see from this plots is that the random forest model the prefix features are found important. These features tell the model something about the previous activity of the model. However, in the decision tree model this features do not come up at all.



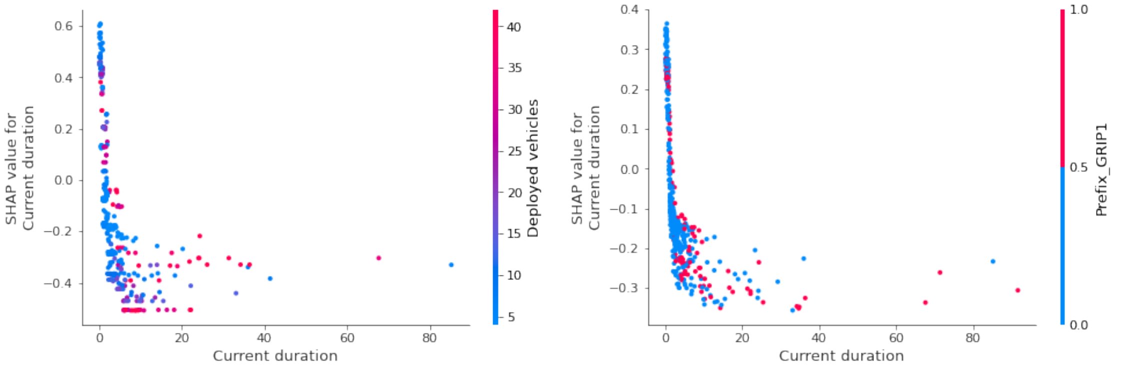
(a) Decision tree model

(b) Random forest model

Figure 6.15: SHAP summary plots for decision moment 3

Therefore, it looks like both models are actually learning different behavior to make predictions. The dependency plots for the *Current duration* and *Deployed vehicles* are explored to gain more insight into these features and when they have positive of negative impact on the prediction.

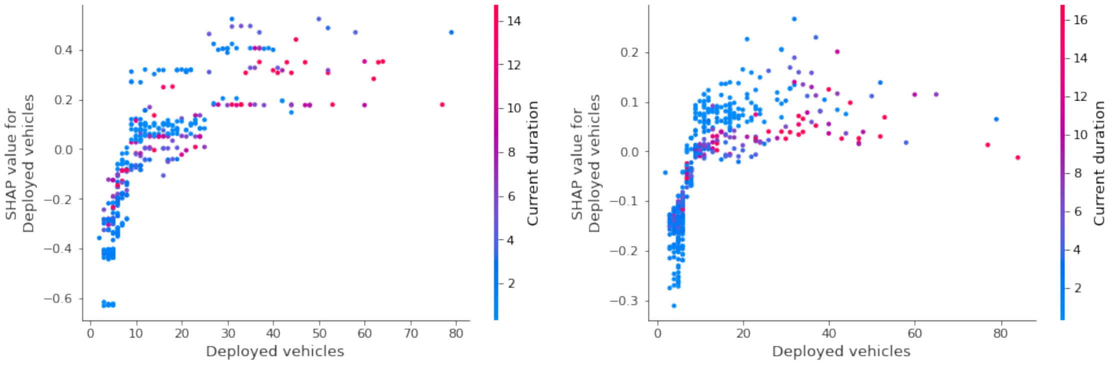
The dependency plots are shown in Figure 6.16 for the *Current duration* and in Figure 6.17 for the *Deployed vehicles*. The plots for the *Current duration* are similar for both models. Also no real conclusions can be drawn from these plots, except that for a longer duration is seems that the *Current duration* has a negative impact on the prediction of 'Additional multi scale up'. What would make sense since these incidents are already longer in the scale up loop, so can have multiple scale ups already. From the plots for the *Deployed vehicles* a similar behavior for both models is visualized as well. The larges difference between the models is in how important this feature is and therefore how large the positive and negative impact is. Both models have an negative impact for less than (approximately) 10 vehicles and a positive impact for more than (approximately) 25 vehicles.



(a) Decision tree model

(b) Random forest model

Figure 6.16: SHAP dependency plots for the feature Current duration of decision moment 3



(a) Decision tree model

(b) Random forest model

Figure 6.17: SHAP dependency plots for the feature Number of deployed vehicles of decision moment 3

Finally, the results for the SHAP explainer are summarized into a top-five of the most important criteria. In the case of the decision tree model, the features *Current duration*, *Deployed vehicles*, *Municipality Amersfoort*, *Fire brigade* and *Municipality Nieuwegein* are found most important. For the *Current duration* a value higher than 4 hours is found to have a negative impact on the prediction. Where for the *Deployed vehicles*, more than 25 vehicles seem to have a positive impact on the prediction for 'Additional multi scale up'. For the binary features *Municipality Amersfoort*, *Fire brigade*, and *Municipality Nieuwegein*, a feature value of 1 results in a positive impact on the prediction. In the case of the random forest model, the features *Current duration*, *Deployed vehicles*, *Prefix* IM, *Prefix instantie assigned* and *Prefix* GRIP *1* are found most important. For the *Current duration* a value higher than 4 hours is found to have a negative impact on the prediction. Where for the *Deployed vehicles*, less than 10 vehicles seem to have an impact on the prediction for 'Additional multi scale up'. While *Deployed vehicles* higher than 10 seem to have mostly a positive impact on the prediction. Furthermore, the prefixes IM, instantie assigned and GRIP 1, all have a positive impact on the prediction of 'Additional multi scale up' for the feature value 1.

6.5 Evaluation of Results

In the previous section, for each decision moment the most important features are identified. These features depend on the modeling technique used. In this section, the mentioned features are evaluated for their relevance in practice. For each feature their meaning in multidisciplinary scale up is of interest. Therefore, the comparison with the questionnaire is made and the features are discussed with the VRU. The discussed features are: *Current duration*, *Deployed vehicles*, *Classification criteria 1*, *Classification criteria 2*, *Municipality*, and *Prefix*. Additionally is discussed which model is preferred based on the findings of these most important features.

The first feature discussed is the *Current duration* of an incident. In the questionnaire this feature is not explicitly mentioned. What is mentioned in the questionnaire are the criteria 'Expected duration' and 'Duration of incident unknown'. However, in the organizational objectives of the VRU is mentioned that they 'make an effort to quickly repair disturbances of daily life', indicating that time and therefore duration of an incident are important for the VRU. The *Current duration* provides insight in how long an incident is already ongoing before arriving in the specific decision point. Notice that according to the model, a short current duration has a positive impact on predicting multidisciplinary scale up in all decision points. From the perspective of the VRU, the suggestion is made, that these incidents have underlying reasons for scale up. However, based on the process model, these incidents seem to arrive in this decision point earlier than incidents that do not require scale up. So the interpretation received from this feature is that a shorter current duration of the incident is an indication for multidisciplinary scale up. Besides, it can be concluded that the current duration is the most important feature in all decision points and models with similar interpretation, shorter current duration has higher positive impact on predicting multidisciplinary scale up. Besides, a longer current duration has a high negative impact on predicting multidisciplinary scale up.

The second feature is *Deployed vehicles* at the incident. This feature can be linked with the criteria 'Own disciplines involved' from the questionnaire. If there are more vehicles deployed by the incident, this results in more people that have to collaborate. Obviously, more people working together also require adequate coordination. The feature *Deployed vehicles* is mentioned as the second most important in all of the models and decision points. However, the specific interpretation for this criterion changes with the different decision points. Additionally, between the two models for the same decision point, the interpretation changes only slightly. Later in the process model, the relevant number of deployed vehicles has a positive impact on predicting multidisciplinary scale-up increases. In other words, the first decision point found results in decision moment 1. In decision moment 1, the highest positive impact is for 0 deployed vehicles. While for decision moment 2 this positive impact is after more than 5 vehicles are deployed and for decision moment 3 after more than 10 vehicles are deployed. This increase is logical since the sequence of the decision moments is also the sequence in time. Over time more vehicles are assigned to the incident. Overall, this feature is logical in the context and the contribution makes sense to the VRU.

The third and fourth feature identified are *Classification criteria 1* and *Classification criteria 2*. Both features relate to the criterion 'incident type' mentioned in the questionnaire. Since these are categorical feature, we look at the specific categories. For *Classification criteria 1* this are the categories 'Healthcare' and 'Accident' mentioned in decision moment 1, and the categories 'Service' and 'Fire' mentioned in decision moment 2. In discussion with the VRU about the interpretation of these categories, the insight was gained that 'Healthcare' is often related with *Classification criteria 2* 'Resuscitation'. For these incidents, all disciplines are alarmed because time is of the essence. However, these incidents are often small and do not need all of these people there for long. Therefore, it is logical that the category healthcare has a negative impact on the prediction of multidisciplinary scale up. Furthermore, it is logical that this category is important in decision moment 1 only because there the decision is made between routine and scale-up. In the case of the category 'Accident' an OvD Fire brigade is often alarmed together with a rescue vehicle. However, in general, that is all scale-up required for a normal accident. Therefore, this classification criterion has a positive impact on predicting multidisciplinary scale up in the first decision moment, but negative in the second decision moment. The category 'Service' is often combined with *Classification criteria 2* Police or Fire brigade. This incident type refers to alarming for assistance of that discipline. The category 'Service' does not provide insight in incident specifications like the category 'Fire' does, for example. The positive impact of 'Service' is remarkable and cannot completely be placed in context. The category 'Fire' is often related to *Classification criteria 2* 'Building'. This category has positive impact on 'Additional multidisciplinary scale up'. Possibly because a fire in a building might results in many different tasks for all disciplines. Overall, the classification criteria is relatable with the incident type. However, for the category 'Service' this is not the case, it is more of a registration name. Moreover, the combination of *Classification criteria 1* and *Classification criteria 2* often provides more insight than approaching these criteria separately.

