MASTER'S THESIS

Preferred Explanations of Algorithmic Decision-Making Systems in the Dutch Property & Casualty Insurance Industry

Schotman, E.H.H.

Award date:
2020

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal ?

Take down policy
If you believe that this document breaches copyright please contact us at:

pure-support@ou.nl

providing details and we will investigate your claim.

Downloaded from https://research.ou.nl/ on date: 09. Sep. 2021

Open Universiteit
www.ou.nl

Preferred Explanations of Algorithmic Decision-Making Systems in the Dutch Property & Casualty Insurance Industry

Education:	Open Universiteit, Faculty of Management, Science & Technology Business Process Management & IT master’s program
Degree programme:	Open University of the Netherlands, Faculty of Management, Science & Technology Business Process Management & IT master's program
Course:	IM0602 BPMIT Graduation Assignment Preparation IM9806 Business Process Management and IT Graduation Assignment
Student:	Schotman, E.H.H.
Identification number:	
Date:	September 9th, 2020
Thesis supervisor	Dr. Iren, D.
Second reader	Dr. Bromuri, S.
Version number:	0.9
Status:	Final draft

ABSTRACT

There are multiple challenges relating to the explainability of algorithmic decision-making (ADM) systems, such as that there is not a clear consensus on what a ‘good’ explanation is and what sort of explanation fit the different types of ADM-systems. Therefore, in this research, the Delphi-method was used to study which explanation would fit the different types of ADM-systems best. This was done in the Dutch property & casualty insurance industry. This report will show that often the data that is used to reach a decision should be included in the explanation because of the privacy perspective. Also, for most ADM-systems in this study, can be said that the general idea behind the algorithm and/or data should be included in the explanation. The main conclusion, however, is that there is not a one-size-fits-all explanation for ADM-systems and that it depends on the type of ADM-system, for what it is used and in which social context.

KEY TERMS

Algorithmic decision-making, explainability, insurance, Delphi-method

SUMMARY

Decisions that were historically made by humans are now made by these so-called algorithmic decision-making (ADM) systems. Despite the benefits, there are various anxieties on the secrecy, lack of transparency, and lack of technical expertise of these ADM-systems. These challenges are increasingly recognized by the public through books like ‘Weapons of Math Destruction’ wherein Cathy O’Neil multiple case studies provides on the harms and risks of ADM-systems, e.g. on insurance. A ‘right to explanation’ is seen as a promising way for accountability. There are however also multiple challenges with explainability such as that there is not a clear consensus on what a ‘good’ explanation is. Also, little study has been done to research what sort of explanation fit the different types of ADM-systems. Therefore, the objective of this research was to understand what kind of explanation would fit the different types of ADM-systems. This was studied for the Dutch property & casualty insurance industry from the perspective of the insurer.

First, a survey was carried out to determine what types of ADM-systems are used. Then for five ADM-systems different scenarios with explanation elements were submitted to a group of industry experts through the Delphi-method. The aim was to reach consensus on what combination of ADM-system and explanation types are preferred from the perspective of the insurer. This Delphi-study consisted of three rounds of questionnaires in which the participants would rank the different scenarios, for the different ADM-systems, based on preferability. Main limitations of both the survey and the Delphi-study, were the low response rate and low external validity. Therefore, the conclusions represent only the synthesis of the opinion of this group of participants and are not statistically meaningful.

On the following page, table 1 shows which explanation elements were preferred for these ADM-systems. Recurring during the study was that often the data that is used to reach a decision should be included in the explanation because of the privacy perspective. This should be done whether it is provided for a specific decision, such as the car registration number for motor insurance, or already resides in the knowledge base of the ADM-system. A second reoccurring explanation element was that, for most ADM-systems, the general idea behind the algorithm and/or data should be included in the explanation instead of providing the decision inference process information for a specific decision. The main conclusion, however, is that there is not a one-size-fits-all explanation for ADM-systems and that it depends on the type of ADM-system, for what it is used and in which social context.

Table 1. ADM-systems and their preferred explanation elements

ADM-System	**Explanation elements**	**Consensus reached**
GLM and random forest - Generalized linear models (GLM) used for pricing. To enhance these GLM's, machine learning random forest models are used in the background. These random forest models combine different decision trees to obtain an aggregated prediction/regression.	- Input parameters - General idea behind the data.	Yes
Price optimization - Machine learning churn models for price optimization purposes. Price optimization refers to the practice of adjusting the premiums, paid by different groups of consumers, based on the behaviours and economic characteristics of the consumer, in ways unrelated to their risk or cost.	- General idea behind the algorithm.	No
Recommendation engine - A recommendation engine used for the 'next best action'. This is used to evaluate the customer's past behaviour, recent actions, and needs to deliver the right message, at the right time, and via the right channel.	- User knowledge base - General idea behind the data.	No
Rule-based fraud detection - A rule-based model used to assess claims and evaluate whether they present anomalies and flag potentially fraudulent claims.	- Input parameters - Specific procedural decision information.	No
Optical character recognition (OCR) - Deep learning networks (Artificial Neural Networks) used to extract information from scanned documents such as images from damaged cars to estimate repair costs.	- Input parameters - General idea behind the algorithm.	No

CONTENTS

1. INTRODUCTION

1.1. **Background**

More and more data is captured through a variety of devices, which are then often processed by algorithms (Newell & Marabelli, 2015). Specifically, machine learning (ML) and artificial intelligence (AI) techniques are becoming more prevalent. With this increase, there is a mounting concern on the use and explainability of the underlying algorithms, data and context of the broader process (Singh, Walden, Crowcroft, & Bacon, 2016).

1.2. Exploration of the topic

Algorithmic decision-making (ADM) is used increasingly through ML and AI techniques in all sorts of industries and governments (Diakopoulos, 2016). Insurance is one of the industries where ADM-systems are used throughout the value chain. For the insurance market to function, it is essential to price risks appropriately, and for coverage to be extended to those in need. The increase of available data and analytic techniques, including ML and AI, could lead to improvements in pricing these risks appropriately (Rumson & Hallett, 2019), and in analysing the profitability of the insured on an individual customer level (Fang, Jiang, & Song, 2016). However, these are not the only application areas. According to Gartner, AI and ML techniques can be applied to many areas in the insurance industry (see figure 1), such as claim handling (Harris-Ferrante, 2017), detecting fraudulent claims (Viaene, Ayuso, Guillen, Van Gheel, & Dedene, 2007), and Usage-Based Insurance, whereby the premium is based on when and how the insured object is used (Arumugam & Bhargavi, 2019).

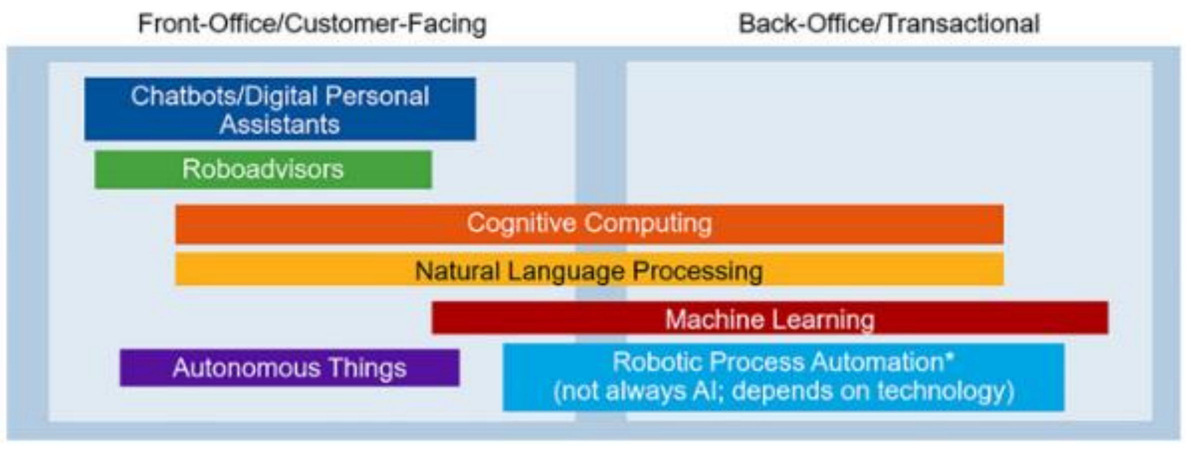

Figure 1. Application of AI in the insurance sector (Harris-Ferrante, 2017)

As these examples show, there are multiple advantages to applying ADM. When there is however a gap between the design, operation, and understanding of the algorithms used, this can have severe consequences for the society as a whole and the individuals involved (Mittelstadt, Allo, Taddeo, Wachter, & Floridi, 2016). Furthermore, the challenges grow as algorithms become more complex and interact with each other's outputs to come to a decision (Tutt, 2017). Challenges like these led to a call for algorithmic accountability: "laws governing decision-making by complex algorithms, or AI", which was recognized by the EU and led to the General Data Protection Regulation (GDPR). Within the GDPR-law there are four articles which address ADM. Article 22 addresses "automated individual decision-making, including profiling" (European Parliament, 2016), and articles 13 -15 address the transparency rights around ADM. To invoke the rights as stated in these articles it is necessary that an individual has the right to an explanation of a specific decision (Kaminski, 2019). A legally binding right to explanation, however, does not exist (Wachter, Mittelstadt, & Russell, 2018). Additionally, even if that would exist, a "meaningful explanation about the logic of processing" is unlikely to be provided (Edwards & Veale, 2017).

Finally, as the industry association (Verbond van Verzekeraars) states in their ethical framework, only when regulators and society have a sufficient level of trust in the use of data and ADM-systems, can insurers use these in their processes (2020).

1.3. Problem statement

Currently, the ADM-systems used by insurers are minimally explained to data subjects. Furthermore, it is not clear what different types of explanations exist and which ones are desirable from an insurers perspective, taking into account avoiding disclosing trade secrets, violating privacy rights of others, and data subjects gaming the insurer (Wachter et al., 2018). However, it is without a doubt that some form of explanation and accountability is required in building trust and societal acceptance of ADM.

1.4. Research objective and questions

Because of the low number of studies focussing on ADM-systems used by insurers and their explainability, this study aims to clarify which combinations of ADM-system and explanation are preferred from a property & casualty insurers perspective. Therefore, the following sub questions need to be answered:

- Which types of ADM-systems are used in the Dutch property & casualty insurance industry?
- Which combinations of ADM-system and explanation are possible?

- What combinations of ADM-system and explanation are preferred from a property & casualty insurers perspective?

1.5. Motivation/relevance

In the literature, there has been done a lot of research on how ADM-systems are used and what the impact is on society. Also, the necessity of transparency and explainable systems has been studied. For the insurance industry, the application, and challenges of ADM-systems are explored, but the relation between ADM-system and explanation, however, has not been studied. Finally, the social function and the historical application of statistical analysis in the insurance industry leads to the social and practical relevance of explaining ADM-systems.

1.6. Main lines of approach

The remainder of this thesis is structured as follows. Chapter 2 provides the theoretical framework in which the available literature is set out which leads to the objective of this research. Then, in chapter 3 the research methodology is stated. Following, the results of the research are described in chapter 4. Finally, the results are discussed and recommendations for practice and research are provided in chapter 5.

2. THEORETICAL FRAMEWORK

2.1. Research approach

The theoretical framework aims to explore the problem statement and make clear what knowledge is already available in the existing literature. Based on the research objective and sub-questions the following questions were defined on which an answer had to be found using the scientific literature.

- What are ADM-Systems?
- What types of ADM-systems are there?
- What are the pros and cons of ADM-systems?
- Are ADM-systems used in the insurance sector?
- Why is there a need for an explanation?
- What is understood by an explanation?

To find an answer to these questions the below described queries were used in the library portal of the 'Open Universiteit' (OU). Additionally, ten relevant articles were found using 'backward snowballing'.

Table 2. Research queries

Nr.	*Query*	*Keywords*	*Database*	*Relevant articles*
1.	ADM	(ADM) OR (Algorithmic decision making)	OU Library Portal	14
2.	Automated decision making	(Automated decision making)	OU Library Portal	8
3.	Types	(ADM) OR (Algorithmic decision making) OR (Automated decision making) AND (Type) OR (Category) OR (Class) OR (Sort)	OU Library Portal	5

4.	Explainability	(ADM) OR (Algorithmic decision making) OR (Automated decision making) AND (Explanation) OR (Interpretable) OR (Transparent)	OU Library Portal	6
5.	Insurance	(Big Data) OR (Data) OR (AI) OR (Machine Learning) OR (ADM) OR (Algorithmic decision making) OR (Automated decision making) OR (Explanation) OR (Interpretable) OR (Transparent) AND (Insurance) OR (Insurer)	OU Library Portal	2

2.2. Implementation

For every query, the first 150 results were assessed on relevance. Hereby,

- The first distinction was based on the 'Title' and the first few lines of the abstract.
- Following, the remainder of the abstract, introduction, conclusion, and when relevant other parts of the articles as well, were studied. This was done by taking notes, highlighting specific sections and for the most relevant articles synthesizing this in an annotated bibliography.

An article was considered relevant when it was related to either ADM-systems or different types of explanations.

2.3. Results and conclusions

Much of the decisions which were historically made by humans are now made by algorithms (Kroll et al., 2016). Decisions on prioritization, classification, association, and filtering are made by these so-called ADM-systems (Diakopoulos, 2016). From an economic perspective, this development seems desirable (Waltl & Vogl, 2018). There are however also various anxieties around the use and impact of these ADM-systems, namely the secrecy, lack of transparency, and lack of technical expertise of these systems (Selbst & Barocas, 2018). Another problem is the bias in these systems, both through discriminating historical data but also because these systems are created by a relatively homogenous group of people (Allen, 2019). These challenges show the tension between the interests of individuals (Newell & Marabelli, 2015) and 'ad hoc' created groups (Mittelstadt, 2017) on the one hand, and businesses and governments on the other. Hereby the individuals and 'ad hoc' created groups are willing to give up their privacy, freedom, and independence for new opportunities, and businesses and governments are keen on exploiting these new opportunities, but sometimes with costs to some individuals and 'ad hoc' created groups (Newell & Marabelli, 2015). These challenges are increasingly recognized by the public through books like 'Weapons of Math Destruction'. Herein Cathy O'Neil (2016) provides multiple case studies, e.g. on insurance, on the harms and risks of ADM-systems. In the insurance industry, ADM-systems are used throughout the value chain (Harris-Ferrante, 2017). A thematical review by EIOPA on the use of big data analytics in motor and health insurance shows that these tools are mostly used within the following parts of the value chain: pricing and underwriting (35%), claims management including fraud prevention (30%) and sales and distribution (24%). Historically, statistical analyses on traditional data sources like demographic data lie at the core of insurance. More currently, these traditional data sources are combined with newer sources like telematics and online media, through ML techniques. It should be noted however that these ML techniques, even the more advanced ones, may not be more complex in terms of explainability than the 'traditional' generalised linear models (EIOPA, 2019).

To reveal how these systems operate, Waltl & Vogl (2018) distinguish three different dimensions which are part of every ADM-system, including AI and ML systems. Every ADM-system can be viewed from a process, model, and classification dimension, which all call for a different level of transparency. Hereby, the model level consists of the decision structures that are used to come to a decision which can vary in interpretability by humans. The authors distinguish three types of models: the deductive and rule-based systems, statistical probabilistic models, and artificial neural networks. Guidotti et al. (2018) separate models more detailed and argue that the more interpretable ones are: decision tree, decision rule, and linear model. Examples of less interpretable models or black boxes are Neural Network, Tree Ensemble, Support Vector Machine, Deep Neural Network, and/or Non-linear models. In understanding how these black boxes work they introduced a taxonomy, wherein they make a separation in explaining the black box and designing a transparent box. When these models are ML-algorithms they can learn in the following three ways: supervised, unsupervised, or reinforced (Karanasiou & Pinotsis, 2017).

Kroll et al. (2016) challenge the position that transparency will solve the problems with ADM-systems because disclosure of the model is not necessary nor sufficient and may even be undesirable. Some of the challenges with transparency are the loss of privacy of others; perverse effects like 'gaming the system'; loss of the competitive advantage (Zarsky, 2016); changing systems over time (e.g. machine learning algorithms (Lepri, Oliver, Letouzé, Pentland, & Vinck, 2018); unclarity onto whom the ADM-system should be transparent (Kemper & Kolkman, 2018); and whether it is fair to impose a higher standard of transparency on ADM then on human decision making (Zerilli, Knott, Maclaurin, & Gavaghan, 2018). To address these challenges Kroll et al. (2016) suggest that ADM-systems should be designed to comply with legal and policy objectives. Selbst & Barocas (2018) add to this that it needs to be revealed what value judgements were made in the design of these systems.

A right to explanation is seen as a promising way for accountability in algorithms. Some argue that there is a right to an *ex-post* explanation of specific decisions within the GDPR regulation (Wachter, Mittelstadt, & Floridi, 2017). Hereby they combine the non-binding Recital 71 with binding articles 13, 14, and 22 to make the argument that "The law will […] effectively create a "right to explanation," whereby a user can ask for an explanation of an algorithmic decision that was made about them" (Goodman & Flaxman, 2017). Wachter et al. (2017) however argue that there is no meaningful right to an explanation and only a "right to be informed" because it is restricted to an explanation of system functionality.

Whether it is legally binding or not, the increase in the number of published research papers on explainability indicates the importance and relevance of this topic (Nunes & Jannach, 2017). There are however multiple challenges with explainability: there is not a clear consensus on what a 'good' explanation is (Nunes, Miles, Luck, & De Lucena, 2012); the type of explanation could affect the decision-making process (Tintarev & Masthoff, 2011) and it is hard to come to a user-tailored explanation without any domain-specific knowledge (Zanker & Ninaus, 2010). To address these challenges Nunes & Jannach (2017) performed a systematic literature review which resulted in a taxonomy of the different aspects to be considered when determining an explanation.

As described, there has been done a lot of research on what ADM-systems are and what the effects are on society. Also, it has been studied how these models could be made more transparent both at the input stage (ex-ante) as at the output stage (ex-post). Furthermore, what is understood by an explanation has been explored. However, what sort of explanation fits the ADM-systems best has not been given that much attention.

2.4. Objective of the follow-up research

The objective of the follow-up research is to understand what kind of explanation would fit the different types of ADM-systems, used in the Dutch property & casualty insurance industry, best from an insurers perspective. To achieve this understanding, it is necessary to study which types of ADM-systems are used, and then which combinations of ADM-system and explanation are preferred.

3. METHODOLOGY

The objective of the research to be conducted is to clarify which combinations of explanation and ADM-system are preferred from a Dutch property & casualty insurers perspective. This chapter describes the research strategy on 'what', 'why', and 'how' this objective will be achieved. The basis for the methodology was the book 'Research Methods for Business Students' written by Saunders, Lewis, and Thornnhill (2016).

3.1. Conceptual design: Select the research method(s)

To achieve the objective of this research there are three questions on which an answer had to be found:

- Which types of ADM-systems are used in the Dutch property & casualty insurance industry?
- Which combinations of ADM-system and explanation are possible?
- What combinations of ADM-system and explanation are preferred from a property & casualty insurers perspective?

To answer the first question, information from the Dutch insurance industry must be gathered regarding which ADM-systems are used, in which processes and for what purpose. Because of the descriptive character of this research question, a questionnaire will be used to gather this information. Some advantages are that it enables comparison and is easy to explain. The disadvantages are that it is not as wide-ranging as other methods and the quality depends on the quality of the questionnaire itself. The design, piloting, and response rate, therefore, deserves extra consideration.

The gathered information will then be combined with the user interface components out of the taxonomy of explanations (see Appendix I), from Nunes and Jannach (2016), to draw up scenarios. This will be done for each combination of ADM-system and explanation type.