Furthermore, the feature *Municipality* is found as important feature. This feature is added to provide information about the location of an incident which is mentioned in the questionnaire. The categories found important by the model are 'Bunnik' in decision moment 2 and 'Amersfoort', 'Nieuwegein' in decision moment 3. However, when the context of this feature is explored and discussed with the VRU, the feature *Municipality* does not seem to match the the aim of adding this feature. According to the VRU the number of incident is a *Municipality* is related with the number of inhabitants. In that context it is logical that large municipalities have more multidisciplinary incidents because there are more incidents in general. Overall, is concluded that the feature *Municipality* does not represent the criteria 'Incident location' well.

Finally, the feature *prefix* is discussed. This feature describes the previous executed activities for this incident. For example, 'Instantie assigned' is one of the activities that is sometimes executed. As can be expected, the prefix is mainly relevant in bucket 3 because of the loop in this decision decision moment results in repetition of incident in this point. These prefixes are found important in the random forest model more than the decision tree model. This feature is logical in the context of the incident because the previous taken actions do indicate what would be a valuable next step to handle the incident.

To conclude, the features *Deployed vehicles* and *Classification criteria 1* and *2* are most recognized by the VRU. For the interpretation of the *Current duration* they are a bit hesitating. However, this feature is included with good evidence so should be kept. The feature *Municipality* is not convincing to be objective in discussion with the VRU. Therefore, this is not found a good feature to draw decision on. The *Prefix* is intuitive for the VRU. Overall, the most important features found by the random forest model are easier to explain in the context of the incidents. Therefore, the random forest model is selected as best overall performing model in this research.

6.6 Chapter Overview

In this chapter, the interesting decision points found in chapter 5 are transformed to a three separate binary learning problems for the three identified decision moments. For each of these decision moment, features are added that are created based on the found important criteria in chapter 4. For each decision moment, two models are fitted; the decision tree model and the random forest model. The models are compared based on their performance with the confusion matrix, f1 score and MCC. Furthermore, the model explainability is important. The top 5 features used by each model at each decision point are evaluated. In this evaluation the practical relevance is explored and discussed with the VRU. Therefrom can be concluded that not all found features are relevant in practice, such as the feature *Municipality.* However, the features *Current duration*, *Deployed vehicles*, *Classification criteria 1*, *Classification criteria 2* and *Prefix* are found valuable and aligned with the implicit knowledge.

Chapter 7

Conclusion and Recommendations

The aim of this thesis was to discover explicit criteria for the implicit knowledge used by decision makers for scale up in multidisciplinary incidents. Therefore, the decision mining approach is applied. The research conclusion is provided in section 7.1. Also, some recommendations are provided in section 7.2. To conclude, a discussion about the scientific contribution and limitations of this research is provided in section 7.3.

7.1 Conclusion

To conclude this research, the formulated research questions are answered based on the knowledge gathered in this research. To start with the first question.

Q1: *Which data is available about the multidisciplinary scale up of incidents?*

In chapter 4 is described how three different data sets are selected. Two of these data sets contain incident-based information and the other vehicle-based information. It was found that these data sets contain mainly data related to the fire brigade. However, an attribute defined the involved disciplines which made it possible to select all incident numbers related to multidisciplinary incidents from the data. Only these incident numbers selected as multidisciplinary incidents are used for further research.

The data contains information about time at which the incident started and disciplines are alarmed. Additionally, information about scale ups is available with timestamps. Furthermore, some basic incidents descriptions are provided as well, like the incident type or location.

Q2: *How can implicit knowledge be acquired and taken into account for explicit criteria?*

With the aim to capture implicit knowledge, this implicit knowledge needs to be explored. Based on the decision-aware mining approach described in Petrusel et al. (2011) a questionnaire is used for knowledge acquisition. In this questionnaire incident descriptions were provided. Each decision-maker had to judge this incident and define their decision, but also justify why they make this decision rather than another decision. The justification was in open text as well as prepared criteria. During analysis, it can be concluded that the prepared criteria captured most of the provided answers. These criteria describe the consideration of the decision maker. The criteria taken into account most are: ‘Incident location’, ‘Incident size’, ‘Incident type’, ‘(number of) injured and injury classification’, ‘Sensitivity incident on social media’, ‘Own disciplines involved’, ‘Expected duration incident’, ‘Possible effects on people/material’, ‘Safe/unsafe area’, ‘Involved partners’ and ‘Duration of incident unknown’.

Q3: *How does the current multidisciplinary scale up process look like and what activities are involved?*

With the timestamps in the data, an event log is created to apply process mining. The activities found in the data related to multidisciplinary scale up are 'Start incident', 'Partner assigned', 'IM', 'OvD Fire Brigade', 'OvD Medical', 'OvD Police', 'CAC' and 'End incident'. However, it should be noticed that the distribution of these events refers that not all OvD activities are in the data. Because for the OvD Fire brigade many more activities are found than for the other OvDs. This was expected due to the fact that the data is fire brigade related.

Furthermore, the best process model was discovered with the inductive miner. The best model was found with the subset where the start activity is 'Start incident' and the different OvD activities are renamed by one activity 'OvD'. This model has as advantage that it is sound and understandable. The disadvantage is that the model contains a loop. From this model five decision points can be identified.

Q4: *Which machine learning techniques can be used to predict multidisciplinary scale up while the model is explainable to distract decision rules?*

From the five identified decision points, two decision points are selected in discussion with the VRU to explore further. These decision points are transformed to a binary learning problem to apply machine learning techniques. Therefore, bucketing in applied to divide the data in three data sets first. Next, the possible machine learning techniques are explored.

In previous decision mining application is chosen for the decision tree model, because of the glass box structure. The decision tree model learns decision rules that can be extracted by visualizing the trained tree. However, decision trees might suffer from low accuracy, therefore, the model is compared with a more complex model, the random forest model. Nevertheless, random forest model might have a better accuracy in general, they are also less explainable. Therefore, the model should be combined with the SHAP explainer to extract decision rules.

After fitting both models on the data, they are compared to each other. Both seem to have good performance on the three different data sets (buckets). In terms of explainability, from the tree of the decision tree model it is easy to extract simple and concrete decision rules. For the SHAP explainer these decision rules are less concrete. It is possible to define features that have an positive (or negative) impact on the prediction.

Q5: *Do the found explicit criteria with machine learning match with implicit knowledge of the decision-makers?*

The performance of the model shows that the human decision can be predicted by the model accurately. Since the input of the model is based on the criteria found in the questionnaire only, the explicit criteria match with the implicit knowledge. However, there was some translation required from the found criteria in the questionnaire towards the data features. Therefore, the criteria are not one on one. The features found important by the models are discussed with the VRU for interpretation purposes. Therefrom can be concluded, that most features do represent the mentioned implicit knowledge, except for the feature *Municipality.* The features found representative are *Current duration*, *Deployed vehicles*, *Classification criteria 1*, *Classification criteria 2* and *Prefix.*

For the three different data sets, different decision rules and important features are extracted. The features *Current duration* and *Deployed vehicles* are identified as most important features for all data sets and models. However, the interpretation of these features is different for the decision points.

RQ: *Is it possible to define explicit criteria to support decision makers, leading to the most appropriate multidisciplinary scale up in crisis management?*

To answer the main research question, it is possible to define explicit criteria to support decision-makers. If these criteria lead to the most appropriate multidisciplinary scale up is not answered. The criteria found in this research are generic guidelines for how the decisions are made by decision-makers right now. With the found explicit criteria it is possible to apply the current judgement on new situations. This provides a general perspective based on the perspectives of all different decision-makers. Therewith, the aim is met to capture criteria independent of the decision-makers view point.

The found criteria are distinct in three decision moments. For each decision moment other criteria sets are found, which are summarized here. In the first decision moment, a short current duration, 0 or more than 4 deployed vehicles and the classification criterion accident, have all a positive impact on the outcome 'Multidisciplinary scale up required'. While a current duration longer than 1 hour, number of deployed vehicles between 0 and 4, the classification criteria resuscitation and healthcare have a negative impact on the outcome 'Multidisciplinary scale up required'. In the second decision moment, a short current duration, more than 6 deployed vehicles, the classification criteria fire and building have a positive impact on the outcome 'Additional multidisciplinary scale up'. Furthermore, a duration longer than 1 hour, less than 4 deployed vehicles and the classification criteria accident have a negative impact on the outcome 'Additional multidisciplinary scale up'. Finally, for the third decision moment, a short duration, more than 10 deployed vehicles and the prifix 'IM', 'Instantie assigned' and 'GRIP 1' have all a positive impact on the outcome 'Additional multidisciplinary scale up'. Additionally, a current duration longer than 4 hours and less than 10 deployed vehicles have a negative impact on the outcome 'Additional multidisciplinary scale up'.

7.2 Recommendations

In this thesis, criteria are found that capture the current decisions of decision-makers within multidisciplinary scale up for the VRU. These criteria might not align with the desired scale-up procedure by the VRU. The executed research provides insight and evidence for the executed multidisciplinary scale up process rather than improvements for this process. Insight is the first step towards improvements. In order for the VRU to use these insight for improvement in the scale up procedure it is recommend to evaluate the explicit criteria. They should consider the aim to solve incidents as soon as possible while looking for improvements. Subsequently, the explicit criteria can be refined based on an evaluation with the desired incident outcomes.

Thereby, the VRU should consider what 'wrong' decision should be penalize more. Whether the decision to scale up early while the incident can be handled with the disciplines on site, or no scale up with a long lasting incident in return. Both situations do not necessarily have to be the wrong decision, but could be less appropriated for the situation on hand. When considering what situation they want to avoid, they could take more risk in one of these directions. Currently it seems that no scale up is preferences by decision-makers to avoid having more people involved than necessary. The question for the VRU is, if they agree with this preferred decision, or would like to change it. The decision rules can be adapted based on the preferences of the VRU in this cases. The awareness of this consideration could help in appropriate decision-making.

The explicit criteria found do not include all knowledge of decision-makers. In the questionnaire more important criteria were found than implemented in the machine learning model. Mainly because there was no appropriate data available to define this knowledge. Besides, some of this knowledge is not quantifiable. The criteria found important but not included are: 'Incident size', '(number of) injured and injury classification', 'Sensitivity incident on social media', 'Expected duration incident', 'Possible effects on people/material' and 'Safe/unsafe area'. For the criteria 'Incident size', 'Sensitivity incident on social media', 'Possible effects on people/material' and 'Safe/unsafe area', it is not directly possible to express these criteria as decision rules. These criteria are highly depending on human interpretation. Possible for the VRU to see if these criteria are translatable to data features in the future. However, these criteria are not found in the data used for this research so other data sources should be checked. The current data also did not contain information about the number of injured or injury classification. When data is available for these features, they are interesting to test for their predictive value.

Furthermore, the data used to create the feature *Deployed vehicles* is mainly based on the vehicles of the fire brigade only. Because this is one of the main features, it is of interest to include data from the other disciplines as well. Expected is that this will increase the value of this feature even more. Especially for incidents that are less fire brigade related. The feature is more general when the vehicles of other disciplines are known as well.

7.3 Discussion

In this research a few limitations were encountered. First of all, the incident data used is provided by the fire brigade. Therefore, the data is incomplete from a multidisciplinary viewpoint. This results in missing OvD activities in the event log from the medical team, police and population care. This incomplete data is also influencing one of the most important features of the model, namely the *Deployed vehicles*. The feature is now based on deployed vehicles by the fire brigade only, which is incomplete from a multidisciplinary viewpoint. Secondly, in discussion with the VRU it came up that the deployment of IM, CAC, and ROL is not always registered. It occurs that there is contact between the CaCo (or VIC) and these people but this is not registered. In practice this works for them. However, when we look at the data only this information is lacking while it could be valuable to consider. Thirdly, this research could have provided more insights by deriving specific criteria for each multidisciplinary activity separately. Indented are the activities: 'IM', 'OvD', 'CAC', 'GRIP 1', and 'GRIP 2'. However, there is a large data imbalance between these activities. Therefore, the choice is made to focus on binary classification only. Finally, the found criteria are a bit superficial, while the questionnaire did provide sufficient suggestions for good criteria. This information provided by the questionnaire is not appreciated enough by this research. It seems that there is not enough emphasis on consistent data gathering within the VRU yet to retrieve these criteria as features. Suggested is that more insights can be gained from knowledge gathering.

The scientific contribution of this research is in twofold. First of all, this research has explored a new application domain for decision mining. With this research is demonstrated that explicit scale-up criteria can be found for crisis response. Therefore, the dynamic and unpredictable domain of crisis response is identified as application domain for decision mining. Secondly, the loop structure in the process model that was defined as challenging by Rozinat and van der Aalst (2006) is handled. In this research this is handled by creating buckets, as used in predictive process monitoring.

Future research could be focused on extending the found criteria by discovering specific criteria for each multidisciplinary activity (IM, OvD, CAC, GRIP 1 and GRIP 2). Moreover, this research insight is gained in explicit scale-up criteria but no improvement steps are taken. Future research could search for possible improvements based on the discovered knowledge in this research, to improve the crisis response procedure of the VRU.

Bibliography

Batoulis, K., Meyer, A., Bazhenova, E., Decker, G. & Weske, M. (2015). Extracting decision logic from process models. *International conference on advanced information systems engineering*, 349–366.

Bazhenova, E., Buelow, S. & Weske, M. (2016). Discovering decision models from event logs. *255*, 237–251. https://doi.org/10.1007/978-3-319-39426-8_19

Bazhenova, E. & Weske, M. (2016). Deriving decision models from process models by enhanced decision mining. *International conference on business process management*, 444–457.

Bennet, B. (2011). Effective emergency management: A closer look at the incident command system. *Professional Safety*, *56*(11), 28–37.

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5–32.

Campos, J., Richetti, P., Baião, F. A. & Santoro, F. M. (2017). Discovering business rules in knowledge-intensive processes through decision mining: An experimental study. *International Conference on Business Process Management*, 556–567.

Chicco, D. & Jurman, G. (2020). The advantages of the matthews correlation coefficient (mcc) over f1 score and accuracy in binary classification evaluation. *BMC genomics*, *21*(1), 1–13.

data science group, P. (n.d.). *Conformance checking.* http://www.processmining.org/conformance.html#what

De Leoni, M. & van der Aalst, W. M. (2013). Data-aware process mining: Discovering decisions in processes using alignments. *Proceedings of the 28th annual ACM symposium on applied computing*, 1454–1461.

De Smedt, J., vanden Broucke, S. K., Obregon, J., Kim, A., Jung, J.-Y. & Vanthienen, J. (2016). Decision mining in a broader context: An overview of the current landscape and future directions. *International Conference on Business Process Management*, 197–207.

Di Ciccio, C., Marrella, A. & Russo, A. (2015). Knowledge-intensive processes: Characteristics, requirements and analysis of contemporary approaches. *Journal on Data Semantics*, *4*(1), 29–57.

Ernst, M. D., Cockrell, J., Griswold, W. G. & Notkin, D. (2001). Dynamically discovering likely program invariants to support program evolution. *IEEE transactions on software engineering*, *27*(2), 99–123.

Fahland, D. (2021). Extracting and pre-processing event logs.

Herrera, M. P. R. & Díaz, J. S. (2019). Improving emergency response through business process, case management, and decision models. *ISCRAM.*

Hertwig, R., Barron, G., Weber, E. U. & Erev, I. (2004). Decisions from experience and the effect of rare events in risky choice. *Psychological science*, *15*(8), 534–539.

Kotsiantis, S. B., Zaharakis, I., Pintelas, P. et al. (2007). Supervised machine learning: A review of classification techniques. *Emerging artificial intelligence applications in computer engineering*, *160*(1), 3–24.

Kushnareva, E., Rychkova, I. & Le Grand, B. (2015). Modeling business processes for automated crisis management support: Lessons learned. *2015 IEEE 9th International Conference on Research Challenges in Information Science (RCIS)*, 388–399.

Leemans, S. J., Fahland, D. & van der Aalst, W. M. (2013). Discovering block-structured process models from event logs-a constructive approach. *International conference on applications and theory of Petri nets and concurrency*, 311–329.

Leewis, S., Berkhout, M. & Smit, K. (2020). Future challenges in decision mining at governmental institutions.

Leewis, S., Smit, K. & Zoet, M. (2020). Putting decision mining into context: A literature study. *Digital business transformation* (pp. 31–46). Springer.

Leoni, M. d., Dumas, M. & García-Bañuelos, L. (2013). Discovering branching conditions from business process execution logs. *International Conference on Fundamental Approaches to Software Engineering*, 114–129.

Liu, H., Gegov, A. & Cocea, M. (2017). Rule based networks: An efficient and interpretable representation of computational models. *Journal of Artificial Intelligence and Soft Computing Research*, *7*.

Lundberg, S. (2018). *Shap documentation.* https://shap.readthedocs.io/en/latest/index.html

Luque, A., Carrasco, A., Martín, A. & de Las Heras, A. (2019). The impact of class imbalance in classification performance metrics based on the binary confusion matrix. *Pattern Recognition*, *91*, 216–231.

Mannhardt, F., De Leoni, M. & Reijers, H. A. (2015). The multi-perspective process explorer. *BPM (Demos)*, *1418*, 130–134.

Mannhardt, F., De Leoni, M., Reijers, H. A. & Van Der Aalst, W. M. (2016). Decision mining revisited-discovering overlapping rules. *International conference on advanced information systems engineering*, 377–392.

Meeuwis, J. (2021). *Predicting crises with real-time incident information.*

Mertens, S., Gailly, F., Van Sassenbroeck, D. & Poels, G. (2020). Integrated declarative process and decision discovery of the emergency care process. *Information Systems Frontiers*, 1–23.

Mitchell, T. (1997). *Machine learning.* McGraw-Hill.

Myles, A. J., Feudale, R. N., Liu, Y., Woody, N. A. & Brown, S. D. (2004). An introduction to decision tree modeling. *Journal of Chemometrics: A Journal of the Chemometrics Society*, *18*(6), 275–285.

Petrusel, R. (2010). Decision mining and modeling in a virtual collaborative decision environment. *Decision Support Systems, Advances in*, 215.

Petrusel, R., Vanderfeesten, I., Dolean, C. C. & Mican, D. (2011). Making decision process knowledge explicit using the decision data model. *International Conference on Business Information Systems*, 172–184.

Rozinat, A. & der Aals, W. V. (2006). *Decision mining in business processes.* Technische Universiteit Eindhoven.

Rozinat, A. & Van der Aalst, W. M. (2008). Conformance checking of processes based on monitoring real behavior. *Information Systems*, *33*(1), 64–95.

Rozinat, A. & van der Aalst, W. M. (2006). Decision mining in prom. *International Conference on Business Process Management*, 420–425.

Shahrah, A. Y. & Al-Mashari, M. A. (2017). Emergency response systems: Research directions and current challenges. *Proceedings of the Second International Conference on Internet of things, Data and Cloud Computing*, 1–6.