To answer the third question insight must be given in which of the drawn-up scenarios are preferred by the industry. Because of the exploratory character of the research question, the scenarios will be submitted to a group of industry experts through the Delphi-method. This is a method "for structuring a group communication process so that the process is effective in allowing a group of individuals, as a whole, to deal witch a complex problem" (1975), This has as advantage that it increases the construct validity, and is well suited for exploratory questions. It has, however, the same design issues as a questionnaire (Okoli & Pawlowski, 2004). Also, it is a time-consuming process which requires proper planning and management (Hsu & Sandford, 2007).

3.2. Technical design: Elaboration of the method

In this section, it is described what data is required, which sources can provide these data, what requirements these sources must meet and how this data will be gathered. This is described per question, apart from the final question which will be described in two separate sub-paragraphs.

3.2.1. Types of ADM-systems used

To address the first question demographic and factual data is required which can be provided by the industry. To ensure that respondents have sufficient knowledge of ADM and reduce the chance of uninformed responses, the questionnaire will be held under policymakers, decisionmakers, privacy-experts, data analysts, and/or actuaries. The demographic data is on education, company, and occupation, this is needed to check if the data collected are representative of the total population. The factual data is on what types of ADM-systems are used in which processes.

The design of the questionnaire will differ according to how it will be delivered, returned, or collected, and the amount of contact the researcher will have with the respondents. Because of the geographical spread of the respondents, the ease and low-cost character, and to reduce the chance of socially desirable responses a self-completed internet questionnaire will be used. A downside of this type is that the response rate is normally lower than an interviewer-completed questionnaire. Other points of attention are the length and complexity of the questionnaire. The questionnaire will include a combination of open and closed questions. The closed questions are used to gather demographic data. To gather the factual data a combination of both open and closed will be used.

The options for the types of ADM-system in the closed questions will be based on Waltl & Vogl (2018):

- Rule-based systems.
- Statistical probabilistic models.
- Artificial neural networks.

Also, there will be made a distinction between learner and traditional ADM-systems. Finally, there will be the possibility to give an example of the model used in the form of an open question. The choice for three types of ADM-systems, instead of a more comprehensive list, is made to minimalize the length and complexity of the questionnaire.

The choice of the process in the closed questions will be based on the process types provided in the 'Industry survey-Big Data thematic review' of EIOPA (2018):

- Product development.
- Pricing and underwriting.
- Sales and distribution.
- Post-sales services and assistance.
- Fraud and claims management.

The questionnaire will be held using the online survey tool 'SoGoSurvey' which enables to embed the questionnaire into an email and a free data export functionality. At the start of the questionnaire, it will be explained clearly and concisely why the researcher would like the respondent to complete the questionnaire. Before using the questionnaire, it will be pilot tested to refine the questionnaire so that respondents will have no problems answering and there will be no problems in recording the data. The suggestion of Bell and Waters (2014) will be followed to find out:

- How long the questionnaire took to complete.
- The clarity of instructions.
- Which, if any, questions were unclear or ambiguous.
- Which, if any, questions the respondent felt uneasy about answering.
- Whether in their opinion there were any major topic omissions.
- Whether the layout was clear and attractive.
- Any other comments.

3.2.2. Combinations of ADM-system and explanation

To draw-up the different scenarios the following data is required:

- Which types of ADM-systems are used.
- Which sorts of explanations can be defined.

The types of used ADM-systems are a result of the before described questionnaire and will therefore not be discussed in any further detail in this paragraph. The sorts of explanations can be considered factual data.

To keep the list of scenarios comprehensible it is required that the sorts of explanations be clear. Therefore, only the ‘User Interface Components’ of the taxonomy will be used. This consist of four ‘content’ related types of information and three ‘presentation’ facets.

- Content: Input parameters, Knowledge base, Decision inference process, and Decision output.
- Presentation: Baselines, Formats, and Perspective.

These explanation ‘content’ types and ‘presentation’ facets will be combined with the types of ADM-systems used to draw up scenarios for each combination. Criteria for each scenario are:

- ADM-system and process must be based on the results of the closed and open questions in the questionnaire.
- Should contain a ‘content’ and ‘presentation’ element for the explanation.
- Should be no longer than a few sentences.

3.2.3. Preferred scenarios: Delphi inquiry design

To set up the Delphi-study the ‘toolkit’ of Day and Bobeva (2005) is followed. Hereby the key stages are Exploration, Distillation, and Utilisation. In the exploration stage, the study is planned, participants are selected, and a pilot is conducted. In the Distillation stage, the different iterative rounds are held. Finally, in the Utilisation stage, the results are analysed The design choices are summarized in table 3 and are further explained in the remainder of this paragraph.

Table 3. Delphi inquiry design (Day & Bobeva, 2005)

Criteria	***Choice***
Purpose of the study	Exploration
Number of rounds	Three
Participants	Heterogeneous group
Mode of operation	Remote
Anonymity of panel	Full
Communication media	Computerized
Concurrency of rounds	sequential

The purpose of this study is to find out what combinations of ADM-system and explanation are preferred by the insurance industry, which is exploratory in character.

When choosing the number of rounds, it is considered that the higher the number of rounds is, the slower the observed convergence is. Also with more than two rounds, there is the risk that experts will abandon the study or will shift their evaluations towards the mean position (Gallego & Bueno, 2014). Therefore, the chosen number of rounds is three.

The criteria for participants are that they have sufficient knowledge of ADM, are willing to commit to the multiple rounds of the study and are willing to revise their judgements to reach consensus (Hsu & Sandford, 2007). Between fifteen and 35 people are common in a Delphi-panel (Gordon, 1994), but it should be noted that the dropout rate could be high, so the initial sample should be on the higher end of the range (Day & Bobeva, 2005). For this study, the panel participants could be experts working:

- within the insurance industry.
- at regulators or supervisors.
- for the 'Verbond van Verzekeraars' (industry association).
- within the financial services consultancy.

The panellist should ideally be heterogeneous in terms of nationality, occupation/role, and age. Of importance is the expertise and knowledge level of the participants. The panel members can be policymakers, decisionmakers, privacy-experts, data analysts, and actuaries.

Computerized remote access is chosen because of the geographical spread of the respondents and the time independency. Also, this will ensure full anonymity.

Finally, a sequential concurrency of rounds is chosen because it is easier in design and enables analysis in between the rounds.

3.2.4. Preferred scenarios: Survey design

For the survey design, the paper of Hsu & Sandford on the Delphi technique is used (2007). In between the different rounds, the results will be analyzed, and the areas of agreement and disagreement will be identified. The respondents will be given two weeks in between each round to respond (Delbecq, Van de Ven, & Gustafson, 1975).

The goal of the first questionnaire is to initiate the process of coming to a consensus on what the most preferred explanations are for the different types of ADM-systems. It also serves as a starting point for their thoughts. Hereby, for every ADM-system, multiple different scenarios are presented which each represent the different sorts of explanations. The respondents are then asked to rank the different scenarios based on preferability. Also, for each ADM-system, there will be an open question to substantiate the chosen order.

In the second round, each respondent will receive a questionnaire that includes the rankings and is asked to revise his/her judgments or to specify why they remain outside of the consensus. In the third and final round the remaining items, their rankings, and minority opinions are shared with the respondents. This will provide a final opportunity to revise their judgements.

3.3. Data analysis

In this section, it is described how the collected data will be analysed. This is described per question.

3.3.1. Types of ADM-systems used

The closed questions in the questionnaire will result in categorical descriptive (nominal) and ranked (nominal) data which will require quantitative data analysis. To enable analysis the first step will be to code the categorical data and to check for any errors. Then the results will be explored using data presentation methods like a table, pie, and/or bar chart. Based on the results, the types of ADM-systems will be linked with the different processes.

3.3.2. Combinations of ADM-system and explanation

The open questions of the questionnaire, wherein examples are asked for the models used will also be analysed but requires qualitative analysis. Hereby the given answers will be coded to identify what are the more commonly used models, which will form the basis of the scenarios.

3.3.3. Preferred scenarios

The different rounds will result in ranked (ordinal) data and will be analyzed to measure the degree of consensus. The two most widely used correlation analysis methods are the ‘Spearman’s rank correlation coefficient’ and the ‘Kendall’s rank correlation coefficient’. Because the data contains tied ranks, Kendall’s rank correlation is considered to be more appropriate (Saunders et al., 2016). Consensus is reached when the coefficient is 0.7 or higher because Schmidt (1997) considers this a high level of agreement for Delphi studies (a Kendall coefficient of 1.0 means that there is full agreement on the ranking).

The open questions, in which the ranking is substantiated, will be analyzed by coding the given answers.

3.4. Reflection w.r.t. validity, reliability, and ethical aspects

In this section, it is described why the chosen methodology will allow for valid and reliable results. This is described per question. Furthermore, it is described why the research is sound from an ethical perspective.

3.4.1. Types of ADM-systems used

The internal validity will be increased by defining the types of ADM-systems and processes in the insurance value chain based on the literature review. Also, this will be strengthened by pilot-testing the questionnaire to assess whether the questions are essential. The external validity will be low because of the relatively low number of respondents and all work in different companies and therefore in a differing context. To strengthen the reliability a rationale will be provided on how the questions were determined and the audit trail, in the data analysis stage, will be logged.

The main issue is that both validity and reliability could be strengthened further for example by using:

- statistical analysis to increase criterion-related validity.
- different scales to measure the same constructs to increase the construct validity.
- test re-test, calculation of internal consistency, and check questions to increase reliability.

However, to be able to use these methods the group of respondents must be higher, and the questionnaire extended. Because the main goal of this part of the research is to enable drawing up the scenarios and perform the Delphi-study, these limitations are accepted.

3.4.2. Combinations of ADM-system and explanation

The main challenge in drawing up the different scenarios regarding internal validity is that they should be concise but complete and recognizable. The conciseness of the scenarios has an impact on the internal validity because when these are too long, the respondents could lose motivation which negatively impacts the output. To address this issue the scenarios should be no longer than a few sentences. When the scenarios do not contain all relevant items on the type of ADM-system and sort of explanation, the Delphi-study cannot answer the research questions. To address this issue the scenarios will be based on the literature review. Finally, when the scenarios are not recognizable the respondents will not be able to rank them during the Delphi-study, therefore the scenarios will be based on the closed and open questions of the questionnaire.

To strengthen the reliability a rationale will be provided on how the scenarios were determined and why these scenarios were chosen.

3.4.3. Preferred scenarios

The main shortcomings of the Delphi-study which have an impact on the internal validity, external validity and reliability are (Hsu & Sandford, 2007):

- Potential low response rates and high time consumption
  To address the first two issues, special attention must go to the motivation of the respondents in which the investigator must play an active role and maintain a high level of communication. Also, the time between the rounds should be minimized to mitigate the risk that the respondents' circumstances, knowledge, and situational context changes to much (Day & Bobeva, 2005).
- Potential of moulding opinions
  To address this issue the questionnaire will be pilot tested so that the formulated questions are clear, concise, and unambiguous. Part of the pilot-test will be if the structure of the questions implies an answer (Day & Bobeva, 2005).
- Unevenly distributed expertise under respondents.
  To address this issue, it is important that the respondents are encouraged to provide substantiation. An additional challenge hereby is the documentation of the results. Therefore, the recording of different substantiations should be done at a similar level of detail.

Also, to strengthen the internal validity the results of the round must be examined on plausibility and consistency. The Delphi-method allows this through the different rounds of feedback and confirmation of the respondents. Thereby, the scenarios in the questionnaire are based on the literature review which increases the internal validity. With regard to the external validity, Gordon (1994) explained why this is non-applicable to a Delphi study: "Because the number of respondents is usually small, Delphi's do not (are not intended to) produce statistically meaningful results; in other words, the results by any panel predict the response of a larger population or even a different Delphi panel. They represent the synthesis of the opinion of the particular group, no more, or less."

Finally, to increase the reliability a rationale will be provided on how the questions were determined and the audit trail, in de data analysis stage, will be logged.

3.4.4. Ethical aspects

There are multiple ethical issues which are of importance throughout the research and require ethical integrity from the researcher. One of the important stages of the research wherein ethical issues can arise is when access is sought. Hereby, it is essential that no pressure is applied on intended participants and that refusal is accepted as part of the research. Another issue is that consent should be informed. To ensure this, those involved in the research must be provided with sufficient information. The relevant information for both the questionnaire and the Delphi-study will be made available online. This includes information on the nature of the research, what the requirements are of being part of the research, what the rights are of those taking part, and how the collected data will be used.

During the data collection stage, it is important that respondents still have the possibility to withdraw. Also, the data will be collected accurately and fully so that subjective selectivity is avoided. This relates to the validity and reliability of the research design which is described in the previous paragraphs. Finally, confidentiality and anonymity are an issue, especially because of the digital channels of communication and data collection. Therefore, the collection of personal data will be minimized.

The issues related to confidentiality and anonymity also apply to the analysis and reporting stage. Hereby the anonymity of individuals will be maintained, and the reported results should not be traceable to an individual.

4. RESULTS

To achieve the objective of this research there were three questions to be answered, namely:

- Which types of ADM-systems are used in the Dutch property & casualty insurance industry?
- Which combinations of ADM-system and explanation are possible?
- What combinations of ADM-system and explanation are preferred from a property & casualty insurers perspective?

Each question will be addressed in a separate paragraph.

4.1. Types of ADM-systems used

A self-completed internet questionnaire was used to answer the question of which types of ADM-systems are used in the Dutch property & casualty insurance industry. This was done to collect demographic and factual data. Before the questionnaire was held it was pilot tested and refined based on the results. The responses were collected through the online survey tool ‘SoGoSurvey’. The questionnaire, including a knowledge assessment, was based on the literature found during the theoretical framework and is included in Appendix VIII.

In addition to the proposed technical design, a knowledge assessment was included, and the participants were asked to assess their knowledge on a scale of one to ten. This decision was taken mainly to train the participants but also to ensure sufficient knowledge of ADM and reduce the chance of uninformed responses. The assessment was based on literature from Barredo Arrieta et al. (2020), Grosan & Abraham (2011), and Bolander (2019). There were twelve questions within the knowledge assessment and each question was weighted equally. The results ranged from zero to ten points and when a participant scored six or more points, it was included in the results. The first three questions of the knowledge assessment asked the participant to select which option described a rule-based model, a statistical probabilistic model, or an artificial neural network. Thereafter, nine statements were provided (three for each ADM-system type) for which the participants had to indicate whether the statement was true or false.

Concerning the response rate, the inherent character of a self-completed internet questionnaire, in combination with a small target group, and dependency on the personal network of the researcher, has led to a low response rate (N=13). Of the thirteen participants, eight were included in the results based on the knowledge assessment. Of these eight participants:

- Six participants worked for an insurer with a gross written premium less than 500 million, and two for an insurer with a gross written premium of more than 500 million.
- Three participants were actuaries, three worked in management positions, one in compliance, and one in marketing intelligence.

The results in figure 2 show that ADM-systems are mainly used within the processes of ‘Product development’ and ‘Pricing and underwriting’. Also, the results show that no artificial neural networks are used and that there is no difference between the use of rule-based or statistical probabilistic ADM-systems.

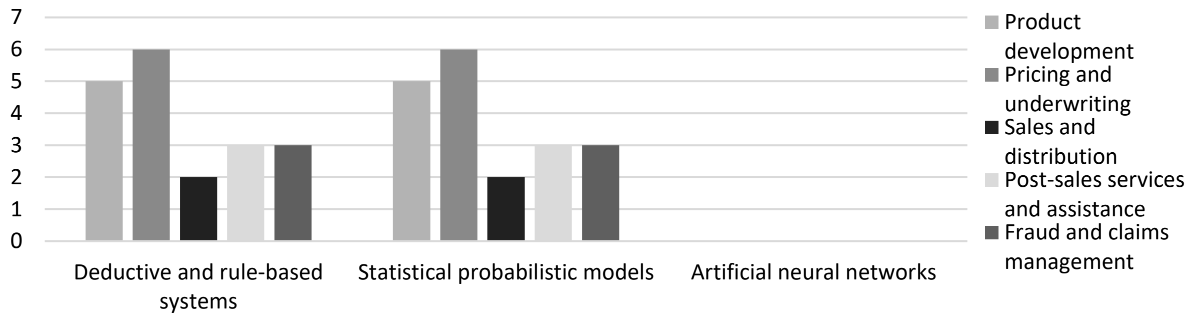

Figure 2. ADM-Systems within P&C Insurers

Finally, it should be noted that:

- The answers to the open questions show that the participants see ‘Product development’ and ‘Pricing and underwriting’ as synonyms for the same process instead of segregated parts of the value chain.
- The described results are not statistically significant and can only be used to better understand the status quo.
- For the following processes the option ‘unknown’ was chosen relatively often: ‘Sales and distribution’ (three times), ‘Post-sales services and assistance’ (four times), and ‘Fraud and claims management’ (four times). For ‘Product development’ and ‘Pricing and underwriting’ the option ‘unknown’ was chosen one time.

Appendix II and III include the full results to respectively the knowledge assessment and the questionnaire, and will not be further discussed.

4.2. Combinations of ADM-system and explanation

To answer the question of which combinations of ADM-system and explanation are possible, the methodology required that the results of the questionnaire would be combined with the user interface components out of the taxonomy of explanations (see Appendix I), to draw up scenarios. This would be done for each combination of ADM-system and explanation. The possible combinations were however too extensive. Hence, this was not an acceptable option, since the conciseness of the number of scenarios impacts internal validity. Another option was that the researcher would choose which elements of an explanation would be included in the different scenarios. This, however, was considered arbitrary. Therefore, it was decided that only the type of ADM-systems would be determined, and the sorts of explanations would be included in the Delphi-study (see paragraph 4.3 for the results to the Delphi-study). Finally, in addition to the results of the questionnaire, literature was used to draw up the scenarios. This deviation from the technical design was chosen because of the limited response to the open questions of the questionnaire. Based on the literature, three ADM-systems were enriched, and two additional ADM-systems were drafted. This resulted in five ADM-systems which would form the basis for the Delphi-study.

4.2.1. Survey-based ADM-systems

GLM and random forest - Generalized linear models (GLM) used for pricing. To enhance these GLM's, machine learning random forest models are used in the background. These random forest models combine different decision trees to obtain an aggregated prediction/regression (Barredo Arrieta et al., 2020).

Recommendation engine - A recommendation engine used for the 'next best action'. This is "used to evaluate the customer's past behaviour, recent actions, and needs to deliver the right message, at the right time, and via the right channel" (EIOPA, 2019).

Rule-based fraud detection - A rule-based model used to assess claims and evaluate whether they present anomalies and flag potentially fraudulent claims (EIOPA, 2019).

4.2.2. Additional ADM-systems

Price optimization - Machine learning churn models for price optimization purposes. Price optimization refers to the practice of adjusting the premiums, paid by different groups of consumers, based on the behaviours and economic characteristics of the consumer, in ways unrelated to their risk or cost (EIOPA, 2019).

Optical character recognition (OCR) – "Deep learning networks (Artificial Neural Networks) used to extract information from scanned documents such as images from damaged cars to estimate repair costs" (EIOPA, 2019).

4.3. Preferred scenarios

To answer the question on what combinations of ADM-system and explanation are preferred from a property & casualty insurers perspective, a Delphi-study was performed. Before the start of the Delphi-study, each round was pilot tested and refined based on the results. The responses were collected through the online survey tool 'SoGoSurvey'. For each round, the results were downloaded, analyzed, and areas of agreement and disagreement were identified. This was used as a basis for the subsequent round. To enable the participants to make a well-founded selection, theory on explainability was provided, which is included in Appendix IV (Nunes & Jannach, 2017).

In the first round, the participants selected the 'content' and 'presentation' elements which should be included in the explanation. Based on the results of round one, scenarios were then drawn up for every ADM-system. There were nine respondents to this round of which, two were data analysts, four were actuaries, one was a policymaker, one was product expert, and one was privacy expert. Of them, eight worked for an insurer and one for the industry association.

Following, in the second round, the participants ranked the drawn-up scenarios based on preferability from the perspective of the insurer. To this round, there were nine respondents of which, two were data analysts, four were actuaries, one was a policymaker, one was product expert, and one was privacy expert. Of them, seven worked for an insurer, one for the industry association, and one in the financial services consultancy. However, only seven responses were recorded, so two of the nine responses were lost. On this will be reflected in chapter 5.

Finally, in the third round, the rankings and minority opinions of round two were presented for the ADM-systems on which no consensus was reached. Then the participants were asked whether they agreed with the presented ranking or substantiate why they did not. There were nine respondents to this round of which, two were data analysts, five were actuaries, one was a policymaker, and one was product expert. Of them, seven worked for an insurer, one for the industry association, and one in the financial services consultancy. However, only eight responses were recorded, so one of the nine responses was lost. On this will be reflected in chapter 5.

The main results are described in the remainder of this chapter. The full results to the first, second, and third-round can be respectively found in Appendix V, VI, and VII. The questionnaires for the three rounds are included in Appendix IX, X, and XI.

4.3.1. GLM and random forest

After the first round, four 'content' elements were combined to draw-up six scenarios. Concerning the 'presentation' elements, a 'natural-language format', and 'positive perspective' apply to every scenario. Whether or not a 'baseline' should be included was still a point of discussion and was therefore included in the second round. During the second round, five out of seven participants gave as a response that a 'baseline' (single or group of alternatives) should not be included for comparison. Finally, consensus was reached in the second round and was therefore not included in the third round. Hereby, scenario b was ranked first (based on the mean ranking) and six out of seven participants ranked it first or second.

Table 4. Results Delphi-study: 'GLM and random forest'

	Ranking per Round	
Scenarios	**2**	**3**
A: Input parameters - background knowledge base.	2	n.a.
B: Input parameters - general idea behind the data.	1	n.a.
C: Input parameters - decision output.	3	n.a.
D: Background knowledge base - general idea behind the data.	4	n.a.
E: Background knowledge base - decision output.	6	n.a.
F: General idea behind the data - decision output.	5	n.a.
Kendall's rank correlation coefficient	0,771	n.a.

4.3.2. Price optimization

After the first round, two 'content' elements were combined to draw-up the first scenario. The second scenario was the 'general idea behind the algorithm' because the substantiations in the first round were most often related to it. Also, the given substantiations showed that some participants found this ADM-system hard to explain, which provided the final two scenarios. Concerning the 'presentation' elements, a 'positive perspective' applies to every scenario. Consensus was not reached because the Kendall's rank correlation coefficient was below the threshold of ≥ 0.7 in both the second and third round. In the third round, of the eight participants, four agreed with the ranking of round two. After round three, scenario b was ranked first (based on the mean ranking) and seven out of eight participants ranked it first or second. Finally, When compared to round two, only scenario c and b switched position in the final round. The rest of the scenarios remained in the same position.

Table 5. Results Delphi-study: 'Price optimization'

	Ranking per Round	
Scenarios	**2**	**3**
A: Input parameters - decision output.	3	3
B: General idea behind the algorithm.	2	1
C: Price optimization should not be used and therefore not explained.	1	2
D: Price optimization can be used but should not be explained towards customers.	4	4
Kendall's rank correlation coefficient	0,233	0,169

4.3.3. Recommendation engine

After the first round, four 'content' elements were combined to draw-up six scenarios. Concerning the 'presentation' elements, a 'group of alternatives', 'multimedia format', and 'positive perspective' apply to every scenario. Consensus was not reached because the Kendall's rank correlation coefficient was below the threshold of ≥ 0.7 in both the second and third round. In the third round, of the eight participants, four agreed with the ranking of round two. After round three, scenario d was ranked first (based on the mean ranking) and six out of eight participants ranked it first or second. Finally, when compared to round two, only scenario a and d switched position in the final round. The rest of the scenarios remained in the same position.

Table 6. Results Delphi-study: 'Recommendation engine'

	Ranking per Round	
Scenarios	**2**	**3**
A: Input parameters - user knowledge base.	1	2
B: Input parameters - general idea behind the data.	3	3
C: Input parameters - decision output.	4	4
D: User knowledge base - general idea behind the data.	2	1
E: User knowledge base - decision output.	5	5
F: General idea behind the data - decision output.	6	6
Kendall's rank correlation coefficient	0,181	0,341

4.3.4. Rule-based fraud detection

After the first round, four 'content' elements were combined to draw-up the first five scenarios. Hereby, the 'knowledge base' elements (both user and background) were combined into a single element to keep the list of scenarios concise and comprehensible. The substantiations in this round were most often related to 'specific procedural decision information', hence this was added as a scenario. Concerning the 'presentation' elements, a 'natural-language format' and 'positive perspective' apply to every scenario. Consensus was not reached because the Kendall's rank correlation coefficient was below the threshold of ≥ 0.7 in both the second and third round. In the third round, of the eight participants, five agreed with the ranking of round two. After round three, scenario b was ranked first (based on the mean ranking) and seven out of eight participants ranked it first or second. Finally, when compared to round two, all scenarios remained in the same position after the final round.

Table 7. Results Delphi-study: ‘Rule-based fraud detection’

	Ranking per Round	
Scenarios	**2**	**3**
A: Input parameters - knowledge base (user and background).	2	2
B: Input parameters - specific procedural decision information.	1	1
C: Input parameters - decision output.	4	4
D: Knowledge base (user and background) - specific procedural decision information.	3	3
E: Knowledge base (user and background) - decision output.	5	5
F: Specific procedural decision information - decision output.	6	6
Kendall’s rank correlation coefficient	0,272	0,563

4.3.5. Optical character recognition (OCR)

After the first round, three ‘content’ elements were combined to draw-up scenarios a, b, and d. The substantiations were most often related to the ‘general idea behind the algorithm’ hence this, in combination with ‘input parameters’, was added as a scenario c. The combination with ‘input parameters’, was however not based on the given substantiations. After the second round, a fifth scenario was added based on the substantiations. Herein the ‘general idea behind the algorithm’ was combined with ‘decision output’. Concerning the ‘presentation’ elements, a ‘multimedia format’ applies to every scenario. Consensus was not reached because the Kendall’s rank correlation coefficient was below the threshold of ≥ 0.7 in both the second and third round. Finally, after round three, scenario c was ranked first (based on the mean ranking) and five out of eight participants ranked it first or second.

Table 8. Results Delphi-study: 'Optical character recognition (OCR)'

	Ranking per Round	
Scenarios	**2**	**3**
A: Input parameters - background knowledge base.	2	5
B: Input parameters - decision output.	4	3
C: Input parameters - general idea behind the algorithm.	1	1
D: Background knowledge base - decision output.	3	4
E: General idea behind the algorithm - decision output.	n.a.	2
Kendall's rank correlation coefficient	0,527	0,072

5. DISCUSSION, CONCLUSIONS AND RECOMMENDATIONS

In this fifth and final chapter, the results will be interpreted. In the first paragraph, the results will be compared with theory and a reflection on the methodology will be provided. Then in the conclusion, a concise summary of the conclusions will be described. Finally, recommendations for practice and further research will be given.

5.1. Discussion - reflection

The research performed can be divided into two main parts (divided into three questions to be answered), which will be addressed in separate paragraphs. These paragraphs will start with a comparison between the research-results and theory. Thereafter, a reflection on the methodology will be provided. During the first part of the research, it was studied what types of ADM-systems are used in the Dutch property & casualty insurance industry. Second, a Delphi-study was held to determine what combinations of ADM-system and explanation are preferred from a property & casualty insurers perspective. Finally, the third paragraph will discuss the ethical aspects of the research.

A general note upfront is that the reliability was strengthened by keeping a research journal in which among others is noted how questions were determined, how feedback from pilots was processed, and wherein an audit trail in the data analysis stage was logged.

5.1.1. Types of ADM-systems used

The results showed that ADM-systems are mainly used within the processes of ‘Product development’ and ‘Pricing and underwriting’. This could however be partly coloured by the occupation of the respondents and their corresponding probable understanding of the use of ADM-systems throughout the value chain. This is also illustrated by the higher number of participants which selected the option ‘unknown’ for the different parts of the value chain, outside of ‘Product development’ and ‘Pricing and underwriting’. The results are in line with the thematical review of EIOPA. Their review, however, showed that big data analytics tools in ‘Product development’ represents only 5% percent of the total use across the value chain (2019). This difference can probably be explained by the fact that the participants, during this study, saw it as a synonym for ‘Pricing and underwriting’. This illustrates that the internal validity would have been strengthened when the different parts of the value-chain were clarified.

To conclude, the results confirmed that ADM-systems are used throughout the value chain (Harris-Ferrante, 2017), but have a focus on pricing.

Methodology reflection

During the data collection phase, demographic data on the company was initially required to check for representativity. After multiple dialogues, however, with a contact of the researcher at the 'Verbond van Verzekeraars' (industry association), this was excluded. This deviation was accepted because it could prevent potential respondents from taking the questionnaire, due to that the researcher worked for an insurer and could provide him with advantageous information on competitors. Instead, a question on the size of the insurer in gross written premium was added.

A self-completed internet questionnaire was used to collect the data. Besides multiple advantages, a recognized and accepted downside of it was the possible lower response rate than an interviewer-completed questionnaire. To increase the response rate, personal contacts of the researcher were actively approached to participate, and a LinkedIn message was drafted in collaboration with a communication specialist. The risk of a low response rate, however, still materialized. This in combination with excluding the demographical question on the company, makes that the results are not statistically significant and cannot be generalized to the Dutch property & casualty insurance market.

As described in the results section, a knowledge assessment was included, and the participants were asked to assess their knowledge on a scale of one to ten. This decision was taken mainly to train the participants but also to ensure sufficient knowledge of ADM and reduce the chance of uninformed responses. This, however, also makes that the results of the questionnaire were very dependent on the quality of the knowledge assessment because it determined which responses were included in the results. To increase the internal validity of the knowledge assessment, it was based on the literature.

Before the questionnaire was held it was pilot tested to increase internal validity and refine the questionnaire. The pilot was based on the questions of Bell and Waters (2014) as described in the technical design. It was pilot tested on fellow students (who have relatively high knowledge of ADM, due to their own research), and on family (who have relatively low knowledge of ADM-systems). In retrospective, both fellow students and family were not representative of the target audience, which negatively impacted the internal validity. This was also made clear by one of the respondents who commented in a personal email to the researcher that some definitions were difficult to interpret and an explanatory note would have helped.

As described in the results chapter, only the type of ADM-systems was determined based on the results questionnaire and the sorts of explanations were included in the first round of the Delphi-study. This decision was taken for two reasons. First, because of the extensive number of possible combinations when all ‘User Interface Components’ of the taxonomy were used. When the number of scenarios would have been too long, respondents could have lost motivation. This would have negatively impacted the output and thus the internal validity. Secondly, another option was that the researcher would have chosen which elements of an explanation was included in the scenarios. This, however, was considered arbitrary which also would have negatively impacted the internal validity.

The responses to the open questions were very concise, thus not sufficient to draw-up the ADM-systems for the Delphi-study. The responses sufficed to partly draw up three ADM-systems and were enriched based on the literature. In addition, to increase the variety in ADM-systems, two ADM-systems were drafted which were based on the thematical review of EIOPA (2019). This combination of questionnaire results and literature improved the internal validity. The drawn-up scenarios formed the basis for the Delphi-study, which will be discussed in the following paragraph.

5.1.2. Preferred combinations of ADM-system and explanation

The main part of this study was the Delphi-study in which different scenarios of explanation elements were ranked, based on preferability, for five ADM-systems. In this paragraph, the results will be reviewed once more and compared with theory, for each ADM-system separately. The table on the next page summarizes the findings and will be further discussed in the following paragraphs. Finally, the methodology will be discussed.

Table 9. Summarized results

ADM-System *Preferred explanation elements*	**Consensus (Kendall's coefficient)**	**Participants which ranked it 1st or 2nd.**
GLM and random forest *Input parameters - general idea behind the data.*	Yes (0,771)	6 out of 7
Price optimization *General idea behind the algorithm.*	No (0,169)	7 out of 8
Recommendation engine *User knowledge base - general idea behind the data.*	No (0,341)	6 out of 8
Rule-based fraud detection *Input parameters - specific procedural decision information.*	No (0,563)	7 out of 8
Optical character recognition (OCR) *Input parameters - general idea behind the algorithm.*	No (0,072)	5 out of 8

GLM and random forest

This type of ADM-system is an example of how 'ad hoc' created groups (Mittelstadt, 2017) are used within the business model of insurers. This shows that the challenges as described by Newell & Marabelli (2015) could therefore be applicable for Dutch insurers. This is especially the case when traditional data sources are combined with newer sources like telematics and online media, through ML techniques (EIOPA, 2019).

The first round showed that the 'input parameters' and 'background knowledge base information' are of importance in the explanation because of the GDPR regulation. This is complementary to that notion that there is a right to an ex-post explanation of specific decisions within the GDPR regulation (Wachter et al., 2017). The general idea behind the data was also preferred because it includes the substantiation of the used pricing variables and should explain why the data is used and why it is objectively justifiable. This addresses a major problem with ADM-systems, namely the bias in these systems (Allen, 2019). After the ranking in the second round, consensus was reached. The 'input parameters' in combination with the 'general idea behind the data' are most important to include in the explanation.

Price optimization

After the first round it seemed like only 'decision output' and 'input parameters' were preferred in an explanation. There was however also discussion on whether price optimization should be used because it is a non-risk related premium surcharge and therefore invalidates the explainability of risk-based pricing. This supplements the concerns of regulators and the insurance industry on the unfair treatment of especially vulnerable consumers (EIOPA, 2019). The first round also showed that it may not be socially acceptable. In the second round, it seemed like the group opinion tended towards not using price optimization. When it will be used, however, the general idea behind the algorithm should be explained, because a procedural explanation seemed to be more fitting than an explanation on an individual basis. The third and final round showed the same conclusion. Consensus, however, was not reached, because there was still discussion on whether price optimization should be used.

Recommendation engine

The first round showed that the 'user knowledge base information' is important in an explanation and should contain the personal information that is used. Also, it is a useful rationale to communicate the effectiveness of the recommendation. The general idea behind the data is also relevant and should explain the importance of the customer's past behaviour on the recommendation. This will lead to more comfort for customers with the recommendation and enables them to make a well-founded choice. The second round had as a result that 'input parameters' and 'user knowledge base information' were preferred. In the third round, this was 'user knowledge base information' in combination with the 'general idea behind the data'. Consensus, however, was not reached.

Rule-based fraud detection

The first round showed that 'input parameters' and 'knowledge base information' (both user and background) are of importance because of the privacy perspective. This should include information regarding the identification of anomalies. The 'specific procedural decision information’ was also seen as important because the social acceptance of differentiation in treatment is very low (e.g. the ‘toeslagenaffaire’ at the Tax Authority), which could lead to reputational damage. This complements the concerns around lack of transparency (Selbst & Barocas, 2018) and biases in ADM-systems (Allen, 2019). Therefore, it is important to be objective and transparent on how a specific decision is made. It is however for this explanation element important to note that there should be a balance between transparency and protection of the capabilities of the insurer. Too much transparency on how fraud is detected gives away an important asset to potential frauds. This confirms the positions of, Kroll et al. (2016) that transparency could be undesirable, and Zarsky (2016) who mentioned perverse effects like ‘gaming the system’. Finally, the ‘decision output’ is of importance. This should include that the output is indicative because if a customer is marked as a fraud by an algorithm, it can be perceived negatively. It should also include the consequences of being identified as a fraud. The second round had as a result that the ‘input parameters’ combined with the ‘specific procedural decision information’ were preferred. There was, however, a minority which would still prefer to include the ‘decision output’ for the reasons described above. The third and final round had the same results, consensus, however, was not reached.

Optical character recognition (OCR)

The first round showed that the ‘input parameters’ and ‘background knowledge base information’ should be included for the substantiation of nonstandard claims. Also, the general idea behind the algorithm should be included. Hereby, it is important to only generally explain how the technique works and consider the differences in knowledge level. This is in line with Zanker & Ninaus (2010) who argued that it is hard to come to a user-tailored explanation without any domain-specific knowledge. The second round had as a result that ‘input parameters’ combined with the ‘general idea behind the algorithm’ were preferred. There was, however, some discussion about whether ‘decision output’ was more important than ‘input parameters’. When considering ‘decision output’, it should be explained what the decision outputs are for specific cases, that expert opinions have the function of a feedback system, and the outputs are used as an indication. The third and final round also showed that the ‘input parameters’ in combination with the ‘general idea behind the algorithm’ are preferred. There was, however, still discussion on whether it should be the ‘input parameters’ or ‘decision output’ in combination with the ‘general idea behind the algorithm’. Therefore, a consensus was not reached.

Methodology reflection

Following the technical design, the 'toolkit' of Day and Bobeva (2005) was followed. Hereby, the first stage was the exploration stage in which the study was planned, participants were selected, and a pilot was conducted. The criteria for participants was that they had sufficient knowledge of ADM, were willing to commit to multiple rounds, and were willing to revise their judgements to reach consensus. To ensure that participants met these criteria an invitation email was sent, and in some cases also a follow-up call was held, wherein an introduction to the research was given including the set-up of the Delphi-study. Secondly, the panellist would be ideally heterogeneous in terms of nationality, occupation/role, and age. Due to the low response rate of the questionnaire, however, the criteria relating to heterogeneity was abandoned in the selection process. Initially, fourteen participants were found who were open to participating in de Delphi Study. Some of them, however, noted that they had doubts about their knowledge level and/or would go on vacation during the study. With these few the agreement was made to try to participate in the first round and drop out during this first round when they assessed their knowledge level as insufficient. Four of the participants dropped out during or before round one.

As opposed to the technical design the participants were given one week instead of two to respond to each round for two reasons: First, to mitigate the risk that the respondents' circumstances, knowledge, and situational context would change too much (Day & Bobeva, 2005). Second, to complete the study before most of the participants would go on their holiday. Before the start of the Delphi-study, the different rounds were pilot tested to increase internal validity and to refine the questionnaires. A second reason for the pilot was to address the potential issue of moulding opinions. The pilot was based on the questions of Bell and Waters (2014). It was pilot tested on friends and family (who had very low knowledge of ADM-systems) because the number of potential participants was too limited to also ask some of them to participate in the pilots. This pilot group was however not representative for the target audience, which negatively impacts the internal validity. This was illustrated by the comments of two participants which said that more examples would have helped them to rank the different scenarios. This shows that the scenarios were not fully recognizable to all respondents. This was also illustrated by the fact that the answers to the open and closed questions were not always in line with each other. This shows that the explanation types were not sufficiently clear and that it has negatively impacted the internal validity.

The second stage of the toolkit was the distillation stage in which the different iterative rounds were held. During the first round, the participants were asked to select which ‘content’ and ‘presentation’ elements they preferred. Initially, the options consisted of four ‘content’ related types of information and three ‘presentation’ facets, which corresponds with the third level of the taxonomy (see Appendix I for the taxonomy). It was however decided to also include the fourth level. This was expected to result in a further clarification on the information that should be included in the explanation; and therefore, have a positive effect on the internal validity. Mainly for the decision inference process (information related to the internal process of the ADM-system), a further clarification was deemed necessary. To be consistent it was decided to not only include the fourth level of the taxonomy for the decision inference process but for all ‘content’ and ‘presentation’ elements.