Slam, N., Wang, W., Xue, G. & Wang, P. (2015). A framework with reasoning capabilities for crisis response decision–support systems. *Engineering Applications of Artificial Intelligence*, *46*, 346–353.

Smirnov, A., Pashkin, M., Levashova, T., Shilov, N. & Kashevnik, A. (2007). Role-based decision mining for multiagent emergency response management. *International Workshop on Autonomous Intelligent Systems: Multi-Agents and Data Mining*, 178–191.

Teinemaa, I., Dumas, M., Rosa, M. L. & Maggi, F. M. (2019). Outcome-oriented predictive process monitoring: Review and benchmark. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, *13*(2), 1–57.

Theeuwes, N. (2019). *Formalization and improvement of ambulance dispatching in brabant-zuidoost.*

Van Der Aalst, W. (2016). *Process mining: Data science in action* (Vol. 2). Springer.

Van der Aalst, W. (2011). *Process mining : Discovery, conformance and enhancement of business processes*. Springer. https://doi.org/10.1007/978-3-642-19345-3

Van Der Aalst, W., Adriansyah, A., De Medeiros, A. K. A., Arcieri, F., Baier, T., Blickle, T., Bose, J. C., Van Den Brand, P., Brandtjen, R., Buijs, J. et al. (2011). Process mining manifesto. *International conference on business process management*, 169–194.

Van der Aalst, W., Adriansyah, A. & van Dongen, B. (2012). Replaying history on process models for conformance checking and performance analysis. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *2*(2), 182–192.

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

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 Dongen, B., Carmona, J. & Chatain, T. (2016). Alignment-based metrics in conformance checking (summary). *Fachgruppentreffen der GI-Fachgruppe Entwicklungsmethoden für Informationssysteme und deren Anwendung*, 87–90.

van Duin, M. & Wijkhuijs, V. (2017). Boom bestuurskunde.

van Duin, M., Wijkhuijs, V. & Jong, W. (2018). Boom bestuurskunde.

VRU, A. P. V. (2020). *Regionaal crisisplan utrecht 2020-2023*.

Weijters, A. J. M. M., van der Aalst, W. M. & de Medeiros, A. K. A. (2006). Process mining with the heuristicsminer algorithm.

Wijkhuijs, V. & van Duin, M. (2020). Boom bestuurskunde.

Wirth, R. & Hipp, J. (2000). Crisp-dm: Towards a standard process model for data mining. *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining, 1*.

Appendix A

Available Data Overview

Table A.1: Basic Data VP

Attribute	**Type**
Incident	Integer
Datum	Temporal
Adres	String
Huisnummer	Integer
Toevoeging	String
Postcode	String
Plaats	String
Meldinsclassificatie 1	Categorical
Meldingsclassificatie 2	Categorical
Meldingsclassificatie 3	Categorical
Prioriteit	Categorical
Gemeente	String
Kazerne	String
OMS	Integer
Object (incident)	Categorical
Naam locatie 1	String
Type locatie 1	Categorical
Naam locatie 2	String
Type locatie 2	Categorical
Uitruk prioriteit	Categorical
Object (verslag)	String
Soort inzet	Categorical
Soort melding	Categorical

Table A.2: General incident data

Attribute	**Type**
DTG start incident	Temporal
DTG einde incident	Temporal
DTG start incident BRW	Temporal
DTG einde incident BRW	Temporal
Incident Nr.	Integer
BRW betrokken	Categorical
POL betrokken	Categorical
CPA betrokken	Categorical
DTG opdracht 1e eenheid BRW	Temporal
Prioriteit BRW	Categorical
Niveau 1	Categorical
Niveau 2	Categorical
Niveau 3	Categorical
Naam locatie 1	String
Huisnummer	Integer
Huisnummer toev.	String
Huisletter	String
X coordinaat	Integer
Y coordinaat	Integer
Naam locatie 2	String
Gemeente naam	String
NaamInstantie	String
DTG aanmaak	Temporal
Naam persoon	String
Funcitesoort	String
DTG opdracht	Temporal
Ploegsoort	Categorical
Ploeg naam	String
DTG opdracht	Temporal

Table A.3: Vehicle data

Attribute	**Type**
Relevant	Categorical
Uitruktijd relevant	Categorical
Opkomsttijd relevant	Categorical
Negeren in rapportages	Categorical
Incidentnummer	Integer
Inzet UID	Integer
Negeren alarmeringtijd	Categorical
Negeren uitgerukttijd	Categorical
Negeren terplaatsetijd	Categorical
Negeren ingerukttijd	Categorical
Negeren terugkazernetijd	Categorical
Normtijd dekkingsplan	Temporal
Normtijd	Temporal
Zorgnorm eenheid	String
Zorgnorm kazerne	String
Partij	String
Uitruknaam	String
Voertuig prioriteit	Categorical
CBS loosalarm	Categorical
CBS soort loosalarm	Categorical
GMS BRW melding CL	Categorical
GMS BRW melding CL1	Categorical
GMS BRW melding CL2	Categorical
GMS BRW soort aflsuiting	Categorical
GMS code voertuigsoort	Categorical
GMS DTG melding	Temporal
GMS DTG start incident	Temporal
GMS BRW DTG start incident	Temporal
GMS DTG brandweerstart	Temporal
GMS DTG alarmering	Temporal
GMS DTG uitgerukt	Temporal
GMS DTG terplaatse	Temporal
GMS DTG ingerukt	Temporal
GMS DTG terugkazerne	Temporal
GMS BRW DTG einde incident	Temporal
GMS DTG einde incident	Temporal
GMS gemeente naam	String
GMS huis paal nr	Integer
GMS huis nr toev	String
GMS kaz naam	Categorical
GMS naam locatie1	String
GMS naam locatie2	String
GMS naam melder	String
GMS naam voertuig	String
GMS nr incident	Integer
GMS object functie naam	Categorical
GMS plaats naam	String
GMS postcode	String
GMS prioriteit incident brandweer	Categorical
GMS T X coord loc	Integer
GMS T Y coord loc	Integer
GMS type locatie1	Categorical
GMS vak nr	Integer

Appendix B

Questionnaire

Dear Sir / Madam,

As part of my graduation assignment at Eindhoven University of Technology towards multidisciplinary scale up in crisis situations within the Veiligheidsregio Utrecht, I would like to ask you to complete this questionnaire.

The list contains questions about your considerations for (or not) multidisciplinary scale up from your own crisis role. Answering the questions will take approximately 25 minutes. The questionnaire outlines three different (crisis) situations. The questions related to these situations relate to your personal assessment of the situation regarding multidisciplinary scaling. Following are a few short general questions.

The questionnaire can be completed anonymously, but there is also the option to leave contact details if you are interested in helping answer any follow-up questions. Finally, I would like to emphasize that there are no right or wrong answers, it is your personal opinion that matters.

Thanks in advance for your cooperation,

Britt Lukassen

Question: What crisis role(s) do you fulfill?

- OvD Fire brigade
- OvD Police
- OvD Population care
- OvD Medical
- CaCo
- HOvD Fire brigade/ Leader CoPi
- ROL of duty

Outline questionnaire

The questionnaire is structured as follows:

- Report description crisis situation 1
  - Two questions
- Additional picture of crisis situation 1

    - Four questions
- Report description crisis situation 2
    - Two questions
- Additional picture of crisis situation 2
    - Four questions
- Report description crisis situation 3
    - Two questions
- Additional picture of crisis situation 3
    - Four questions
- Multidisciplinary scale up of general questions
    - Three questions

The crisis situations described are based on actual incidents with dynamic aspects that may complicate the decision to scale up multidisciplinary. That is why I would like to ask you to approach these crisis situations as much as possible as an incident to which you are (or can be) linked in your crisis role. I ask several times about points for attention and criteria, trying to use clear terminology. Think, for example, of the expected duration of the incident, victims, area of effect, meteorological conditions, etc..

To clarify the term flexible scaling: various scaling ups are possible. You can think of scaling up within the monodisciplinary column, but also involving an operational core team or flexibly scaling up within the GRIP structure. When asked about flexible scaling, you have the opportunity to explain yourself in more detail.

Crisis situation 1 - report description

Around 01:00 am on the night of Tuesday 23 April to Wednesday 24 April, a report was received of a single-vehicle accident in Woerden on the A12 in the direction of The Hague. The car came to a stop against a matrix column that stands between the crash barrier and the noise barrier. According to the reporter, there are several people in the vehicle, and no fire appears to have started. After the reporter went to look, it turns out that there are four victims who cannot be contacted.

Question: Do you currently want to scale up multidisciplinary with the information that is now available?

- No
- Yes, scale up to GRIP
- Yes, flexible scale up
- Other,...

Question: What reason(s) do you have for this?
...

Question: What information would you like to know to determine whether this incident can lead to multidisciplinary scaling up?
...

Crisis situation 1 - additional information

Repeat:

Around 01:00 am on the night of Tuesday 23 April to Wednesday 24 April, a report was received of a single-vehicle accident in Woerden on the A12 in the direction of The Hague. The car came to a stop against a matrix column that stands between the crash barrier and the noise barrier. According to the reporter, there are several people in the vehicle, and no fire appears to have started. After the reporter went to look, it turns out that there are four victims who cannot be contacted.