One of the shortcomings of a Delphi-study is the potential low response rate and high time consumption (Hsu & Sandford, 2007). To address this a high level of communication was maintained during the study, it still, however, manifested itself and negatively impacted both validity and reliability. In the first round, there were nine participants. After the second round, it seemed, based on IP-addresses (to guarantee anonymity), that four participants had dropped out, which would lead to a remaining number of five respondents. This was considered a too low response rate, especially because round three still had to be conducted. Therefore, it was decided to also include the responses of participants who did not seem to have participated in the previous round(s) to the second and third round. After the third round, however, participants were asked to disclose to which round they had participated, which led to a deviating image. Namely, that seven persons had participated in all three rounds, one of them had participated in round one and three, one in round two and three, and finally one in the first and second round. This is also shown in table 10. This, however, does not match with 1. The conclusion on participation based on IP-address and 2. The recorded responses which were nine for round one, seven for round two and eight for round three. The first point, related to the IP-address, can probably be explained by the use of a VPN connection for some participants. The point related to the recorded responses, can either be caused by a failure of the survey tool to record responses, or by participants who did not correctly remember to which round they participated. Regardless of the cause, it leads to an important deficit in this research. Namely, that it is not fully certain who participated to which round and therefor to which group of experts the questionnaires were submitted. This negatively impacts the reliability and internal validity of the research. This could have been prevented by registering in each round who had participated and/or dropped out. Also, the fact that the group of respondents were not identical in every round led to that, the results could not be fully examined on plausibility and consistency, which negatively impacts the internal validity.

Table 10. Participants to the Delphi-study

			Round		
Participant	**Organization**	**Occupation**	**1**	**2**	**3**
1	Insurer	Data analyst	x	x	x
2	Insurer	Data analyst	x	x	x
3	Insurer	Actuary	x		x
4	Insurer	Actuary	x	x	x
5	Insurer	Actuary	x	x	x
6	Industry association	Policymaker	x	x	x
7	Insurer	Product expert	x	x	x
8	Financial services consultancy	Actuary		x	x
9	Insurer	Privacy expert	x	x	
10	Insurer	Actuary	x	x	x
Total number of participants			9	9	9

A specific shortcoming for the ADM-system ‘Optical character recognition (OCR)’ is the inclusion of ‘input parameters’ in scenario c. In retrospective, this was unjustifiably included and could have steered the results, which negatively impacts the internal validity. It does not seem to have negatively impacted the results, however, because in the second and third round it was substantiated ranked first. Finally, to address the issue of unevenly distributed expertise under respondents, the presented aggregated substantiations in round two and three were at a similar level of detail.

The third and final stage of the toolkit was the utilization stage, during which the results were analyzed. During the first round the scenarios were drawn up through the following steps: First, the ‘content’ elements, which were selected by more than 50% of the participants, were included, with exception of the ADM-system ‘Price optimization’. For this ADM-system, it was decided to set the threshold to 40%, because two participants did not select any elements, since they did not consider this ADM-system explainable. These were then combined in pairs of two to make up a scenario. For some ADM-systems, there were also scenarios added based on the provided substantiations. Concerning the ‘presentation’ elements, it was decided not to differentiate in ‘presentation’ elements in the scenarios, because for most ADM-systems there was less discussion on the ‘presentation’ elements. Also, when the ‘presentation’ elements were used to draw-up different scenarios, it would have negatively impacted the conciseness of the number of scenarios. Respondents could have lost motivation, which then would have negatively impacted the output and thus the internal validity. Therefore, all ‘presentation’ elements were included, in every scenario, which was selected by more than 50% of the participants, with again, 40% in the case of ‘Price optimization’.

The results of the second and third round were analyzed using the ‘Kendall’s rank correlation coefficient’ and ranked based on the mean rank. Consensus was reached when the correlation coefficient was ≥ 0.7. For most ADM-systems, no consensus was reached, which can be partly explained by the varying composition of the group of participants.

In the third-round, minority opinions out of round two were presented. These were identified as follows:

- The first step was to identify which participants had a deviating ranking in the second round. It was considered a deviation when the participant ranked the lowest mean ranked scenario as the first, second or third scenario (first and second in case of four scenarios in total; first, second or third in case of six scenarios in total). The reverse applied for the highest mean ranked scenario. When there were six scenarios in total, the same was done for the second to lowest mean ranked scenario. Then it was considered deviating when the participant ranked the second to lowest mean ranked scenario as the first or second scenario. Here also, the reverse applied for the second to the highest mean ranked scenario.
- Following, the substantiations were studied to identify the arguments that support the deviation and the associating explanation elements with it.

These arguments were then considered a minority opinion and were presented in the third round. The initial goal of providing minority opinions was to provide a final opportunity for participants to revise their judgement. This aim was reached partly, because of the varying composition of the group of participants.

Finally, for all three rounds applies that the substantiations, provided by the participants, were coded based on the ‘User Interface Components’ of the Taxonomy. The raw responses of the participants are included in Appendix V, VI, and VII and the coded results can be requested from the researcher.

5.1.3. Ethical aspects

One of the important stages of the research wherein ethical issues could have arisen was when access was sought. Hereby no pressure was applied on intended participant and refusal was accepted as part of the research. This was done ethically and is illustrated by multiple participants who felt comfortable enough to refuse or withdraw from the study. A second issue was that consent should be informed. This was done by providing relevant information for both the questionnaire and the Delphi-study, in the questionnaires and all invitation communication through mail and LinkedIn. This included information on the nature of the research, what the requirements were of taking part in the research, what the rights were of those taking part, and how the collected data would be used. Finally, during the data collection stage, the collection of personal data was minimized, and the results are not traceable to individuals.

5.2. Conclusions

There has been done a lot of research on what ADM-systems are and what the effects are on society. Also, it has been studied how these models could be made more transparent both at the input stage (ex-ante) and the output stage (ex-post). Furthermore, what is understood by an explanation has been explored. However, what sort of explanation fits the ADM-systems best has not been given that much attention. Therefore, the objective of this research was to understand what kind of explanation would fit the different types of ADM-systems, used in the Dutch property & casualty insurance industry best, from an insurers perspective. This was studied by submitting this problem to a group of industry experts through the Delphi-method. For five ADM-systems, used in the insurance value-chain, an explanation was sought. Main limitations were the low response rate and low external validity. Therefore, the conclusions represent only the synthesis of the opinion of this group of participants and are not statistically meaningful.

Recurring during the study was that the data that is used to reach a decision should be included in the explanation because of the privacy perspective. This should be done whether it is provided for a specific decision, such as the car registration number for motor insurance, or already resides in the knowledge base of the ADM-system. A second reoccurring explanation element was that, for most ADM-systems, the general idea behind the algorithm and/or data should be included in the explanation instead of providing the decision inference process information for a specific decision. The main conclusion, however, is that there is not a one-size-fits-all explanation and it depends on the type of ADM-system, for what it is used, and in which social context.

5.3. Recommendations for practice

This study has shown for five ADM-systems what type of explanation could be fitting, and that the explanation depends on different factors, such as, for what the ADM-system is used, and what the social context is. Hence, the Dutch insurance industry is recommended to consider including the described explanation elements in their explanation towards customers. Also, it is recommended to consider different types of explanation for other ADM-systems than the ones used in this study.

5.4. Recommendations for further research

This research was done by performing a Delphi-study, which has an exploratory character and cannot be generalized externally. The results should therefore be a starting point for follow-up research to study whether the conclusions can be generalized to the Dutch property & casualty insurance industry as a whole. The major limitation of this research was the low response rate to both the questionnaire and Delphi-study. Therefore, it is recommended to broaden this research by a bigger and heterogeneous group of participants. Finally, during this study, the Taxonomy from Nunes & Jannach (2017) was used to determine the types of explanation that would fit the used ADM-systems in the Dutch property & casualty insurance industry. Other researchers are encouraged to do this for other industries as well, to learn from practice what type of explanation fit differing ADM-systems.

REFERENCES

Allen, J. A. (2019). The Color of Algorithms: An Analysis and Proposed Research Agenda for Deterring Algorithmic Redlining. *Fordham Urban Law Journal*, *XLVI*.

Arumugam, S., & Bhargavi, R. (2019). A survey on driving behavior analysis in usage based insurance using big data. *Journal of Big Data*. https://doi.org/10.1186/s40537-019-0249-5

Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., … Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, *58*(October 2019), 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

Bell, J., & Waters, S. (2014). *Doing Your Research Project* (6th ed.). Open University Press.

Bolander, T. (2019). What do we loose when machines take the decisions? *Journal of Management and Governance*, *23*(4), 849–867. https://doi.org/10.1007/s10997-019-09493-x

Day, J., & Bobeva, M. (2005). A generic toolkit for the successful management of delphi studies. *Electronic Journal of Business Research Methods*, *3*(2), 103–116.

Delbecq, A. L., Van de Ven, A. H., & Gustafson, D. H. (1975). *Group techniques for program planning*.

Diakopoulos, N. (2016). Accountability in algorithmic decision making. *Communications of the ACM*, *59*(2), 56–62. https://doi.org/10.1145/2844110

Edwards, L., & Veale, M. (2017). Slave to the Algorithm? Why a “Right to an Explanation” Is Probably Not the Remedy You Are Looking For. In *Duke Law & Technology Review* (Vol. 16). https://doi.org/10.2139/ssrn.2972855

EIOPA. (2018). *Industry survey-Big Data thematic review*. Retrieved from http://eur-lex.europa.eu/legal-

EIOPA. (2019). *Big data analytics in motor and health insurance : a thematic review.*

European Parliament. (2016). EU Directive 2016/679 - General Data Protection Regulation (GDPR). *Official Journal of the European Union*. Retrieved from https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32016R0679

Fang, K., Jiang, Y., & Song, M. (2016). Customer profitability forecasting using Big Data analytics: A case study of the insurance industry. *Computers and Industrial Engineering*, *101*, 554–564. https://doi.org/10.1016/j.cie.2016.09.011

Gallego, D., & Bueno, S. (2014). Exploring the application of the Delphi method as a forecasting tool in Information Systems and Technologies research. *Technology Analysis* *and Strategic Management*, *26*(9), 987–999. https://doi.org/10.1080/09537325.2014.941348

Goodman, B., & Flaxman, S. (2017). European union regulations on algorithmic decision making and a “right to explanation.” *AI Magazine*, *38*(3), 50–57. https://doi.org/10.1609/aimag.v38i3.2741

Gordon, T. J. (1994). The Delphi Method in Futures Research Methodology. *AC/UNU Project*.

Grosan, C., & Abraham, A. (2011). Rule-Based Expert Systems. *Intelligent Systems Reference Library*, *17*, 149–185. https://doi.org/10.1007/978-3-642-21004-4_7

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A survey of methods for explaining black box models. *ACM Computing Surveys*, *51*(5). https://doi.org/10.1145/3236009

Harris-Ferrante, K. (2017). *To the Point: Leveraging AI for Success in Digital Insurance.* Barcelona, Spain: Presentation Gartner Symposium ITXPO, Nov. 5-7.

Hsu, C.-C., & Sandford, B. A. (2007). The Delphi Technique: Making Sense of consensus. *Practical Assessment, Research & Evaluation*, *12*(10). https://doi.org/10.1080/02688867.1988.9726654

Kaminski, M. (2019). THE RIGHT TO EXPLANATION, EXPLAINED. *Berkeley Technology Law Journal*, *34*(189), 189–218.

Karanasiou, A. P., & Pinotsis, D. A. (2017). A study into the layers of automated decision-making: emergent normative and legal aspects of deep learning. *International Review of Law, Computers and Technology*, *31*(2), 170–187. https://doi.org/10.1080/13600869.2017.1298499

Kemper, J., & Kolkman, D. (2018). Transparent to whom? No algorithmic accountability without a critical audience. *Information Communication and Society*, 1–16. https://doi.org/10.1080/1369118X.2018.1477967

Kroll, J. A., Huey, J., Barocas, S., Felten, E. W., Reidenberg, J. R., Robinson, D. G., & Yu, H. (2016). Accountable Algorithms. *University of Pennsylvania Law Review*, (633), 66. https://doi.org/10.1002/ejoc.201200111

Lepri, B., Oliver, N., Letouzé, E., Pentland, A., & Vinck, P. (2018). Fair, Transparent, and Accountable Algorithmic Decision-making Processes: The Premise, the Proposed Solutions, and the Open Challenges. *Philosophy and Technology*, *31*(4), 611–627. https://doi.org/10.1007/s13347-017-0279-x

Linstone, H. A., & Turoff, M. (1975). *The Delphi Method: Techniques and Applications*.

Addison-Wesley.

Mittelstadt, B. (2017). From Individual to Group Privacy in Big Data Analytics. *Philosophy and Technology*, *30*(4), 475–494. https://doi.org/10.1007/s13347-017-0253-7

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, *3*(2), 205395171667967. https://doi.org/10.1177/2053951716679679

Newell, S., & Marabelli, M. (2015). Strategic opportunities (and challenges) of algorithmic decision-making: A call for action on the long-term societal effects of "datification." *Journal of Strategic Information Systems*, *24*(1), 3–14. https://doi.org/10.1016/j.jsis.2015.02.001

Nunes, I., & Jannach, D. (2017). A systematic review and taxonomy of explanations in decision support and recommender systems. *User Modeling and User-Adapted Interaction*, *27*, 393–444. https://doi.org/10.1007/s11257-017-9195-0

Nunes, I., Miles, S., Luck, M., & De Lucena, C. J. P. (2012). Investigating explanations to justify choice. *Lecture Notes in Computer Science*, *7379*, 212–224. https://doi.org/10.1007/978-3-642-31454-4_18

O'Neil, C. (2016). *Weapons of math destruction: how big data increases inequality and threatens democracy*. Crown.

Okoli, C., & Pawlowski, S. D. (2004). The Delphi method as a research tool: An example, design considerations and applications. *Information and Management*, *42*(1), 15–29. https://doi.org/10.1016/j.im.2003.11.002

Rumson, A. G., & Hallett, S. H. (2019). Innovations in the use of data facilitating insurance as a resilience mechanism for coastal flood risk. *Science of the Total Environment*, *661*, 598–612. https://doi.org/10.1016/j.scitotenv.2019.01.114

Saunders, M., Lewis, P., & Thornhill, A. (2016). *Research Methods for Business Students*. Pearson.

Schmidt, R. C. (1997). Managing Delphi Surveys Using Nonparametric Statistical Techniques. *Decision Sciences*, *28*(3), 763–774.

Selbst, A. D., & Barocas, S. (2018). The intuitive appeal of explainable machines. *Fordham Law Review*, *87*(3), 1085–1139. https://doi.org/10.2139/ssrn.3126971

Singh, J., Walden, I., Crowcroft, J., & Bacon, J. (2016). *Responsibility & Machine Learning: Part of a Process*. https://doi.org/10.1038/nature14539.12

Tintarev, N., & Masthoff, J. (2011). Designing and Evaluating Explanations for Recommender Systems. *Recommender Systems Handbook*, 479–510.

https://doi.org/10.1007/978-0-387-85820-3_15

Tutt, A. (2017). An FDA for Algorithms. *SSRN Electronic Journal*, 83–123. https://doi.org/10.2139/ssrn.2747994

Verbond van Verzekeraars. (2020). *Toelichting Ethisch Kader*.

Viaene, S., Ayuso, M., Guillen, M., Van Gheel, D., & Dedene, G. (2007). Strategies for detecting fraudulent claims in the automobile insurance industry. *European Journal of Operational Research*, *176*(1), 565–583. https://doi.org/10.1016/j.ejor.2005.08.005

Wachter, S., Mittelstadt, B., & Floridi, L. (2017). European union regulations on algorithmic decision making and a "right to explanation." *International Data Privacy Law*, *7*(2). https://doi.org/10.1609/aimag.v38i3.2741

Wachter, S., Mittelstadt, B., & Russell, C. (2018). Counterfactual Explanations Without Opening The Black Box: Automated Decisions and the GDPR. *Harvard Journal of Law & Technology*, *31*(2).

Waltl, B., & Vogl, R. (2018). Increasing Transparency in Algorithmic- Decision-Making with Explainable AI. *Datenschutz Und Datensicherheit - DuD*, *42*(10), 613–617. https://doi.org/10.1007/s11623-018-1011-4

Zanker, M., & Ninaus, D. (2010). Knowledgeable Explanations for Recommender Systems. *Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology - Volume 01*, 657–660. https://doi.org/10.1109/WI-IAT.2010.131

Zarsky, T. (2016). The Trouble with Algorithmic Decisions: An Analytic Road Map to Examine Efficiency and Fairness in Automated and Opaque Decision Making. *Science Technology and Human Values*, *41*(1), 118–132. https://doi.org/10.1177/0162243915605575

Zerilli, J., Knott, A., Maclaurin, J., & Gavaghan, C. (2018). Transparency in Algorithmic and Human Decision-Making: Is There a Double Standard? *Philosophy & Technology*. https://doi.org/10.1007/s13347-018-0330-6

APPENDICES

Appendix I – Explanation Taxonomy

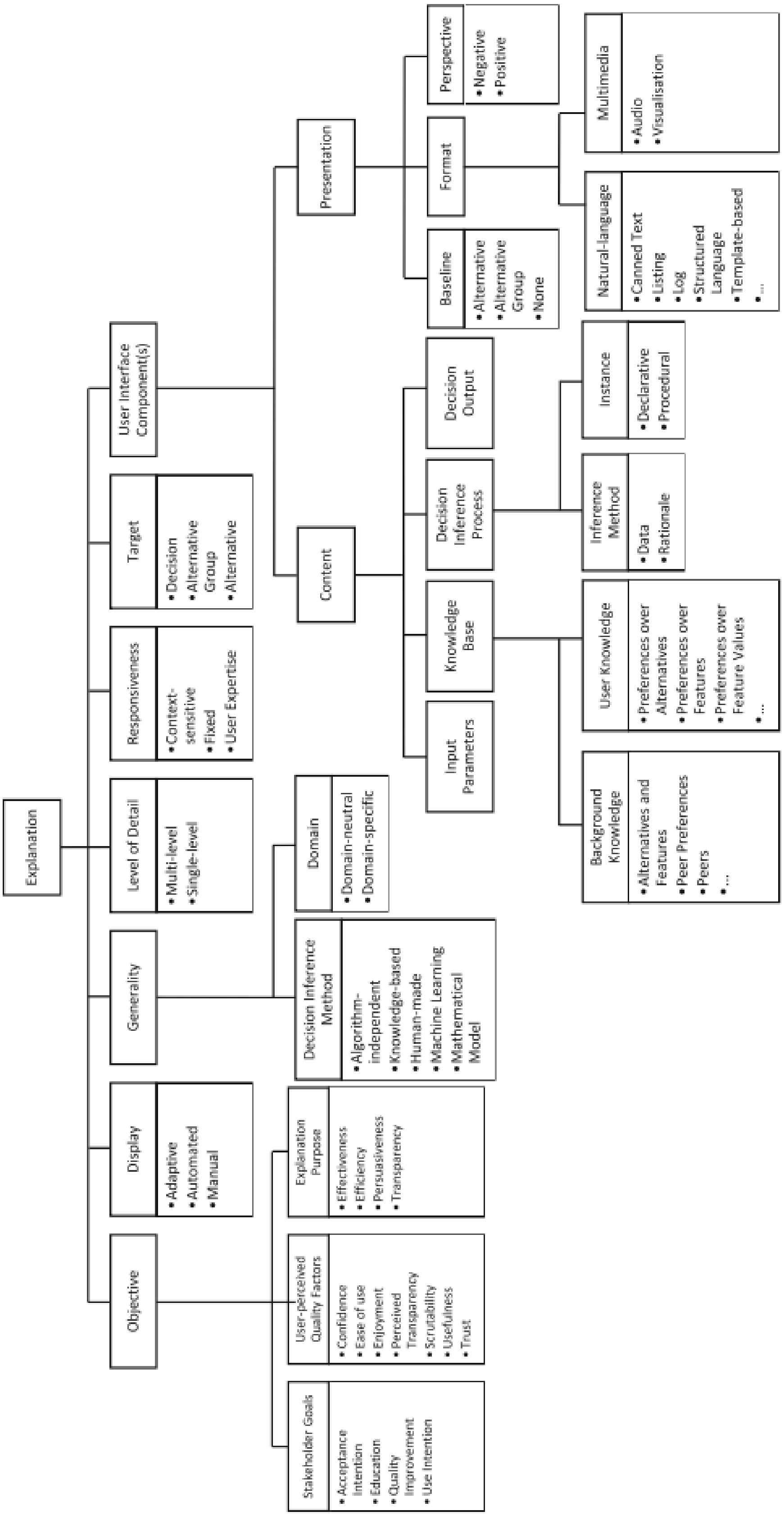

Figure 3. Explanation Taxonomy (Nunes & Jannach, 2017)

Appendix II – Results: Knowledge assessment

Table 11. Knowledge assessment: answers to question 6

Respondent	6. How would you assess your current knowledge level on the use of Algorithmic Decision Making systems in the insurance industry?
1	3
2	1
3	7
4	1
5	3
6	7
7	6
8	6
9	7
10	6
11	6
12	6
13	3

Table 12. Knowledge assessment: answers to question 7

Respondent	7. Which of the choices below describe a rule-based model:
1*	Rule-based models refer to every model that generate statistical rules.
2*	Rule-based models mimic the functioning of the brain by generating rules.
3*	Rule-based models refer to every model that generate statistical rules.
4	Rule-based models mimic the reasoning of a human expert in solving a knowledge-intensive problem.
5*	Rule-based models refer to every model that generate statistical rules.
6	Rule-based models mimic the reasoning of a human expert in solving a knowledge-intensive problem.
7*	Rule-based models refer to every model that generate statistical rules.
8*	Rule-based models refer to every model that generate statistical rules.
9	Rule-based models mimic the reasoning of a human expert in solving a knowledge-intensive problem.
10*	Rule-based models refer to every model that generate statistical rules.
11	Rule-based models mimic the reasoning of a human expert in solving a knowledge-intensive problem.
12*	Rule-based models mimic the functioning of the brain by generating rules.
13*	Rule-based models refer to every model that generate statistical rules.

* Incorrect answer.

Table 13. Knowledge assessment: answers to question 8

Respondent	8. Which of the choices below describe a statistical probabilistic model:
1*	Statistical probabilistic models refer to statistical models whose rules give a probability distribution as a solution.
2	Statistical probabilistic models refer to probabilistic models whose links represent the conditional dependencies between a set of variables.
3*	Statistical probabilistic models refer to statistical models whose rules give a probability distribution as a solution.
4*	Statistical probabilistic models refer to statistical models whose rules give a probability distribution as a solution.
5*	Statistical probabilistic models refer to statistical models whose rules give a probability distribution as a solution.
6*	Statistical probabilistic models refer to statistical models whose rules give a probability distribution as a solution.
7*	Statistical probabilistic models refer to statistical models whose rules give a probability distribution as a solution.
8*	Statistical probabilistic models refer to statistical models whose rules give a probability distribution as a solution.
9	Statistical probabilistic models refer to probabilistic models whose links represent the conditional dependencies between a set of variables.
10*	Statistical probabilistic models refer to statistical models whose rules give a probability distribution as a solution.
11*	Statistical probabilistic models refer to statistical models whose rules give a probability distribution as a solution.
12	Statistical probabilistic models refer to probabilistic models whose links represent the conditional dependencies between a set of variables.
13*	Statistical probabilistic models refer to statistical models whose rules give a probability distribution as a solution.

* Incorrect answer.

Table 14. Knowledge assessment: answers to question 9

Respondent	9. Which of the choices below describe an artificial neural network:
1*	Artificial neural networks refer to statistical models that simulate the atomic processes of the human brain.
2*	Artificial neural networks create rules to simulate the human brain.
3*	Artificial neural networks create rules to simulate the human brain.
4*	Artificial neural networks create rules to simulate the human brain.
5*	Artificial neural networks refer to statistical models that simulate the atomic processes of the human brain.
6	Artificial neural networks try to simulate the atomic processes of the human brain.
7	Artificial neural networks try to simulate the atomic processes of the human brain.
8*	Artificial neural networks create rules to simulate the human brain.
9	Artificial neural networks try to simulate the atomic processes of the human brain.
10*	Artificial neural networks create rules to simulate the human brain.
11*	Artificial neural networks refer to statistical models that simulate the atomic processes of the human brain.
12*	Artificial neural networks refer to statistical models that simulate the atomic processes of the human brain.
13*	Artificial neural networks refer to statistical models that simulate the atomic processes of the human brain.

* Incorrect answer.

Table 15. Knowledge assessment: answers to question 10a

Respondent	10(a). Which statements are true for a rule-based model: Rule-based models are relatively interpretable models.
1*	Don't know
2*	False
3	True
4	True
5*	Don't know
6	True
7	True
8	True
9	True
10	True
11*	False
12	True
13*	Don't know

* Incorrect answer.

Table 16. Knowledge assessment: answers to question 10b

Respondent	10(b). Which statements are true for a rule-based model: Rule-based systems are also known as production systems or expert systems.
1*	Don't know
2	True
3	True
4*	Don't know
5*	Don't know
6	True
7	True
8*	False
9*	False
10*	False
11	True
12*	False
13*	Don't know

* Incorrect answer.

Table 17. Knowledge assessment: answers to question 10c

Respondent	10(c). Which statements are true for a rule-based model: Rule-based models are implicit.
1*	Don't know
2	False
3*	True
4	False
5*	True
6*	Don't know
7	False
8	False
9*	True
10	False
11*	Don't know
12*	True
13*	Don't know

* Incorrect answer.

Table 18. Knowledge assessment: answers to question 11a

Respondent	11(a). Which statements are true for a statistical probabilistic model: Statistical probabilistic models present the relationships between features and target.
1	True
2	True
3	True
4*	Don't know
5	True
6	True
7	True
8	True
9*	False
10	True
11*	Don't know
12	True
13*	Don't know

* Incorrect answer.

Table 19. Knowledge assessment: answers to question 11b

Respondent	11(b). Which statements are true for a statistical probabilistic model: Statistical probabilistic models are explicit.
1*	False
2	True
3*	False
4*	False
5	True
6*	Don't know
7	True
8	True
9	True
10	True
11*	Don't know
12	True
13*	Don't know

* Incorrect answer.

Table 20. Knowledge assessment: answers to question 11c

Respondent	11(c). Which statements are true for a statistical probabilistic model: Statistical probabilistic models are implicit.
1*	True
2	False
3*	True
4*	True
5	False
6*	Don't know
7	False
8	False
9	False
10	False
11*	Don't know
12	False
13*	Don't know

* Incorrect answer.

Table 21. Knowledge assessment: answers to question 12a

Respondent	12(a). Which statements are true for artificial neural networks: Artificial neural networks can never be 100% predictable, error-free or explainable.
1*	False
2*	False
3	True
4	True
5	True
6	True
7	True
8	True
9	True
10*	False
11	True
12	True
13*	Don't know

* Incorrect answer.

Table 22. Knowledge assessment: answers to question 12b

Respondent	12(b). Which statements are true for artificial neural networks: Artificial neural networks are implicit.
1	True
2*	False
3	True
4	True
5	True
6*	Don't know
7	True
8	True
9	True
10*	False
11*	Don't know
12	True
13*	Don't know

* Incorrect answer.

Table 23. Knowledge assessment: answers to question 12c

Respondent	12(c). Which statements are true for artificial neural networks: Artificial neural networks are explicit.
1	False
2*	True
3*	True
4	False
5	False
6*	Don't know
7	False
8	False
9	False
10*	True
11*	True
12*	True
13*	Don't know

* Incorrect answer.

Appendix III – Results: Questionnaire on types of ADM-systems

Table 24. Questionnaire on types of ADM-systems: demographic data

Respon-dent	Age	Education	Organi-zation	Size	Occupation	Knowledge assessment
1*	50 - 65	Master's Degree	P&C insurer	> 500 mil. GWP	Data analyst/scien-tist	3
2	35 - 49	HBO	P&C insurer	< 500 mil. GWP	Compliance	6
3*	20 - 34	Master's Degree	P&C insurer	< 500 mil. GWP	Management	5
4	35 - 49	Master's Degree	P&C insurer	< 500 mil. GWP	Actuary	6
5	35 - 49	Master's Degree	P&C insurer	< 500 mil. GWP	Actuary	6
6	35 - 49	Bachelor's degree	Other insurer	> 500 mil. GWP	Management	6
7	50 - 65	Academic Doctorate Degree	P&C insurer	> 500 mil. GWP	Management	10
8	50 - 65	Professional Doctorate Degree	P&C insurer	< 500 mil. GWP	Management	8
9	35 - 49	Master's Degree	P&C insurer	< 500 mil. GWP	Actuary	9
10*	35 - 49	Academic Doctorate Degree	P&C insurer	< 500 mil. GWP	Policy maker	5
11*	50 - 65	MBO	P&C insurer	> 500 mil. GWP	Risk management	3
12	20 - 34	Master's Degree	P&C insurer	< 500 mil. GWP	Strategic Marketeer Intelligence	7
13*	50 - 65	HBO	Other insurer	< 500 mil. GWP	Management	0

* Not included in the results because the score on the knowledge assessment was < 6.

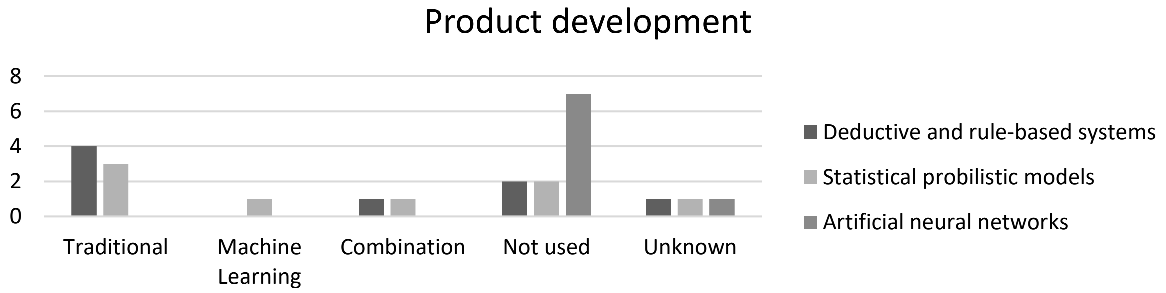

Figure 4. Questionnaire on types of ADM-systems: product development - results

Table 25. Questionnaire on types of ADM-systems: product development - substantiations

Respondent	Substantiation
1*	No models are used
2	?
3*	niet van toepassing
4	Most commonly generalized linear models (GLM)
5	NA
6	pricing GLM/random forrest
7	,
8	Rescue and Igloo applications of WTW for P&C
9	setting a initial pricing based on available information from comparable products or market prices
10*	Opmerking: product ontwikkeling en pricing zie ik als in elkaar doorwerkende activiteiten.
11*	claim handeling business rules
12	.
13*	Beslisbomen in acceptatievraagstukken, randomforest in pricing en segmentatie.

* Not included in the results because the score on the knowledge assessment was < 6.

Table 26. Questionnaire on types of ADM-systems: product development – closed questions

Respondent	Use of ADM-systems	Deductive and rule-based systems	Statistical probabilistic models	Artificial neural networks
1*	Unknown	Unknown	Unknown	Unknown
2	Yes	Used (traditional)	Used (traditional)	Not used
3*	Yes	Used (traditional)	Used (traditional)	Not used
4	No	Not used	Not used	Not used
5	Yes	Used (traditional)	Used (combination of machine learning and traditional)	Not used
6	Yes	Used (combination of machine learning and traditional)	Used (traditional)	Not used
7	Yes	Used (traditional)	Used (traditional)	Not used
8	Yes	Used (machine learning)	Used (traditional)	Not used
9	Yes	Used (traditional)	Used (traditional)	Not used
10*	Unknown	Unknown	Unknown	Unknown
11*	Yes	Used (traditional)	Used (combination of machine learning and traditional)	Not used
12	Yes	Used (traditional)	Not used	Unknown
13*	Yes	Used (traditional)	Unknown	Unknown

* Not included in the results because the score on the knowledge assessment was < 6.

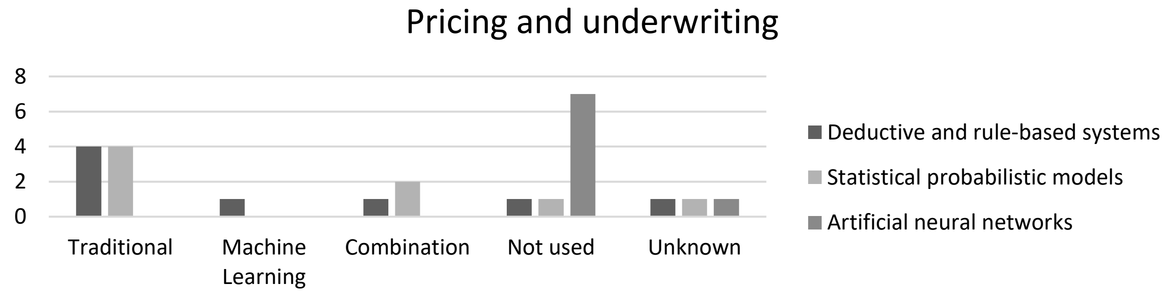

Figure 5. Questionnaire on types of ADM-systems: pricing and underwriting - results

Table 27. Questionnaire on types of ADM-systems: pricing and underwriting - substantiations

Respondent	Substantiation
1*	For pricing and risk analysis regression models are used.
2	?
3*	Voor wat Pricing van onze schadeverzekeringen maken we gebruik van traditionele GLM modellen. Om deze GLM modellen te verbeteren wordt er op de achtergrond gebruik gemaakt van Machine Learing technieken om het traditioneel model te verbeteren. Dit omdat we aan de voorkant nog niet met niet-traditionele modellen willen/mogen/kunnen werken. Verder wordt er ook veel geëxperimenteerd met bovenstaande modellen. Dit is meer voor eigen inzichten binnen het team en voor ervaring opdoen met deze technieken.
4	GLM
5	NA
6	GLM + random forrest
7	.
8	See before for claims amount en frequency
9	used for defining and selecting the risk variables for premium formularia
10*	GLM-methoden voor pricing.
11*	?
12	.
13*	klantgroep segmentatie

* Not included in the results because the score on the knowledge assessment was < 6.

Table 28. Questionnaire on types of ADM-systems: pricing and underwriting – closed questions

Respondent	Use of ADM-systems	Deductive and rule-based systems	Statistical probabilistic models	Artificial neural networks
1*	Unknown	Unknown	Unknown	Unknown
2	No	Not used	Not used	Not used
3*	Yes	Used (traditional)	Used (traditional)	Not used
4	No	Not used	Not used	Not used
5	Yes	Used (traditional)	Used (machine learning)	Not used
6	Yes	Used (combination of machine learning and traditional)	Not used	Not used
7	Yes	Used (traditional)	Used (traditional)	Not used
8	Yes	Not used	Used (traditional)	Not used
9	Yes	Used (traditional)	Used (traditional)	Not used
10*	Yes	Used (traditional)	Unknown	Unknown
11*	Yes	Used (traditional)	Used (combination of machine learning and traditional)	Not used
12	Yes	Used (traditional)	Unknown	Unknown
13*	Yes	Used (traditional)	Not used	Unknown

* Not included in the results because the score on the knowledge assessment was < 6.

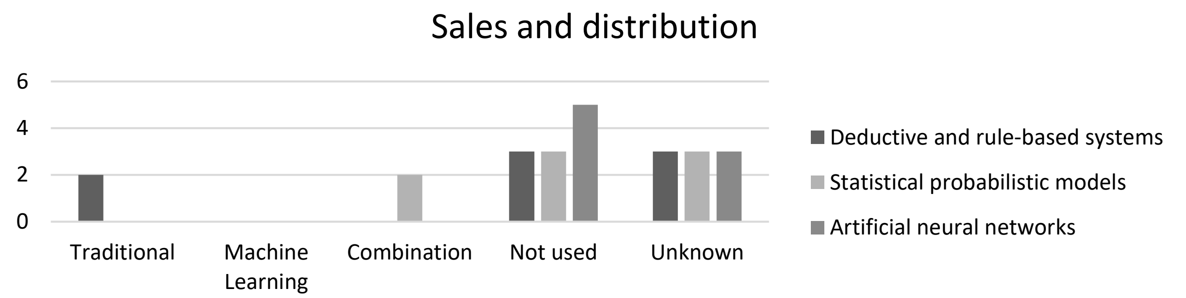

Figure 6. Questionnaire on types of ADM-systems: sales and distribution - results

Table 29. Questionnaire on types of ADM-systems: sales and distribution - substantiations

Respondent	Substantiation
1*	Sales and distribution is the task of ABN AMRO Bank, our joint venture partner. So our company doesn't perform theses tasks.
2	?
3*	Nvt
4	Not that I am ware of, but simple statisitical models might be used.
5	NA
6	recommendation engine for next best action
7	.
8	Not aplicable
9	..
10*	Pricing gebruikt GLM om relevante factoren vast te stellen die in het risicomodel en na kostenopslagen en commerciële opslagen/kortingen leidt tot commercieel model. Dit wordt qua logica in het pricing systeem ingebouwd en vervolgens via de website aan de klant aangeboden. In die zin zit hier geen ADM-systematiek in bij mijn weten.
11*	?
12	.
13*	Next best salesactie

* Not included in the results because the score on the knowledge assessment was < 6.

Table 30. Questionnaire on types of ADM-systems: sales and distribution – closed questions

Respondent	Use of ADM-systems	Deductive and rule-based systems	Statistical probabilistic models	Artificial neural networks
1*	Unknown	Unknown	Unknown	Unknown
2	No	Not used	Not used	Not used
3*	Unknown	Unknown	Unknown	Unknown
4	Unknown	Unknown	Unknown	Unknown
5	Yes	Used (traditional)	Used (combination of machine learning and traditional)	Not used
6	No	Not used	Not used	Not used
7	No	Not used	Not used	Not used
8	No	Not used	Not used	Not used
9	No	Not used	Not used	Not used
10*	Unknown	Unknown	Unknown	Unknown
11*	Yes	Used (traditional)	Used (combination of machine learning and traditional)	Not used
12	Yes	Not used	Not used	Unknown
13*	Yes	Not used	Not used	Unknown

* Not included in the results because the score on the knowledge assessment was < 6.

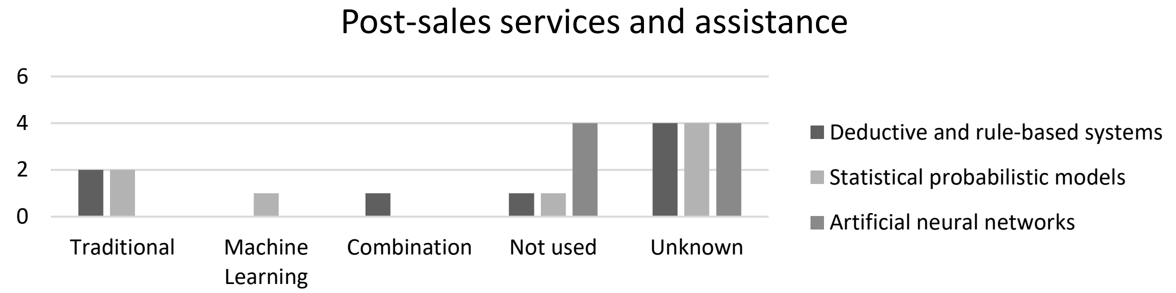

Figure 7. Questionnaire on types of ADM-systems: post-sales services and assistance - results

Table 31. Questionnaire on types of ADM-systems: post-sales services and assistance - substantiations

Respondent	Substantiation
1*	This is not conducted at our firm.
2	?
3*	nvt
4	Not used as far as I am aware.
5	NA
6	classifier for email routing
7	,
8	Predication call traffic and number of claims
9	...
10*	Toetsing van fraude met systemen die gebruik maken van patroon herkenning en dergelijke.
11*	?
12	.
13*	Onbekend

* Not included in the results because the score on the knowledge assessment was < 6.