New information:

Involved:

- OvD Police with two emergency assistance units and Traffic Accident Analysis
- OvD Medical with three ambulances and a trauma team
- OvD Fire Brigade with a tanker sprayer and emergency vehicle

The duty officer (OvD) Medical who is alerted after the accident rushes to the scene. The highway to The Hague is already closed to traffic by Rijkswaterstaat at that time. Because a child seat was found on the road that may have come from the car, it is taken into account that there was still a fifth occupant. The location of the accident is being explored using a police helicopter. Meanwhile, it appears that four young men have died on the spot. Relatives of the victims arrive at the scene around 2 a.m., partly via the opposite lane, which is still open to traffic at that time. There is some commotion and because more family members may be on their way to the scene of the accident, the Public Prosecutors of the involved (emergency) services decide to close the A12 completely. In addition, the Turkish relatives indicate that one of the killed victims would get married within a few days.

Question: Do you currently want to scale up multidisciplinary with the information that is now available?

- No
- Yes, scale up to GRIP
- Yes, flexible scale up
- Other,...

Question: What reason(s) do you have for this?
...

Question: Which points of attention, in the described situation, give reason for multidisciplinary scale up from your perspective?
...

Question: From your perspective, which points of attention, in the described situation, give no reason for multidisciplinary scale up?
...

Question: Which of the incident characteristics listed below do you take into account when considering multidisciplinary scaling?

Incident:

- Incident type
- Incident size

- Expected duration incident
- Incident location
- Timestamp incident

Risks and safety:

- Possible involvement of hazardous substances/smoke
- Possible perpetrators/fleeing suspects
- Possible effects on people/material
- Safe/unsafe area
- Safe/unsafe approach route
- Applicable (safety) procedures

Meteo:

- Wind direction/ wind speed
- Temperature
- Rainfall/ clouds

Victims/population:

- (Number of) injured and injured classification
- Population specifications (for example: in a certain neighbourhood)
- Reception locations/ relatives

Environmental analysis:

- Vulnerable objects in the environment (for example: nursery/town hall)
- Care objects
- Events/demonstrations and/or other activities
- Municipal or regional border crossing incident
- Busy location/flow-through location

Communication:

- Sensitivity incident on social media
- Use NL-alert and/or other means of communication
- Action perspective communication

Services involved:

- Own disciplines involved
- Involved partners (also nationally)

Missing information:

- Cause of incident
- Duration of incident unknown
- Planning and scenarios unknown

Question: If desired, you can provide a brief explanation here and name any missing features.
...

Crisis situation 2 - report description

On Sunday 28 August, at 4.26 pm, the emergency room received an automatic report of a fire allegedly raging in a nursing home in Nieuwegein. According to the procedure that applies to automatic fire alarms, the dispatcher of the control room contacts the supervisor on duty in the care home. The fire alarm comes from room 280 on the second and top floor. When an emergency response officer arrives on the floor, she encounters a closed compartment door behind which nine apartments are located. There is now a thick layer of black smoke in the hallway behind the door.

Question: Do you currently want to scale up multidisciplinary with the information that is now available?

- No
- Yes, scale up to GRIP
- Yes, flexible scale up
- Other,...

Question: What reason(s) do you have for this?
...

Question: What information would you like to know to determine whether this incident can lead to multidisciplinary scaling up?
...

Crisis situation 2 - additional information

Repeat:

On Sunday 28 August, at 4.26 pm, the emergency room received an automatic report of a fire allegedly raging in a nursing home in Nieuwegein. According to the procedure that applies to automatic fire alarms, the dispatcher of the control room contacts the supervisor on duty in the care home. The fire alarm comes from room 280 on the second and top floor. When an emergency response officer arrives on the floor, she encounters a closed compartment door behind which nine apartments are located. There is now a thick layer of black smoke in the hallway behind the door.

New information:

Involved:

- OvD Police with three emergency aid units
- OvD Medical with three ambulances and trauma team
- OvD Fire Brigade with three tanker guns, rescue vehicles and support units

From the telephone conversation, the operator has enough information to alert the fire brigade. The nearest tanker sprayer arrives at 4:33 PM, followed by another tanker sprayer and a rescue vehicle. In the meantime, the eviction of the residents is started by the staff and the emergency response officers. The compartment door remains closed and all other residents are brought to safety. After a few minutes, the staff is assisted by local residents who have noticed that there is an enormous amount of smoke coming from the care home. This also causes some nuisance in the area. All apartments have been evacuated by 5 p.m., with the exception of the wing where the fire is raging. In a restaurant next door, about 75 residents are waiting for what will happen next. A reporter from RTV-Utrecht, among others, reports live on the fire and social media attention has been high from the start of the incident.

In order to extinguish the fire on the second floor and to bring the residents of the relevant wing to safety, an effort is being made from both the inside and the outside. The firefighters who go in, upon arrival on the second floor, find that the corridor is filled with black smoke. They literally don't see a hand in front of their eyes. Using thermal imaging cameras, they are able to locate two people in the hallway who are clearly unconscious. Shortly afterwards, a third victim is found. It turns out residents who had opened their apartment doors and were overcome by the smoke and heat. Residents of the relevant wing who are still in their room (behind a closed and fire-resistant door) are relatively safe there, because they are shielded from the smoke. Firefighters make it clear to them that they should stay in their room until they are picked up. The firefighters take care of the three victims; they are taken to the roof of the building, where an attempt is made to resuscitate them.

Question: Do you currently want to scale up multidisciplinary with the information that is now available?

- No
- Yes, scale up to GRIP
- Yes, flexible scale up
- Other,...

Question: What reason(s) do you have for this?
...

Question: Which points of attention, in the described situation, give reason for multidisciplinary scale up from your perspective?
...

Question: From your perspective, which points of attention, in the described situation, give no reason for multidisciplinary scale up?
...

Question: Which of the incident characteristics listed below do you take into account when considering multidisciplinary scaling?

Incident:

- Incident type
- Incident size
- Expected duration incident
- Incident location
- Timestamp incident

Risks and safety:

- Possible involvement of hazardous substances/smoke
- Possible perpetrators/fleeing suspects
- Possible effects on people/material
- Safe/unsafe area
- Safe/unsafe approach route
- Applicable (safety) procedures

Meteo:

- Wind direction/ wind speed

- Temperature
- Rainfall/ clouds

Victims/population:

- (Number of) injured and injured classification
- Population specifications (for example: in a certain neighbourhood)
- Reception locations/ relatives

Environmental analysis:

- Vulnerable objects in the environment (for example: nursery/town hall)
- Care objects
- Events/demonstrations and/or other activities
- Municipal or regional border crossing incident
- Busy location/flow-through location

Communication:

- Sensitivity incident on social media
- Use NL-alert and/or other means of communication
- Action perspective communication

Services involved:

- Own disciplines involved
- Involved partners (also nationally)

Missing information:

- Cause of incident
- Duration of incident unknown
- Planning and scenarios unknown

Question: If desired, you can provide a brief explanation here and name any missing features.
...

Crisis situation 3 - report description

Background information:

On Saturday 10 and Sunday 11 June, the two-day event 'Houten Stormt' will take place, an obstacle course at the Down Under event site, close to the municipality of Utrecht and Nieuwegein. The event and the obstacles have been thoroughly discussed with all emergency services and the mayor has issued an event permit. Based on this, the event is classified as high-risk (C category) and various requirements have been set, which are met by the event organization.

Notification:

Saturday, June 10 at 12.30 pm, a report was received that during the competition for competitive runners, at the HoutenStormt event, a participant, (presumably) unconscious, disappeared underwater. This happened at the Superslide obstacle where participants are only allowed to go down one at a time. A contestant slides down the Superslide and is launched into the water. As soon as she resurfaces, she makes a swimming stroke towards the buoy, but at that moment a competitor coming from behind comes from the slide into the water and hits her on top of her head. The contestant shoots away underwater. Immediately the rescue team and the divers start a search in the murky water; the obstacle is immediately stopped. There are many spectators in the stands who are filming and are becoming more restless. Several reports are also received at the control room from spectators who report the incident.

Question: Do you currently want to scale up multidisciplinary with the information that is now available?

- No
- Yes, scale up to GRIP
- Yes, flexible scale up
- Other,...

Question: What reason(s) do you have for this?
...

Question: What information would you like to know to determine whether this incident can lead to multidisciplinary scaling up?
...

Crisis situation 3 - additional information

Repeat:

On Saturday 10 and Sunday 11 June, the two-day event 'Houten Stormt' will take place, an obstacle course at the Down Under event site, close to the municipality of Utrecht and Nieuwegein. The event and the obstacles have been thoroughly discussed with all emergency services and the mayor has issued an event permit. Based on this, the event is classified as high-risk (C category) and various requirements have been set, which are met by the event organization.

Saturday, June 10 at 12.30 pm, a report was received that during the competition for competitive runners, at the HoutenStormt event, a participant, (presumably) unconscious, disappeared underwater. This happened at the Superslide obstacle where participants are only allowed to go down one at a time. A contestant slides down the Superslide and is launched into the water. As soon as she resurfaces, she makes a swimming stroke towards the buoy, but at that moment a competitor coming from behind comes from the slide into the water and hits her on top of her head. The contestant shoots away underwater. Immediately the rescue team and the divers start a search in the murky water; the obstacle is immediately stopped. There are many spectators in the stands who are filming and are becoming more restless. Several reports are also received at the control room from spectators who report the incident.

New information:

Involved:

- OvD Police with two emergency aid units and investigations related to crime scene
- OvD Medical with ambulance and trauma team
- OvD Fire Brigade with tank car spray and diving team
- Public Prosecution Service still without units, mayor has already been informed

More divers and members of the rescue team come from the aid stations at other obstacles to help search. Meanwhile, bystanders are becoming more restless because the event has been halted and the participant does not immediately surface. Disagreements and a fight among the spectators ensues when someone is approached who is filming the accident. The atmosphere is grim. After about nine minutes, the participant is found and brought ashore, where first aid is given. The emergency services put up fences to protect the CPR from the public. Among the spectators are also relatives of the participant.