Table 32. Questionnaire on types of ADM-systems: post-sales services and assistance – closed questions

Respondent	Use of ADM-systems	Deductive and rule-based systems	Statistical probabilistic models	Artificial neural networks
1*	Unknown	Unknown	Unknown	Unknown
2	No	Not used	Not used	Not used
3*	Unknown	Unknown	Unknown	Unknown
4	Unknown	Unknown	Unknown	Unknown
5	Yes	Used (traditional)	Used (machine learning)	Not used
6	Yes	Used (combination of machine learning and traditional)	Used (traditional)	Not used
7	Yes	Used (traditional)	Used (traditional)	Not used
8	No	Not used	Not used	Not used
9	Yes	Not used	Not used	Used (combination of machine learning and traditional)
10*	Unknown	Unknown	Unknown	Unknown
11*	Unknown	Unknown	Unknown	Unknown
12	Unknown	Unknown	Unknown	Unknown
13*	Unknown	Unknown	Unknown	Unknown

* Not included in the results because the score on the knowledge assessment was < 6.

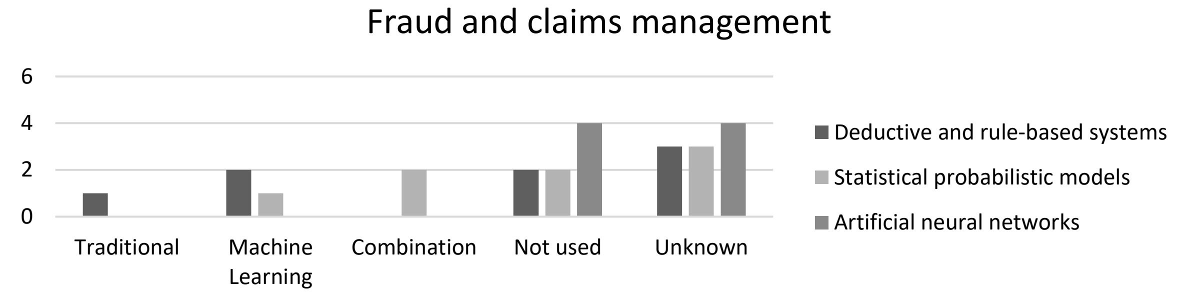

Figure 8. Questionnaire on types of ADM-systems: fraud and claims management - results

Table 33. Questionnaire on types of ADM-systems: fraud and claims management - substantiations

Respondent	Substantiation
1*	The fraud department has started a pilot with machine learning. But I don't know what the status is of that project.
2	?
3*	nvt
4	I would expect so, but I do not know for certain.
5	NA
6	fraud detection and fraud prevention
7	.
8	Fraud signals
9	...
10*	Zie eerdere beantwoording.
11*	business rules for claim handling
12	.
13*	Op groepsniveau zeker.

* Not included in the results because the score on the knowledge assessment was < 6.

Table 34. Questionnaire on types of ADM-systems: fraud and claims management – closed questions

Respondent	Use of ADM-systems	Deductive and rule-based systems	Statistical probabilistic models	Artificial neural networks
1*	Unknown	Unknown	Unknown	Unknown
2	No	Not used	Not used	Not used
3*	Unknown	Unknown	Unknown	Unknown
4	Unknown	Unknown	Unknown	Unknown
5	Yes	Used (traditional)	Used (combination of machine learning and traditional)	Not used
6	No	Not used	Not used	Not used
7	Yes	Used (machine learning)	Used (machine learning)	Not used
8	No	Not used	Not used	Not used
9	Yes	Used (traditional)	Used (traditional)	Used (traditional)
10*	Yes	Used (traditional)	Unknown	Unknown
11*	Unknown	Used (machine learning)	Used (combination of machine learning and traditional)	Unknown
12	Yes	Used (traditional)	Unknown	Unknown
13*	Yes	Used (traditional)	Unknown	Unknown

* Not included in the results because the score on the knowledge assessment was < 6.

Appendix IV – Theory on Explainability

When determining an explanation for an ADM-system there are different aspects to be considered. For this study, 'explainability' is considered to have 'content' and 'presentation' elements.

Regarding 'content' we consider four types of information that can be included in the explanation:

- **Input parameters** refer to the inputs that were provided for a particular decision problem. Such as the car registration number for motor insurance.
- **Knowledge base information** resides in the knowledge base of an ADM-system. The provided information can be:
    - user knowledge (tailored to the specific user) or
    - background knowledge (selected independent of the current user).
- **Decision inference process** information is related to the internal process of the ADM-system. This can refer to the:
    - Specific decision, which can be:
        - Procedural (i.e. describe the steps taken to reach a decision).
        - Declarative (such as the confidence in the decision).
    - General idea behind the
        - Algorithm (e.g. recommendation of alternatives that similar users like);
        - Data (e.g. use of users' driving behaviour to set the premium discount).
- **Decision output** refers to the decision reasoning outcome and, for example, describes the particular features and feature values of the recommended and non-recommended alternatives. An example of this is when the pros and cons of the alternatives are explained.

Looking at the 'presentation' element we consider three facets:

- **Baselines**: The baseline for comparison can be
    - a single alternative to that recommended or
    - a group of alternatives.
- **Formats**: Different output formats can be chosen such as:
    - Natural language or
    - Different types of multimedia (i.e. audio, images or film)
- **Perspective**: The perspective in which an explanation is presented can be either:
    - Positive (why an alternative is suitable for a user) or
    - Negative (why certain negative aspects of an alternative could be acceptable).

Appendix V – Results: 1st round Delphi-study

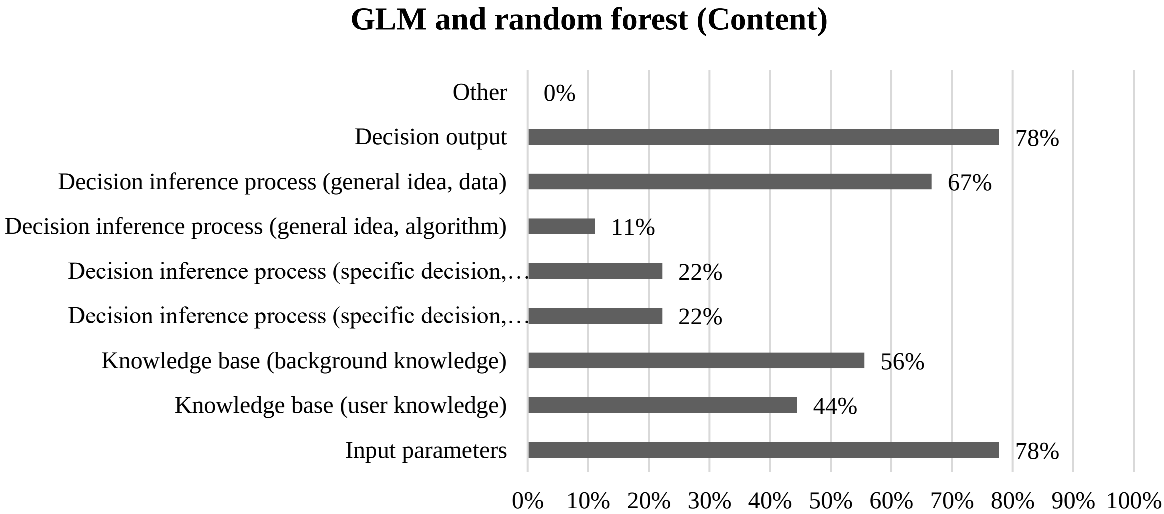

Figure 9. 1st round Delphi-study: GLM and random forest (content) - results

Table 35. 1st round Delphi-study: GLM and random forest (content) – substantiations

Respondent	Substantiation
1	Ik zou me beperken tot het grote verhaal. Van belang is daarbij wel: - Welke marktkennis en gebruikerskennis (en aannames) worden gehanteerd als uitgangspunt - Wat is onze visie op data - Decision output: vooral onderbouwing keuze gebruikte premiefactoren. Alternatieven niet noemen.
2	- Ik denk dat je de info beperkt moet houden om informatie overload te voorkomen - Ik denk dat input parameters noodzakelijk zijn op grond van de AVG (als de klant daar om vraagt sowieso, maar als je dit standaard transparant kunt maken, doe je het heel netjes) - Ik denk dat klanten vooral zullen willen weten qua content dus welke info er in ging, en wat de decision output is - Bij presentation kies ik steeds beide perspective antwoorden, omdat ik denk dat men vooral zal willen weten, waarom men wel/niet in aanmerking komt voor bepaalde alternatieven. - Ik maak geen verschil tussen verschillende typen algoritmen in m'n antwoorden, omdat ik me niet kan voorstellen dat de informatiebehoefte van klanten zo sterk verschilt per type: het gaat steeds om een besluit van de verzekeraar. Enfin: voorbeelden hadden mij als gezegd geholpen om een meer afgewogen oordeel te geven.
3	Voor een klant is het belangrijk dat hij/zij zelf logisch kan beredeneren waarom hij/zij een hogere premie moet betalen. Bijvoorbeeld als hij/zij alle gegevens hetzelfde laat maar het adres van de buurman invult kan niet zomaar de premie veranderen. Dan moet logisch zijn. Om het resultaat te kunnen beredeneren moet je weten wat er in het model gaat. Wat er globaal gebeurt en waarom de keuze is gevallen zoals die is. Daarnaast zou ik wat willen weten over de betrouwbaarheid van de decision. Is die laag, dan zou

	het niet gebruikt mogen worden. Mijn inziens voegt het weinig toe om iets te vertellen over het algoritme zelf omdat te ingewikkeld is.
4	In order to explain the output of the model to customers, it is required to state which input has been used, how the model works (general idea, conceptual explanation of algorithm and way of interference with data) and the decision output.
5	Transparantie over de individuele data, welke wordt gebruikt in het prijsproces, zodat dit voor de klant volstrekt duidelijk is (mede in kader van privacy wetgeving). Het onderliggende rekenmechanisme GLM en/of random forest om te komen tot de prijsstelling vind ik geen content voor klanten. De decision output beschouw ik hier als de uiteindelijke consumentenprijs.
6	Bij het uitleggen hoe een prijs tot stand komt, lijkt mij de objectivering het belangrijkst. Prijsstelling is ook een belangrijke bij het onderscheid maken tussen klanten. dat mag alleen maar obv 'feiten'.
7	as much clarity for client as possible
8	Met name van belang dat toegelicht wordt (of kan worden) welke data gebruikt wordt in uiteindelijke beslissing en dat ook uitgelegd kan worden waarom deze data van belang zijn in de beslissen. Naar mijn mening speelt de techniek die erachter zit hierin een minder grote rol en is het denk ik ook minder relevant om de uitkomst af te wegen tegen alternatieven maar is het belangrijker dat de uitkomst op zichzelf uitlegbaar is.
9	to transparant, you need to share the most important parts of the information. There should be a balance between complete in your information and the overkill of information. In other words, say what the machine does but take into account the difference acknowledge level.

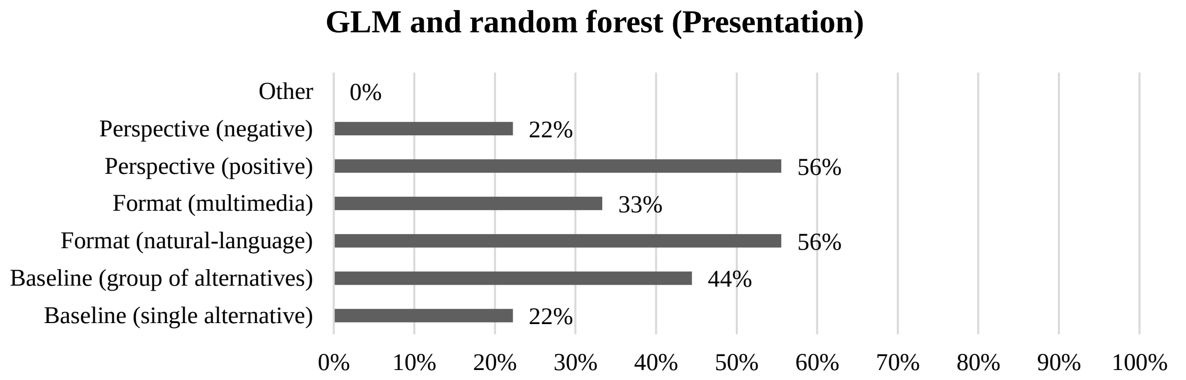

Figure 10. 1st round Delphi-study: GLM and random forest (presentation) - results

Table 36. 1st round Delphi-study: GLM and random forest (presentation) – substantiations

Respondent	Substantiation
1	Ik zou naar de klant niet ingaan op alternatieven. Dit leidt alleen maar tot vragen. Belangrijk is één helder verhaal. Om dit voor iedereen helder te hebben, is het belangrijk dat dit in meerdere formats uitgedrukt wordt.
2	previous answer was lost due to technical problems
3	De prijs is redelijk recht toe recht aan maar het zou wel mooi zijn dat als er alternatieven mogelijk zijn. Dat dat getoond wordt. Bijvoorbeeld bij de dekking van een autoverzekering. All risk versus beperkt casco oid
4	From model perspective, it is required to present the baseline: preferred solution for the customer and possible alternatives. Other elements are in my opinion less important in order to explain the model. In order to increase the sales, field tests should help which form fits best to a customer.
5	De uiteindelijke consumentenprijs kan gepresenteerd worden met alternatieve opties (aanvullende dekkingen, andere eigen behouden en/of output voor vergelijkbare klanten), zodat de klant hierin nog kan kiezen.
6	prijs moet m.i. eenduidig uitgelegd worden. Dit is wat het is. Omdat prijs als complex ervaren kan worden, zou ik kiezen voor de combinatie van beeld, geluid en schrift.
7	as much information for client as possible
8	Voor afweging baselining zie boven. Aangezien je met deze techniek zo specifiek mogelijk prijsbepaling wilt doen leent een persoonlijke benadering en daarom natural language zich in mijn ogen beter dan algemene benadering via multimedia. Daarnaast denk ik per definitie dat beter is een positief perspectief te hanteren ipv negatief. Dit zal in geval de prijs relatief gunstig is uiteraard eenvoudiger zijn. Maar ook op moment dat iemand in hoge risico categorie valt is het waardevol om hier op een goede en positieve manier toelichting op te geven omdat dit juist ook preventief inzicht kan geven richting klant.
9	Be transparant. So from that perspective, show both sides of the Medal. This will prove that the company is not using the model in the own favor.

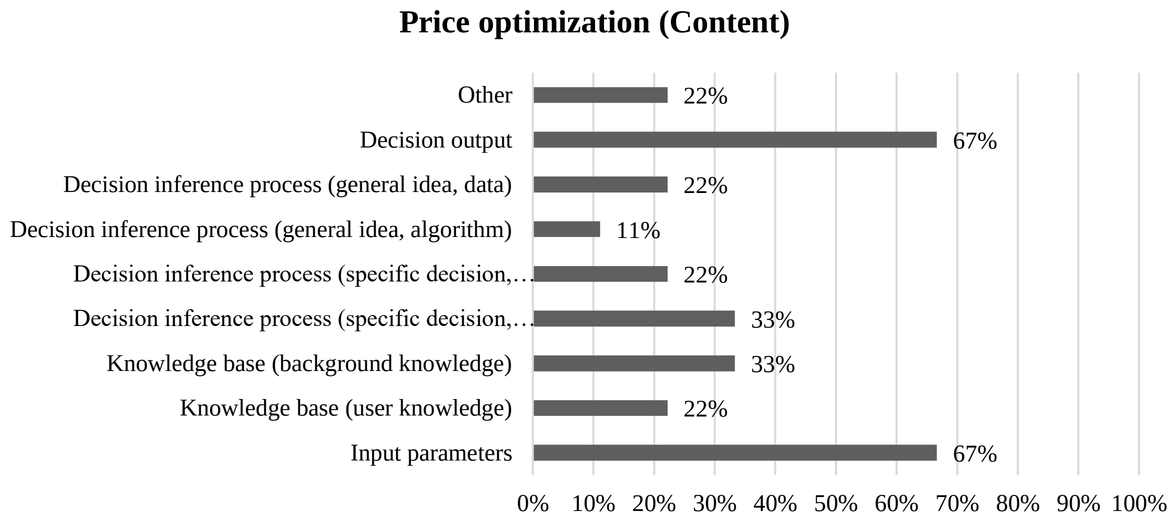

Figure 11. 1st round Delphi-study: Price optimization (content) - results

Table 37. 1st round Delphi-study: Price optimization (content) – substantiations

Respondent	Substantiation
1	Ik vraag me af of dynamic pricing (daar gaat het hier om denk ik) überhaupt uit te leggen is aan de klant. Zo snel dit ook maar enigszins transparant gedaan wordt, dan verliest dit zijn werking bij rationeel polishoudergedrag. Daarnaast botst dit ook met de uitlegbaarheid van het vorige model: daar wordt energie gestoken in het nauwkeurig uitleggen van het risicomodel. Een niet-risico gerelateerde opslag ontkracht deze eerste uitleg.
2	previous answer was lost due to technical problems
3	Voorbeeld dat hier op gaat is dat als je wordt ingeschat als niet prijs sensitief dan krijg je een hogere premie. Dat is op zich een boodschap richting klanten die moeilijk te communiceren is. Want dan is het beste advies om altijd te switchen. Wil je dat wel in de markt? Ik vind wel dat je helder moet communiceren naar klanten wat meegenomen wordt bij prijsbepaling. Maar de pragmatische insteek is dan om dat niet zo expliciet te duiden. Of dat helemaal ethisch is een ander verhaal.
4	The applied techniques are different compared to the previous case. Most important is to explain the conceptual soundness how the model is constructed and how reliable the results are.
5	Prijsoptimalisatie is een intern bedrijfsproces, waarbij groepen klanten worden gecategoriseerd op basis van gedragskenmerken en additioneel concurrentiegegevens worden gebruikt om tot optimale prijs te komen en meer groei te realiseren. Deze kennis is additioneel op de onderliggende GLM of random forest uitkomsten, hetgeen tot de uiteindelijke consumentenprijs leidt. Transparantie over dit interne proces vind ik geen content voor klanten (hoe de bakker zijn brood bakt is zijn geheim). We kunnen als verzekeraar wel transparant zijn, dat we prijsoptimalisatie toepassen en hier algemene knowledge base background informatie aan klanten kunnen verschaffen.

6	Juridisch wespennest om los van het risico onderscheid te maken. Maatschappelijk lijkt me dat ook niet aanvaardbaar. Tenslotte de vraag of het kan in een transparante markt waar de klant zoekt naar de laagste premie.
7	give client insight in how its data is used
8	Allereerst vind ik dat je dit sowieso niet zou moeten willen en daarmee is de toelichting dus ook niet relevant. Als je dit toch wilt doen lijkt mij een meer procedurele uitleg in dit geval belangrijker dan specifieke toelichting op individuele basis. Je kunt bv beter iets zeggen in de trend "we kijken naar een passende prijs voor verschillende klantgroepen op basis van consumentengedrag" ipv "vanwege uw salaris valt u in de categorie die bereid is een hogere premie te betalen dus rekenen we die".
9	to transparant, you need to share the most important parts of the information. There should be a balance between complete in your information and the overkill of information. In other words, say what the machine does but take into account the difference acknowledge level.

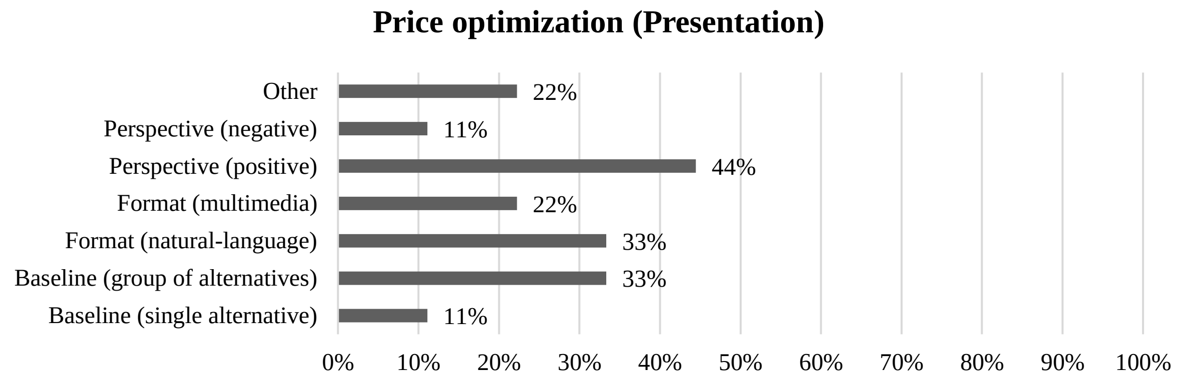

Figure 12. 1st round Delphi-study: Price optimization (presentation) - results

Table 38. 1st round Delphi-study: Price optimization (presentation) – substantiations

Respondent	Substantiation
1	Zie vorige comment
2	previous answer was lost due to technical problems
3	Ook hier altijd goed om alternatieven te duiden. Maar ga je dan ook duiden dat een ander gedrag niet gerelateerd aan het risico to een andere optimalisatie e en prijs leidt. Ik zie dat niet helemaal voor me.
4	See also answer in the previous case.
5	Zie uitkomsten GLM en random forest
6	zie boven
7	give as much unbiased info as possible
8	Zie ook boven. In dit geval zou je communicatie niet te specifiek moeten maken naar mijn idee en dus leent multimedia zich daar beter voor. Perspectief zou positief moeten zijn in de trend van 'wat kun je hier als klant aan hebben'. Het baselinen tov alternatieven zie ik niet veel toegevoegde waarde in.
9	Be transparant. So from that perspective, show both sides of the Medal. This will prove that the company is not using the model in the own favor.

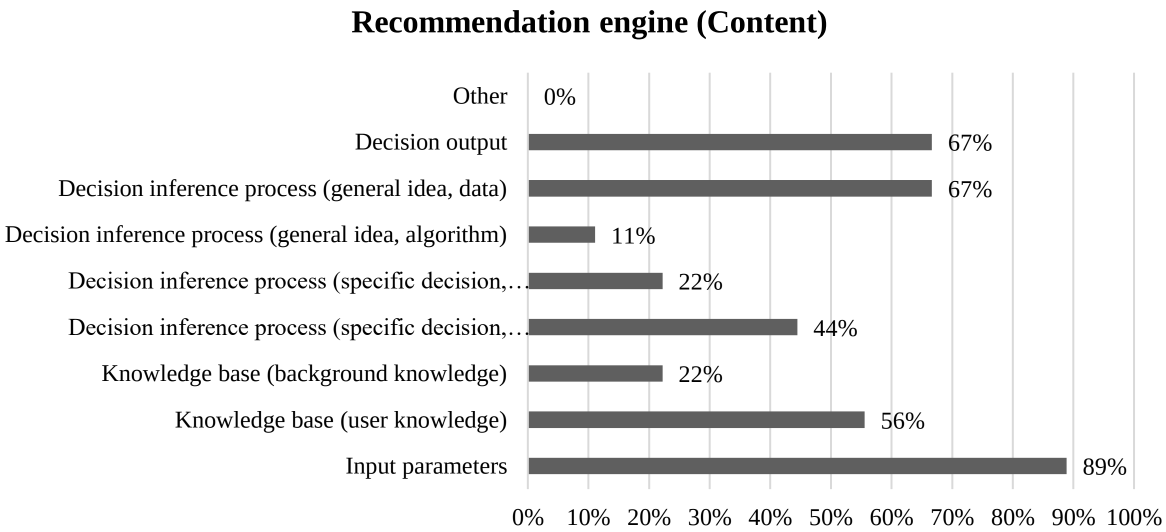

Figure 13. 1st round Delphi-study: Recommendation engine (content) - results

Table 39. 1st round Delphi-study: Recommendation engine (content) – substantiations

Respondent	Substantiation
1	Belangrijkst is dat de uiteindelijke boodschap aankomt en efficient is. Naar de klant toe is onderbouwing van het gekozen kanaal/ tijdstip/ vorm etc niet belangrijk. Er staat geen tegenprestatie van de klant (bv premie) tegenover, dit verplicht ook minder tot onderbouwing. Om effectiviteit boodschap te onderbouwen kan het wel nuttig zijn persoonlijke informatie te communicaren: waarom is een bepaalde uiting gericht op juist de ontvanger?
2	previous answer was lost due to technical problems
3	Ook hier de basis moet helder zijn. Welke data wordt gebruikt en welke keuze is waarom gemaakt
4	Apart from the elements from the previous cases, it is also important to explain the impact of the customer's past behaviour on the price.
5	Bij een advies systeem voor klanten is meer transparantie over de onderliggende data en kennis van het beslisproces nodig. Hierdoor voelen klanten meer comfort bij het advies en hebben meer inzicht, waarom dit advies voor hen de beste keuze is. Zij moeten gefundeerd een keuze kunnen maken..
6	Op grond van jouw……. denken wij dat.... omdat je.... en onze ervaring is dat
7	as much info as possible on reliability of the outcome
8	Hier is het denk ik vooral belangrijk om het data aspect goed te belichten, dus op hoofdlijnen welke data gebruik je en hoe resulteert dit uiteindelijk tot de beslissing.
9	Always easy and clear message to explain to the customer. This has to do with the indirect effect onthe use of the output. It's a recommendation, not a decision with consequences.

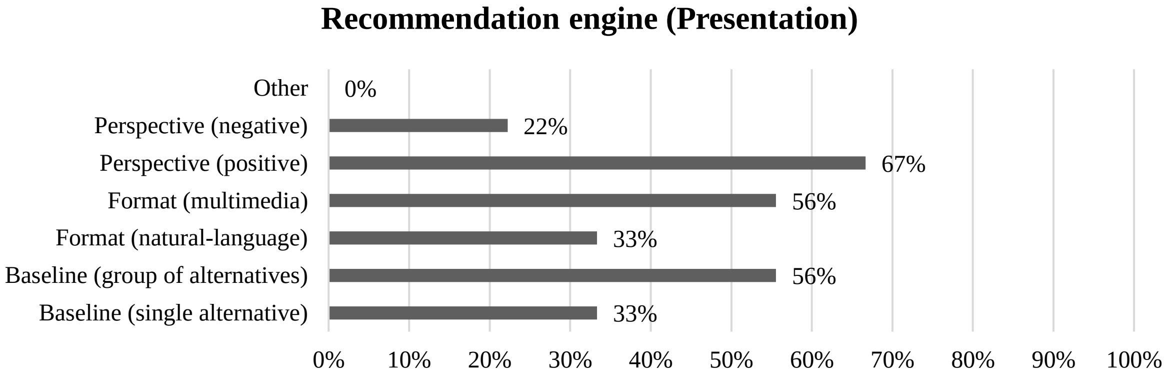

Figure 14. 1st round Delphi-study: Recommendation engine (presentation) - results

Table 40. 1st round Delphi-study: Recommendation engine (presentation) – substantiations

Respondent	Substantiation
1	Alternatieven zijn niet belangrijk, wel flexibiliteit in vorm om maximaal publiek te bereiken.
2	previous answer was lost due to technical problems
3	Dit leent zich om meerdere alternatieven te tonen. Ik zie in de verzekeringsmarkt nog niet helemaal de multimedia optie ed
4	I do not see any difference regarding the presentation to the customer (from a general perspective).
5	Een "next best action" met hooguit een single alternatief aanvullen , omdat dan het idee van de right message weggaat en de klant door de bomen het bos niet meer ziet.
6	kan als 'push' worden ervaren dus keuzes, maar twee lijkt mij voldoende. multimedia, omdat je een (latente) behoefte of snaar wil raken. en uiteraard uitleggen dat je een keus aanbiedt om de klant te helpen
7	easy to understand and put in perspective
8	Belangrijk hierbij is duidelijk benadrukken dat je de klant hiermee werk uit handen neemt. In dit geval is het wel zinvol om een alternatieven mee te nemen in de presentatie om ook inzichtelijk te maken dat hiermee een stuk besluitvorming voor je gedaan wordt die je anders zelf had moeten doen als klant
9	Always easy and clear message to explain to the customer. This has to do with the indirect effect onthe use of the output. It's a recommendation, not a decision with consequences.

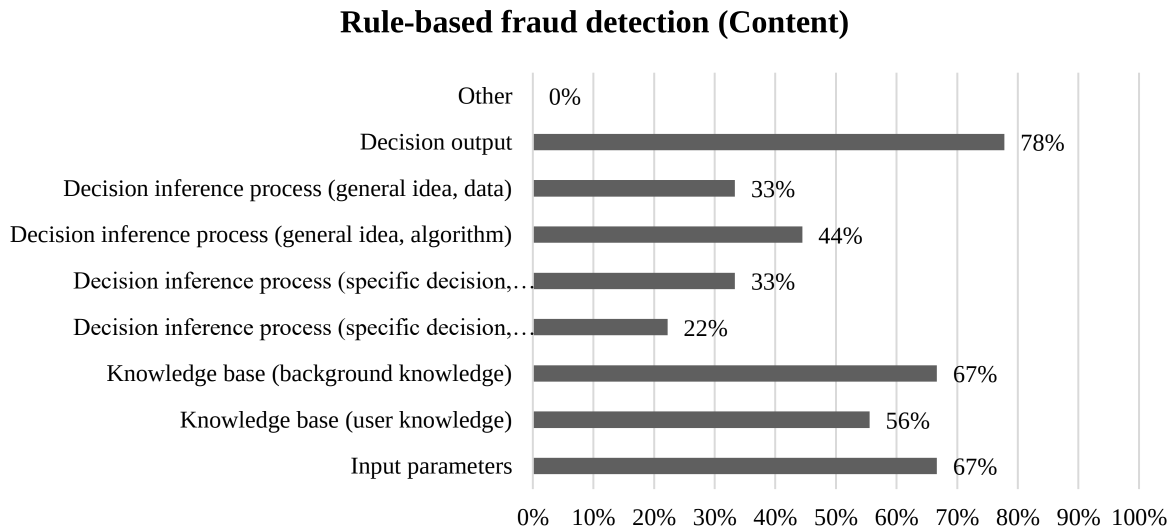

Figure 15. 1st round Delphi-study: Rule-based fraud detection (content) - results

Table 41. 1st round Delphi-study: Rule-based fraud detection (content) – substantiations

Respondent	Substantiation
1	Een vermoeden van fraude moet nauwkeurig en transparant onderbouwd worden. Daarom is het van belang veel informatie te ontsluiten, natuurlijk op een zo duidelijk mogelijke wijze.
2	previous answer was lost due to technical problems
3	Het moet duidelijk zijn welke data over de klant wordt gebruikt om tot een anomaly te komen. Mede vanuit privacy gevoel. Hierin is het ook belangrijk om goed te duiden hoe dan het algoritme werkt. Daarnaast zal naar voren moeten komen dat het indicatief is. Want het zal een klant een negatief gevoel geven als hij als fraudeur wordt gekenmerkt door een algoritme.
4	Here, it is not required to explain the relevant input parameters to the general customer. If fraud has been observed, then the insurer should explain why the come to this conclusion.
5	Fraudedetectie is een intern bedrijfsproces. Transparantie over dit interne proces vind ik geen content voor klanten. Decision output zie ik hier als uitval uit het fraudesysteem op basis waarvan de fraudedesk bepaalt, wat de volgende acties is. We kunnen als verzekeraar wel transparant zijn, dat we fraudedetectie toepassen en hier algemene knowledge base background informatie aan klanten kunnen verschaffen."
6	gevoelige toepassing. zie ook toeslagenaffaire. wellicht had de belastingdienst initieel prima indicatoren. gevaar voor reputatieschade dus. de maatschappelijke aanvaarding voor onderscheid in behandeling ligt laag. dus zo objectief mogelijk brengen en transparant zijn in hoe je gekomen bent tot. Tenslotte zou een uniforme werkwijze marktbreed, de acceptatie verhogen.
7	insight for clients how data is used

8	Hiervan vraag ik me ook af of je dit überhaupt moet willen toelichting aan de klant en ik denk zeker niet op detailniveau. Je zou als je hierin open bent wel mee kunnen geven hoe betrouwbaar een inschatting is daarmee creëer je mogelijk wel een stuk openheid en haal je de uitkomst ook meer uit de perceptie van 'we vermoeden dat jij fraude hebt gepleegd' naar 'op basis van ons algoritme heeft deze claim een verhoogde kans'
9	The purpose of using this type of modelling is different to the previous ones. Very important in this is case, is to create the balance between transparancy and protection of your own capabilities. If you be completely transparant on how we manage to detect fraud, you give away an important asset to the potential frauds.

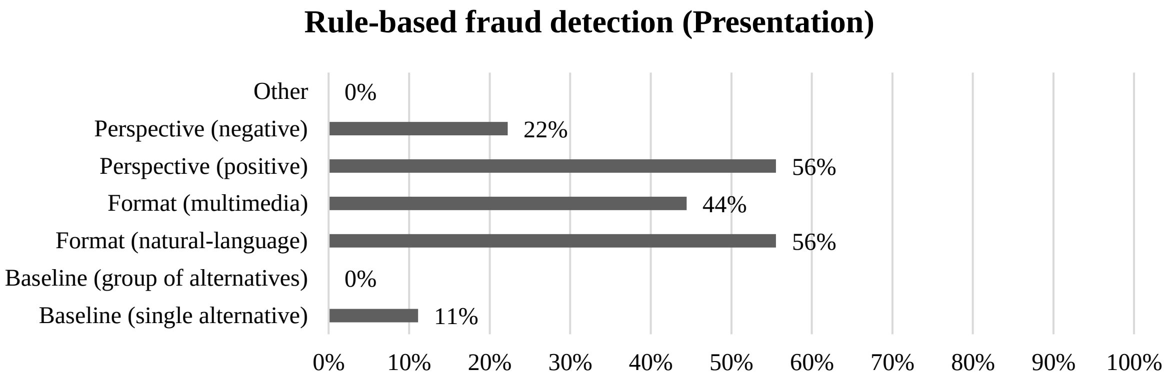

Figure 16. 1st round Delphi-study: Rule-based fraud detection (presentation) - results

Table 42. 1st round Delphi-study: Rule-based fraud detection (presentation) – substantiations

Respondent	Substantiation
1	Ook hier zou ik geen alternatieven aandragen. Dit leidt alleen maar tot ruis in de discussie die belangrijk is: is er wel of geen fraude gepleegd.
2	previous answer was lost due to technical problems
3	Je zal aan moeten geven dat het indicatief is en dat het belangrijk is voor een verzekeraar en voor de klant zelf dat misbruik wordt tegengegaan. Als je dit niet doet dan wordt een verzekering onbetaalbaar.
4	It is important to explain the fraud detection engine and the consequences of being detected as a fraud executing person.
5	Zie boven.
6	vraagt om duidelijke en stevige uitleg
7	make as clear and detailed as possible how decision is made
8	Belangrijk om in de communicatie af te zetten tegen alternatief van niet gebruiken van deze techniek waarbij gevolg dan hogere kosten zouden zijn. Gegeven gevoeligheid van dit soort zaken zou mijn voorkeur uitgaan naar persoonlijke benadering en toelichting ipv via multimedia
9	The purpose of using this type of modelling is different to the previous ones. Very important in this is case, is to create the balance between transparancy and protection of your own capabilities. If you be completely transparant on how we manage to detect fraud, you give away an important asset to the potential frauds.

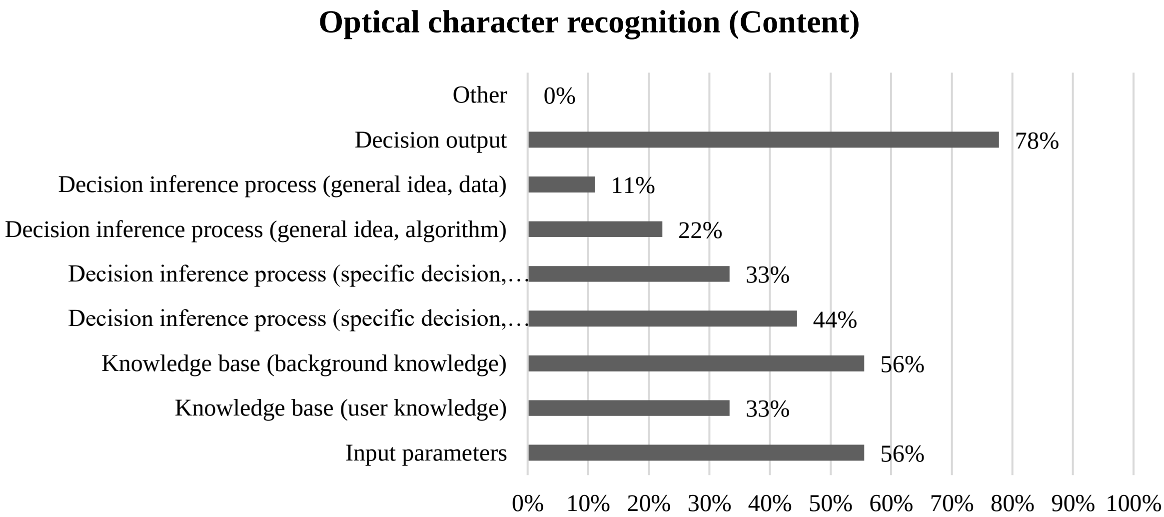

Figure 17. 1st round Delphi-study: Optical character recognition (content) - results

Table 43. 1st round Delphi-study: Optical character recognition (content) – substantiations

Respondent	Substantiation
1	Ik zou hier niet te veel ontsluiten. Belangrijkste is dat claims goed afgehandeld worden, dan is de klant tevreden. Mogelijk bij niet-standaarduitkeringen onderbouwen waarom dit zo is adhv van markt- en userkennis, aangevuld met specifieke redenering waarom dit leidt tot het uiteindelijke resultaat.
2	previous answer was lost due to technical problems
3	Als het goed is hebben klantgegevens hier geen invloed op en dat is dan ook niet relevant. Mijn inziens moet dan uitleggen hoe alles technisch in zijn werk gaat. Het is gewoon een slimme manier van technologie gebruiken. Voor het uitkeren van een schade moet dan wel het vertrouwen er zijn dat het betrouwbaar is. Dat er wel mee bepaald kan worden of er fraude wordt gepleegd zou ik minder op ingaan.
4	For the general public, only a general description of the conceptual theory are required. In a particular case, the insurer should also provide decision output and the way it has been substantiated.
5	Deze systemen ondersteunen het interne fraude, dan wel het claimproces en hiervoor volstaat m.i. algemene kennisgeving wat op dit gebied wordt gedaan, als er bijvoorbeeld additionele input (schadefoto) bij klanten wordt opgevraagd.
6	ook weer fact-based uitleggen. Maar ook uitleggen hoe het werkt en waarom we daar voor kiezen. Het heeft voor- en nadelen.
7	info to assess value of outcome
8	Ook hier zal je klant moeten kunnen uitleggen hoe je in zijn/haar specifieke geval tot een inschatting bent gekomen en wat daaraan ten grondslag ligt. Focus ligt in dit geval denk ik minder op de data maar meer op het beslisproces wat hieronder ligt maar wel toegespitst op het individuele geval.

9	to transparant, you need to share the most important parts of the information. There should be a balance between complete in your information and the overkill of information. In other words, say what the machine does but take into account the difference acknowledge level.