The unrest among the spectators has not improved in the meantime because there is a lot of uncertainty about the continuation of the event. The event organization advocates a modest continuation of the event and is consulting with the victim's husband.

Question: Do you currently want to scale up multidisciplinary with the information that is now available?

- No
- Yes, scale up to GRIP
- Yes, flexible scale up
- Other,...

Question: What reason(s) do you have for this?
...

Question: Which points of attention, in the described situation, give reason for multidisciplinary scale up from your perspective?
...

Question: From your perspective, which points of attention, in the described situation, give no reason for multidisciplinary scale up?
...

Question: Which of the incident characteristics listed below do you take into account when considering multidisciplinary scaling?

Incident:

- Incident type
- Incident size
- Expected duration incident
- Incident location
- Timestamp incident

Risks and safety:

- Possible involvement of hazardous substances/smoke
- Possible perpetrators/fleeing suspects
- Possible effects on people/material
- Safe/unsafe area
- Safe/unsafe approach route
- Applicable (safety) procedures

Meteo:

- Wind direction/ wind speed
- Temperature

- Rainfall/ clouds

Victims/population:

- (Number of) injured and injured classification
- Population specifications (for example: in a certain neighbourhood)
- Reception locations/ relatives

Environmental analysis:

- Vulnerable objects in the environment (for example: nursery/town hall)
- Care objects
- Events/demonstrations and/or other activities
- Municipal or regional border crossing incident
- Busy location/flow-through location

Communication:

- Sensitivity incident on social media
- Use NL-alert and/or other means of communication
- Action perspective communication

Services involved:

- Own disciplines involved
- Involved partners (also nationally)

Missing information:

- Cause of incident
- Duration of incident unknown
- Planning and scenarios unknown

Question: If desired, you can provide a brief explanation here and name any missing features.
...

General questions

The following questions are independent of the crisis situations described above. These specific questions are optional to answer. In these questions you have the opportunity to explain your general considerations in multidisciplinary scaling up.

Question: Are you considering multidisciplinary scaling up if one or more crisis partners play an important role in the incident?
...

Question: Overweegt u multidisciplinair opschalen als er behoefte is aan clustering en duiding van (veel) informatie bij het incident? Bijvoorbeeld, veel processen betrokken waardoor er een veelheid aan informatie bij verschillende diensten zit.
...

Question: Overweegt u multidisciplinair opschalen als er communicatie-uitdagingen zijn?
...

Thank you for your cooperation and time!

It is possible that new questions arise from the answers to this questionnaire. That is why I would like to ask you if you are open to any follow-up questions. If you are interested in this, you can enter your email address below. This is certainly not mandatory.

Appendix C

Performance Analysis Event log

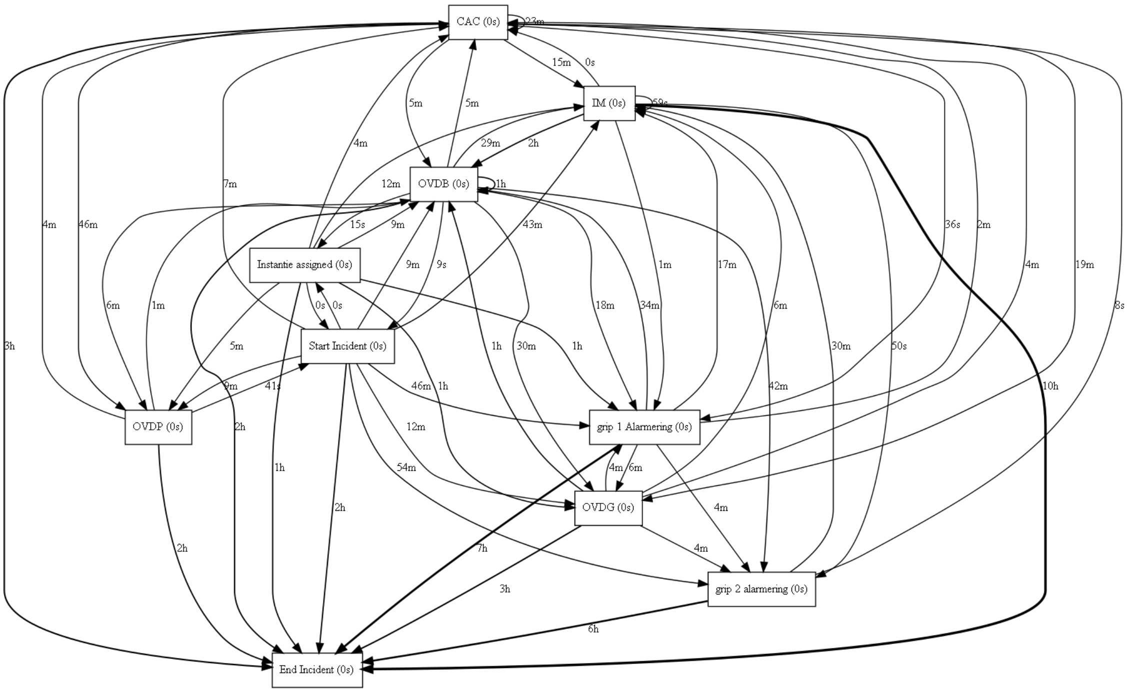
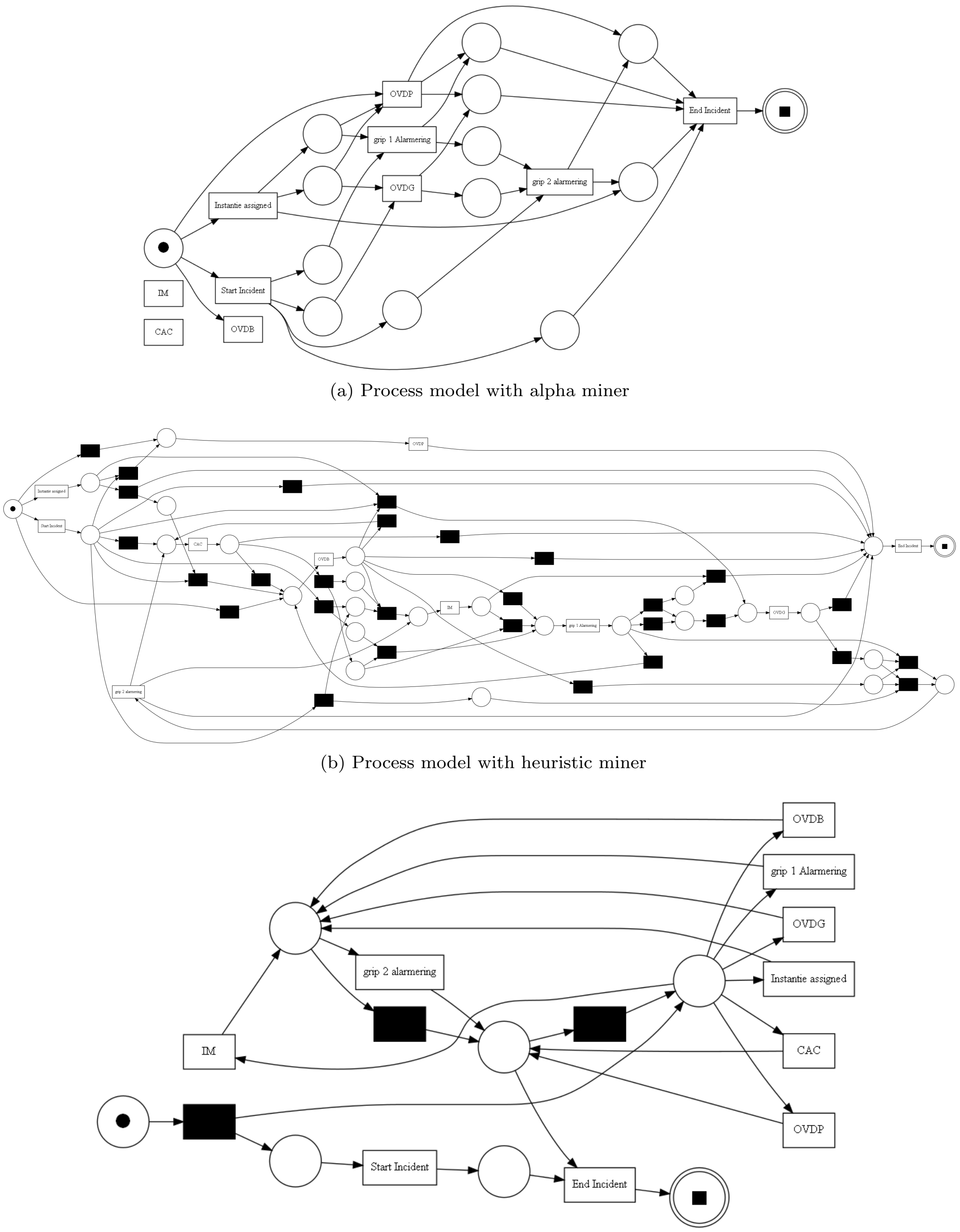


Figure C.1: Directly follow graph with performance measurement

Appendix D

Process Discovery All Subsets

D.1 Subset 1



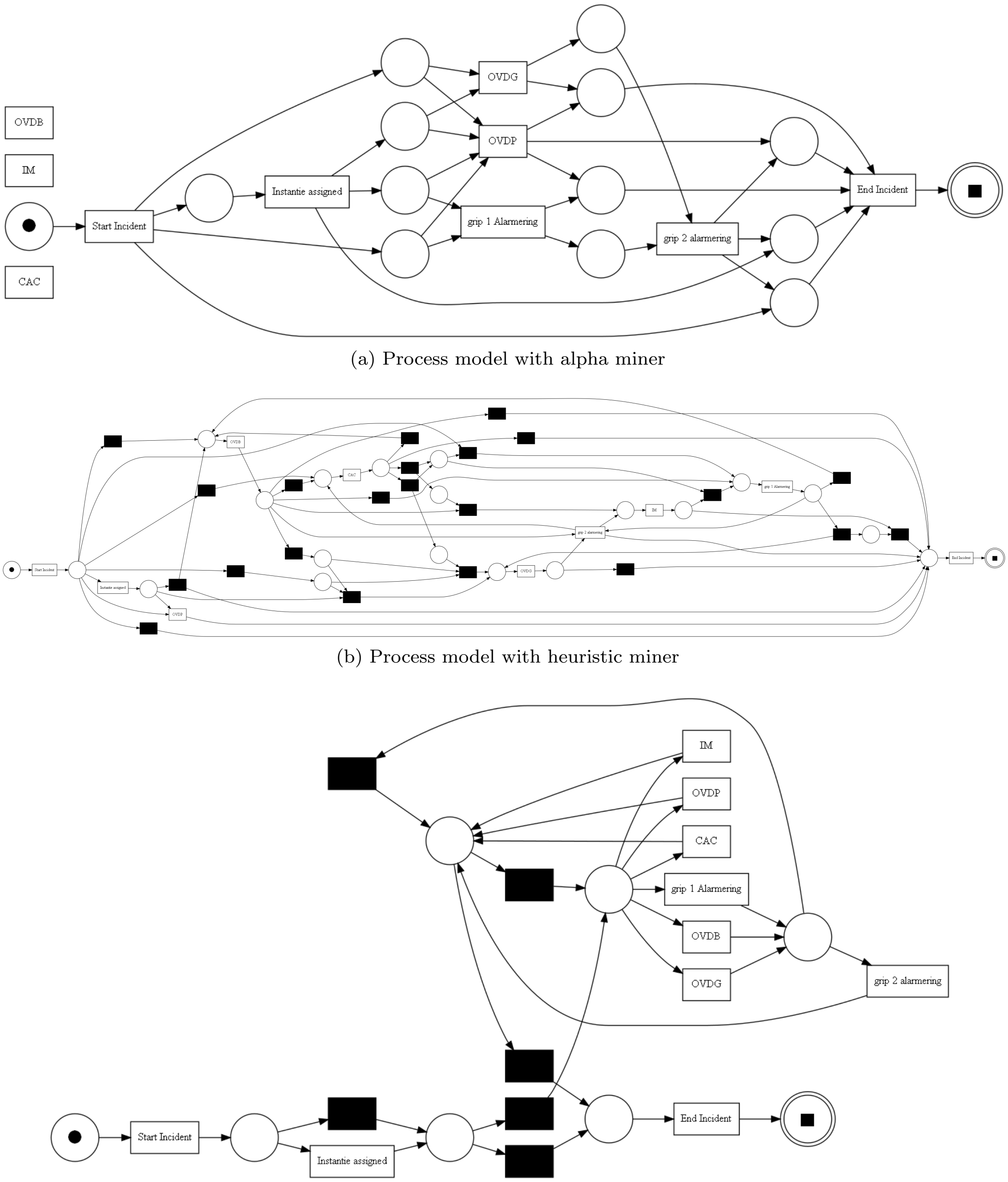
(a) Process model with alpha miner

(b) Process model with heuristic miner

(c) Process model with inductive miner

Figure D.1: Process discovery of subset 1

D.2 Subset 2



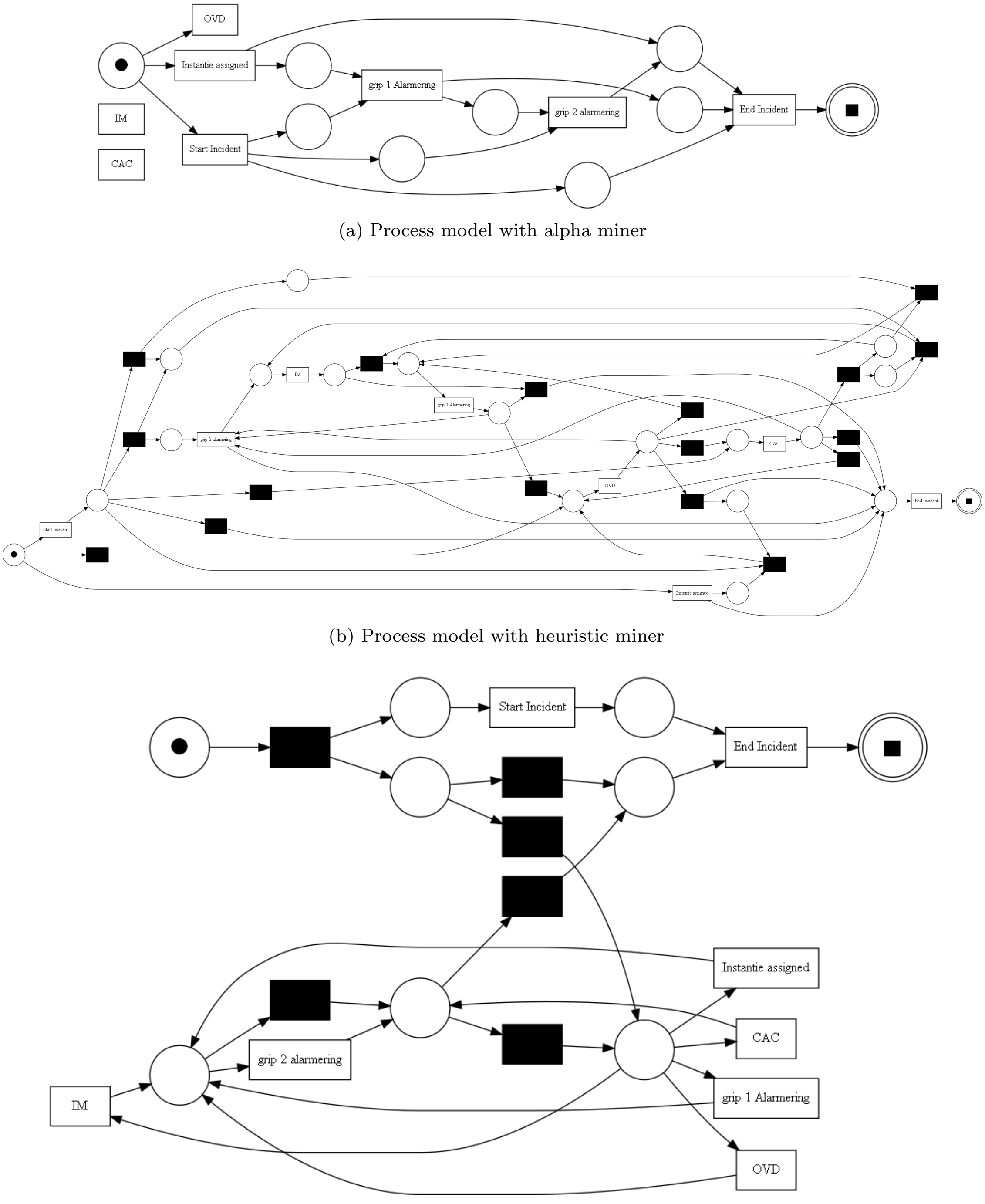
(a) Process model with alpha miner

(b) Process model with heuristic miner

(c) Process model with inductive miner

Figure D.2: Process discovery of subset 2

D.3 Subset 3



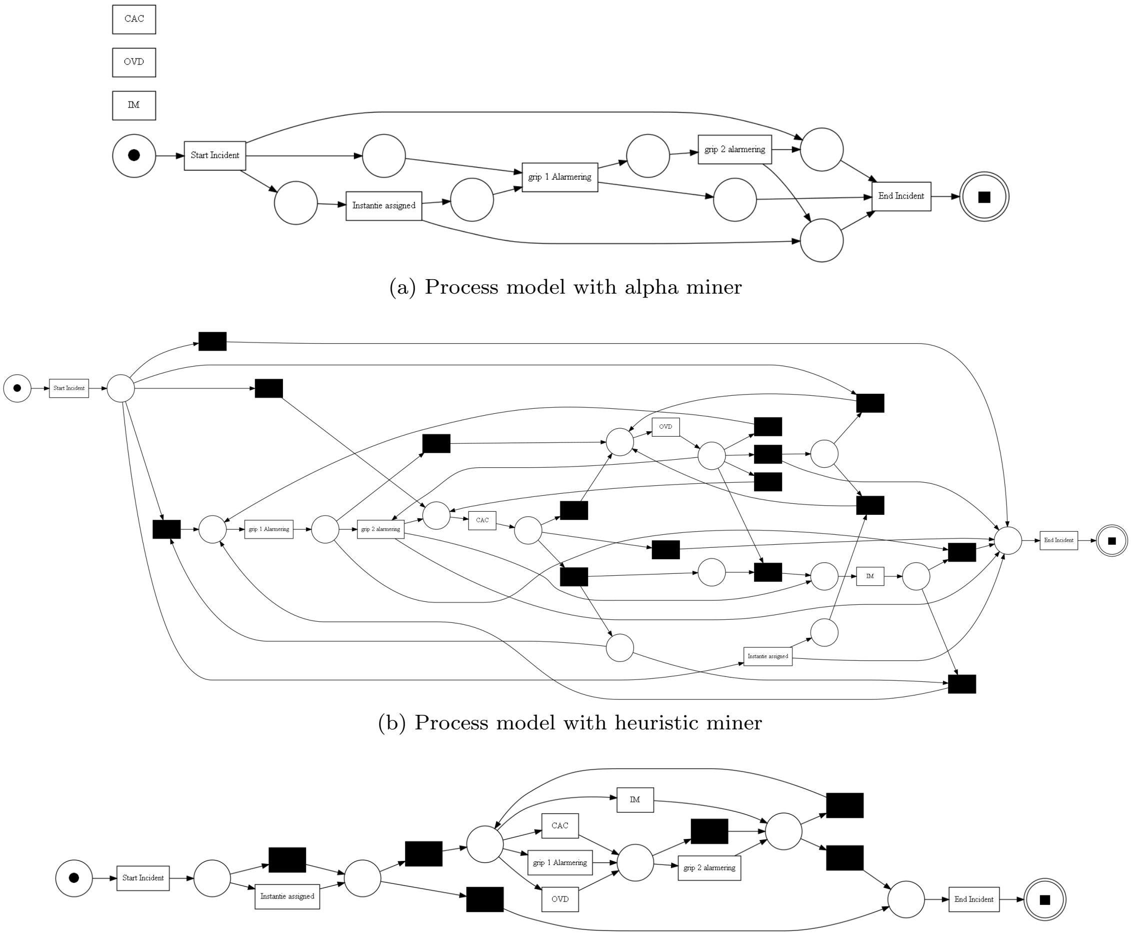
(a) Process model with alpha miner

(b) Process model with heuristic miner

(c) Process model with inductive miner

Figure D.3: Process discovery of subset 3

D.4 Subset 4



(a) Process model with alpha miner

(b) Process model with heuristic miner

(c) Process model with inductive miner

Figure D.4: Process discovery of subset 4

Appendix E

Decision Tree Visualization

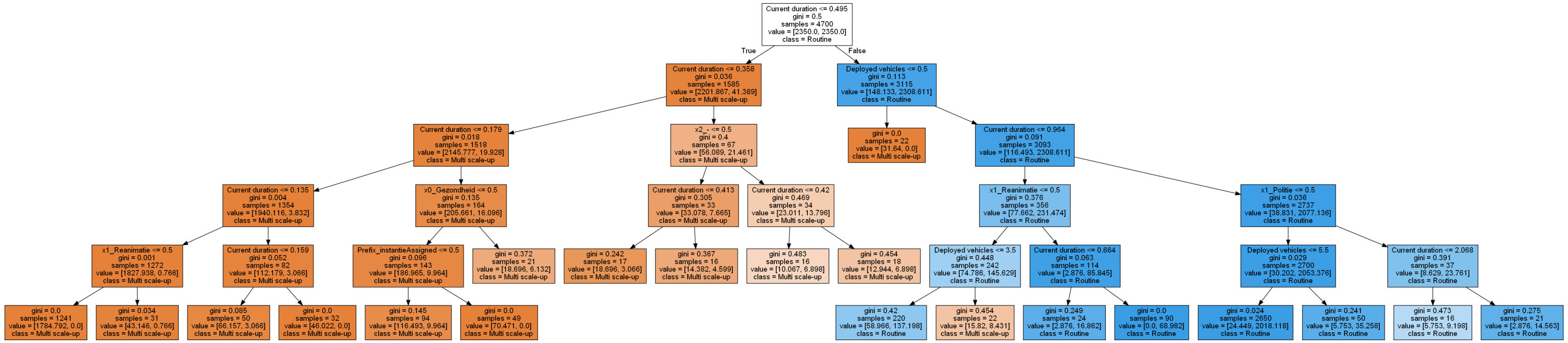


Figure E.1: Decision tree bucket 1

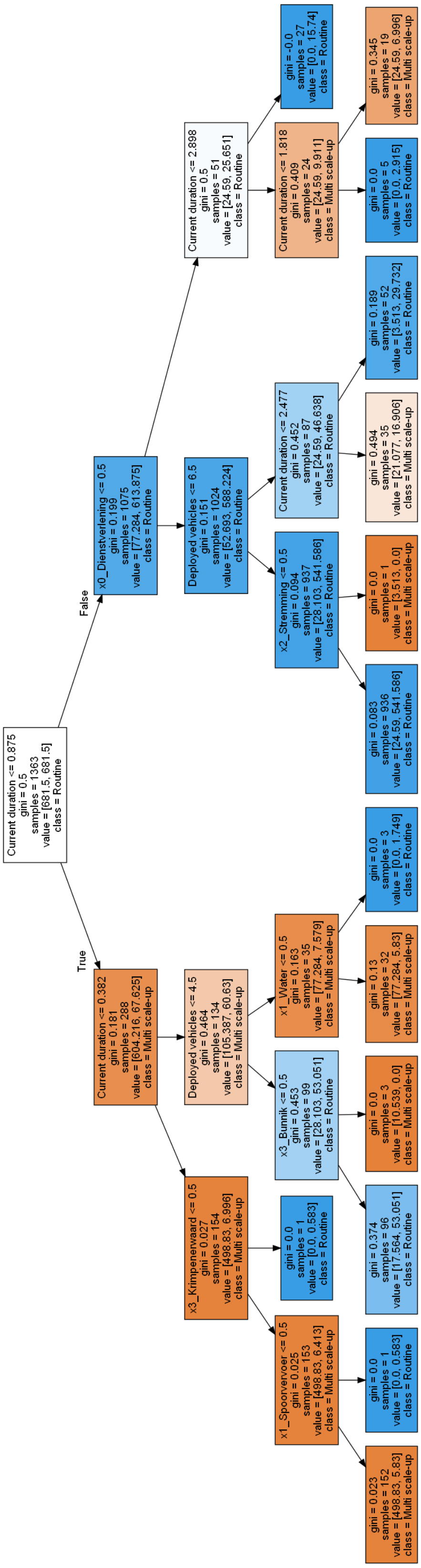


Figure E.2: Decision tree bucket 2

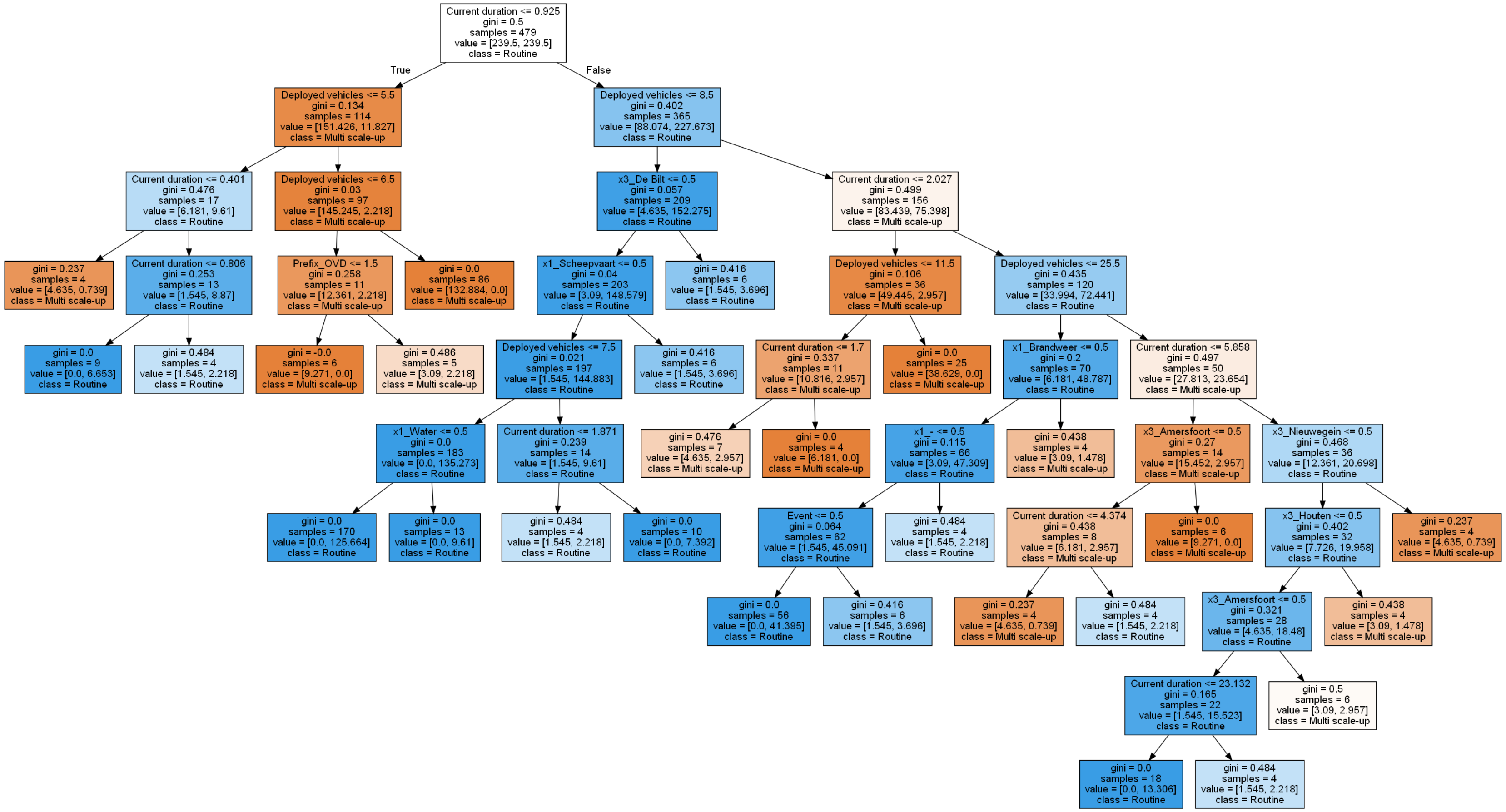


Figure E.3: Decision tree bucket 3