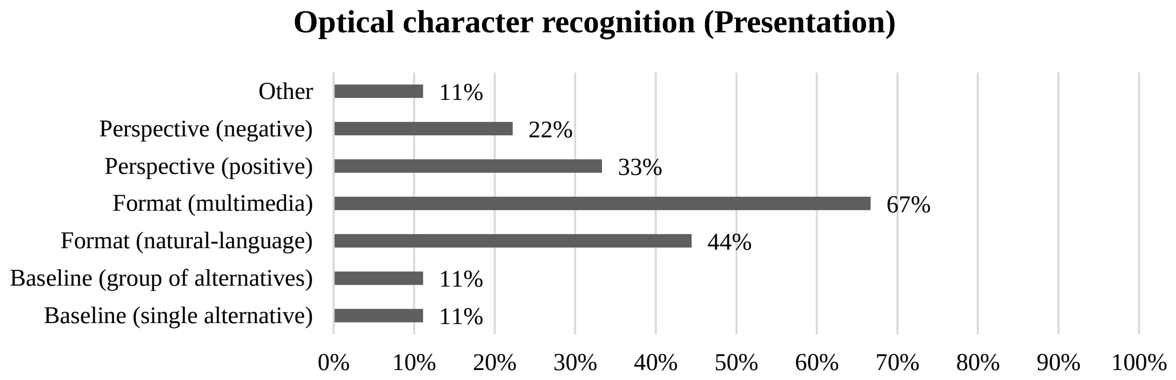

Figure 18. 1st round Delphi-study: Optical character recognition (presentation) - results

Table 44. 1st round Delphi-study: Optical character recognition (presentation) – substantiations

Respondent	Substantiation
1	Ook hier; geen alternatieven noemen, wel zorgen voor flexibiliteit in uitingen.
2	previous answer was lost due to technical problems
3	Hiervoor leent zich natuurlijk bij uitstek multimedia om te duiden hoe het proces werkt om van een plaatje naar schade inschatting te gaan.
4	Only a general description of the applied techniques.
5	Presentatie kan ik hier niet goed plaatsen
6	aangeven dat het een keus is tussen verschillende mogelijkheden. En wat het oplevert om hiervoor te kiezen.
7	easy to understand
8	Denk dat toe kan voegen om een 2nd best alternatief mee te geven om zo de klant ook indruk te geven dat er ruimte is om discussie te hebben over de uitkomst. Je wilt denk ik niet meerdere alternatieven geven om het wel overzichtelijk te houden. Gezien het specifieke karakter van de beslissing zou ook hier natural language naar mijn idee passender zijn dan multimedia.
9	to transparant, you need to share the most important parts of the information. There should be a balance between complete in your information and the overkill of information. In other words, say what the machine does but take into account the difference acknowledge level.

Appendix VI – Results: 2nd round Delphi-study

GLM and random forest

Table 45. 2nd round Delphi-study: GLM and random forest – ranking per respondent

Ranking per respondent						
Respondent	Scenario a	Scenario b	Scenario c	Scenario d	Scenario e	Scenario f
1	1	2	3	4	5	6
2	2	1	3	4	6	5
3	2	1	3	4	6	5
4	2	3	1	4	5	6
5	5	2	1	4	6	3
6	2	1	3	4	6	5
7	2	1	3	5	6	4

Table 46. 2nd round Delphi-study: GLM and random forest – mean ranks

Mean Ranks	
Scenario b	1,57
Scenario a	2,29
Scenario c	2,43
Scenario d	4,14
Scenario f	4,86
Scenario e	5,71

Table 47. 2nd round Delphi-study: GLM and random forest – test statistics

Test Statistics	
N	7
Kendall's W[a]	0,771
Chi-Square	27,000
df	5
Asymp. Sig.	0,000

Table 48. 2nd round Delphi-study: GLM and random forest – inclusion of baseline

Question 2. All scenarios include a natural-language format and are presented from a positive perspective. Should there also be a baseline (single or group of alternatives) included for comparison?	
Yes	5
No	2

Table 49. 2nd round Delphi-study: GLM and random forest – substantiations

Respondent	**Substantiation**
1	Most important are input parameters: this is the information the customer directly provided.
2	Het is moeilijk een goed beeld te krijgen wat nu precies decision output precies inhoud. Daarom lager in de prioriteit gezet. Minimaal is voor mij de parameters daarna de general idea en dan background knowledge. Dus hoogste prio dat je weet waarop iets gedaan wordt. Daarna steeds meer onderbouwing
3	Een prijs is een gegeven output voor de klant, hij kan niet kiezen, hooguit andere dekkingsopties.
4	I think all three (input, output, knowledge base and general idea behind te data) are important. Input/output is most important, then knowledge base, then general idea.
5	Key element is in my opinion being able to explain and justify the decisions made which lead to the decision output. Therefore the input parameters used and decision inference process should be explainable as a minimum. I don't see much added value in sharing background knowledge base (also don't recognize a direct link with AVG). Given that explainability of the decision is key, adding alternatives in the explanation is less relevant and might even cause more confusion.
6	Insight in the rational behind the decision and based on which data / input decisions are made is the most important goal to explain premiums to customers. It is confusing to give to much detailed information at first, however all perspectives have to be explained when asked for.
7	x

Price optimization

Table 50. 2nd round Delphi-study: Price optimization – ranking per respondent

Ranking per respondent				
Respondent	Scenario a	Scenario b	Scenario c	Scenario d
1	3	4	1	2
2	2	1	3	4
3	3	2	4	1
4	2	3	1	4
5	4	2	1	3
6	3	2	1	4
7	4	2	1	3

Table 51. 2nd round Delphi-study: Price optimization – mean ranks

Mean Ranks	
Scenario c	1,71
Scenario b	2,29
Scenario a	3,00
Scenario d	3,00

Table 52. 2nd round Delphi-study: Price optimization – test statistics

Test Statistics	
N	7
Kendall's W[a]	0,233
Chi-Square	4,886
df	3
Asymp. Sig.	0,180

Table 53. 2nd round Delphi-study: Price optimization – substantiations

Respondent	Substantiation
1	Personally I would not favor using price optimization but rather offer a premium which is fairly related to costs and risk.
2	Het moet duidelijk zijn dat het gebeurd. Maar blijft altijd een risico dat klanten ervaren dat switchen dan een goed scenario is omdat je dan een lagere prijs krijgt. Dan is er wel een probleem in de markt omdat (acquisitie) kosten dan voor iedereen gaan stijgen. Vanuit het open communiceren naar klanten optie 2 gekozen. Als je het niet wil uitleggen dan kan je het beter niet doen
3	Internal process we use, but do not need to share with customers, if we do share, then preferably just a general, procedural explanation.
4	I don't think price optimization is a future proof system. If you can't or won't explain it, you shouldn't do it, because it will get out eventually and then we'll be standing there with our pants down.
5	In principle I believe price optimization regardless of risk characteristics is not desirable. If you would use it I think it is wise to not actively explain this to customers. However if a customer would ask how his/her premium is determined you need to explain it at least in a holistic way instead of not explaining at all.
6	I don't believe price optimization is beneficial to society in general. It is not a sustainable solution to share risks. This is different from Risk based pricing, which can be an incentive towards less risky behaviour. However on the other side, from a competition based perspective it could be necessary to use price optimization. To my opinion Government should set rules (based on the general best interest for society) to which extend differentiation in premiums is allowed.
7	x

Recommendation engine

Table 54. 2nd round Delphi-study: Recommendation engine – ranking per respondent

Ranking per respondent						
Respondent	Scenario a	Scenario b	Scenario c	Scenario d	Scenario e	Scenario f
1	1	2	3	4	5	6
2	2	1	3	4	6	5
3	2	3	6	1	4	5
4	1	3	4	2	5	6
5	5	6	4	3	1	2
6	5	6	2	3	1	4
7	2	1	4	3	6	5

Table 55. 2nd round Delphi-study: Recommendation engine – mean ranks

Mean Ranks	
Scenario a	2,57
Scenario d	2,86
Scenario b	3,14
Scenario c	3,71
Scenario e	4,00
Scenario f	4,71

Table 56. 2nd round Delphi-study: Recommendation engine – test statistics

Test Statistics	
N	7
Kendall's W[a]	0,181
Chi-Square	6,347
df	5
Asymp. Sig.	0,274

Table 57. 2nd round Delphi-study: Recommendation engine – substantiations

Respondent	Substantiation
1	Input parameters > knowledge base > general idea > decision output. Transparency on which date is used is most important, then the specific knowledge (and assumptions) used."
2	Ook hier weer de parameters prioriteit. Klant moet weten waar het over gaat daarna wat verdere verdieping.
3	User knowledge base and general idea behind the data are for comfort to client of higher importance than input parameters (customer already knows) or decision output (the next best action is shown).
4	No opinion on the ranking. But I do think a recommendation engine is usefull and if it is used, a customer should be able to know how the recommendation was formed: what input parameters and personal data was used, what the general idea is and what user knowledge base was used.
5	Input parameters are in this case less relevant I think. information on how you come up with a next best action are most important to communicate, in this perspective I think the decision output (and pros and cons compared to alternatives are important to communicatie).
6	Understanding of the rational behind the recommendation is in the best interest of customer and insurer (for certain on the longer term). Insight in behaviour of the customer can lead to more awareness in making choices by the customer and as result more satisfied about those choices.
7	x

Rule-based fraud detection

Table 58. 2nd round Delphi-study: Rule-based fraud detection – ranking per respondent

Ranking per respondent						
Respondent	Scenario a	Scenario b	Scenario c	Scenario d	Scenario e	Scenario f
1	1	2	4	3	5	6
2	2	1	3	4	5	6
3	1	2	5	3	4	6
4	1	2	4	5	3	6
5	6	1	3	4	5	2
6	6	4	2	5	3	1
7	3	1	6	2	5	4

Table 59. 2nd round Delphi-study: Rule-based fraud detection – mean ranks

Mean Ranks	
Scenario b	1,86
Scenario a	2,86
Scenario d	3,71
Scenario c	3,86
Scenario e	4,29
Scenario f	4,43

Table 60. 2nd round Delphi-study: Rule-based fraud detection – test statistics

Test Statistics	
N	7
Kendall's W[a]	0,272
Chi-Square	9,531
df	5
Asymp. Sig.	0,090

Table 61. 2nd round Delphi-study: Rule-based fraud detection – substantiations

Respondent	Substantiation
1	1 and 2 are on an equal level. Most importante are input parameters, but both knowledge base and procedural decision information are important.
2	Zie mijn uitleg hiervoor. Maar decision output is op dit onderwerp minder relevant
3	Because of privacy perspective as transparent as possible in used input and knowledge base.
4	I don't think rule based fraud detection requires a great deal of transparency, as long as it is only used for raising red flags, and a human decision is made before a customer is confronted with actual legal consequences.
5	In this case I think more that 2 elements are required in the explanation. But in ordering them in terms of importance I think being transparent about the decision inference process is most important, followed by explaining decision output in general terms and the input parameters used from privacy perspective.
6	Being clear about the interpretation of the output and to state it is just an indication is very important in the acceptance that ADM systems are used for this goal. This is a perfect example where human intelligence and ADM strengthen each other. The final outcome yes/no fraud must be substantiated very well and objective.
7	x

Optical character recognition

Table 62. 2nd round Delphi-study: Optical character recognition – ranking per respondent

Ranking per respondent				
Respondent	Scenario a	Scenario b	Scenario c	Scenario d
1	1	3	2	4
2	2	3	1	4
3	2	4	1	3
4	1	3	2	4
5	3	4	1	2
6	4	3	1	2
7	3	4	1	2

Table 63. 2nd round Delphi-study: Optical character recognition – mean ranks

Mean Ranks	
Scenario c	1,29
Scenario a	2,29
Scenario d	3,00
Scenario b	3,43

Table 64. 2nd round Delphi-study: Optical character recognition – test statistics

Test Statistics	
N	7
Kendall's W[a]	0,527
Chi-Square	11,057
df	3
Asymp. Sig.	0,011

Table 65. 2nd round Delphi-study: Optical character recognition – substantiations

Respondent	Substantiation
1	Input parameters are again most important, we should be transparent on which data we use. Decision output marks least important, offering alternatives does not add value and will only lead to wrong discussions.
2	Ook hier weer de basis input parameters is het minimum. Daarna een uitleg
3	general idea seems important to explain and decision output seems less important
4	No opinion on this subject again: as long as the system is fair and doesn't have a bias towards certain groups, the information towards customers is less relevant.
5	For me the general idea behind the algorithm is most important to communicate (which only comes back in 1 of the scenarios). And if I need to choose I would say explaining decision output (comparing with alternatives) is more important than explaining the input parameters.
6	I believe this technique have to prove itself first for the settlement of claims. In the evolution towards this stage, it can be used to have a first indication of the costs. This can be communicated towards customer and 'schadeherstel bedrijf' (expert). The expert will use the indication and add his expert opinion for the determination of the exact costs. Following this procedure, the AI network will learn by more examples. Expert opinions can be compared and challenged within this system. For acceptance this ADM system it is necessary to give a brief general idea how the algorithm works and what the outputs are for specific cases (where expert opinions have the function of a feedback system)
7	x

Appendix VII – Results: 3rd round

Price optimization

Table 66. 3rd round Delphi-study: Price optimization – ranking per respondent

Ranking per respondent				
Respondent	Scenario c	Scenario b	Scenario a	Scenario d
1	1	3	4	2
2	4	2	3	1
3	4	2	1	3
4*	1	2	3	4
5*	1	2	3	4
6*	1	2	3	4
7*	1	2	3	4
8	4	1	2	3

* Concurred with ranking of round 2

Table 67. 3rd round Delphi-study: Price optimization – mean ranks

Mean Ranks	
Scenario b	2,00
Scenario c	2,13
Scenario a	2,75
Scenario d	3,13

Table 68. 3rd round Delphi-study: Price optimization – test statistics

Test Statistics	
N	8
Kendall's W[a]	0,169
Chi-Square	4,050
df	3
Asymp. Sig.	0,256

Table 69. 3rd round Delphi-study: Price optimization – substantiations

Respondent	Substantiation
1	I feel that if you chose to employ price optimization, you should not explain its underlying mechanisms to the customers. If a customer understands which information is used in price optimization, it will be less effective as the customer can act upon it to achieve a better price.
2	Ik zie niet in, waarom wij als verzekeraar geen prijsoptimalisatie zouden willen toepassen. Dus ben het niet eens met de ranking, waarin dit bovenaan staat.
3	Prijs optimalisatie moet mogelijk zijn mede omdat je je economisch in de voet schiet maar moet dan wel uitgelegd worden als klanten er naar vragen. Dus transparantie is vereist. Het argument dat het een intern proces is, daar ben ik niet mee eens. Namelijk de output = de prijs gaat naar de klant. Dat is mijn inziens niet intern.
4	Using price optimization is a strategic choice, however can be questioned from an ethical perspective. Does it add value to the customers in the long run and if not it is not a sustainable strategy from an Insurers perspective. I believe Insurers have to be transparent if they use this strategy. That is why I concur with the ranking as described.
5	
6	
7	
8	not using price optimization may be the best choice, but as long as it is used, it should be explained as best as possible. that is why I chose b/a/d/c: c is hypothetical.

Recommendation engine

Table 70. 3rd round Delphi-study: Recommendation engine – ranking per respondent

Ranking per respondent						
Respondent	Scenario a	Scenario d	Scenario b	Scenario c	Scenario e	Scenario f
1*	1	2	3	4	5	6
2	3	1	2	6	4	5
3*	1	2	3	4	5	6
4	6	4	5	3	2	1
5	1	4	2	3	5	6
6	5	1	3	6	4	2
7*	1	2	3	4	5	6
8*	1	2	3	4	5	6

* Concurred with ranking of round 2

Table 71. 3rd round Delphi-study: Recommendation engine – mean ranks

Mean Ranks	
Scenario d	2,25
Scenario a	2,38
Scenario b	3,00
Scenario c	4,25
Scenario e	4,38
Scenario f	4,75

Table 72. 3rd round Delphi-study: Recommendation engine – test statistics

Test Statistics	
N	8
Kendall's W[a]	0,341
Chi-Square	13,643
df	5
Asymp. Sig.	0,018

Table 73. 3rd round Delphi-study: Recommendation engine – substantiations

Respondent	Substantiation
1	
2	Comfort about the outcome from this specific ADM is important. Clients need to feel comfortable with the advised next best action and therefore need to understand why the ADM came with this outcome given the user knowledge base, input parameters and general idea behind the data. Decision output and showing several alternatives is less obvious, because these are not the next best action. The general purpose of the ADM is not to choose, but simply one answer.
3	
4	Understanding the advise and knowing the best action for is the most important for the customer. Based on the minority opinions the above order follows.
5	altijd de input om aan te refereren
6	Blijf van mening dat input parameters in deze minder relevant zijn dan communiceren hoe je op basis van de data tot de next best action komt.
7	
8	

Rule-based fraud detection

Table 74. 3rd round Delphi-study: Rule-based fraud detection – ranking per respondent

Ranking per respondent						
Respondent	Scenario b	Scenario a	Scenario d	Scenario c	Scenario e	Scenario f
1	2	1	3	4	5	6
2*	1	2	3	4	5	6
3*	1	2	3	4	5	6
4	5	6	3	4	2	1
5	1	2	4	3	5	6
6*	1	2	3	4	5	6
7*	1	2	3	4	5	6
8*	1	2	3	4	5	6

* Concurred with ranking of round 2

Table 75. 3rd round Delphi-study: Rule-based fraud detection – mean ranks

Mean Ranks	
Scenario b	1,63
Scenario a	2,38
Scenario d	3,13
Scenario c	3,88
Scenario e	4,63
Scenario f	5,38

Table 76. 3rd round Delphi-study: Rule-based fraud detection – test statistics

Test Statistics	
N	8
Kendall's W[a]	0,563
Chi-Square	22,500
df	5
Asymp. Sig.	0,000

Table 77. 3rd round Delphi-study: Rule-based fraud detection – substantiations

Respondent	Substantiation
1	I feel that most important is to disclose any knowledge we have on the customer.
2	
3	
4	Transparancy about procedure and motivated output as indication of potential fraude is most important
5	ook hier weer altijd refereren aan de input
6	I feel that most important is to disclose any knowledge we have on the customer.
7	
8	

Optical character recognition

Table 78. 3rd round Delphi-study: Optical character recognition – ranking per respondent

Ranking per respondent					
Respondent	Scenario c	Scenario a	Scenario d	Scenario b	Scenario e
1	2	1	4	3	5
2	5	4	2	3	1
3	1	3	4	5	2
4	2	5	3	4	1
5	1	2	5	3	4
6	3	5	2	4	1
7	1	5	4	2	3
8	5	3	2	1	4

Table 79. 3rd round Delphi-study: Optical character recognition – mean ranks

Mean Ranks	
Scenario c	2,50
Scenario e	2,63
Scenario b	3,13
Scenario d	3,25
Scenario a	3,50

Table 80. 3rd round Delphi-study: Optical character recognition – test statistics

Test Statistics	
N	8
Kendall's W[a]	0,072
Chi-Square	2,300
df	4
Asymp. Sig.	0,681

Table 81. 3rd round Delphi-study: Optical character recognition – substantiations

Respondent	Substantiation
1	I would not disclose any information on alternatives (decision output), but start with data on the customer.
2	I agree with minority opinion and alternative scenario.
3	Output alternatieven zijn in dit scenario relevanter omdat het minder binair is en er zijn daadwerkelijk andere mogelijke alternatieven. Zeker bij een schade kan het nogal uitmaken of de bumper kapot is en/of ook een deur oid. Dan is het wel relevant ook voor de klant wat andere inschattingen kunnen zijn.
4	Feedback loop is important
5	idem
6	Input parameters less relevant for this scenarioz focus should be on explaining decisions en general idea on how to come there
7	Decision output is relevant for the customer which car was damaged, in order to explain the amount of money to receive.
8	I don't think the order of these scenario's is particularly relevant: output is probably more relevant, but the customer will have to be able check whether the input was correct (and has a right to do so under GDPR). So all information may be relevant.

Appendix VIII – Questionnaire on types of ADM-systems

*** Required Information**

Dear Participant,

Welcome to my MSc Thesis Research!

I highly appreciate your time and effort for participating in this survey. The focus of this research **is to understand what explanations of Algorithmic Decision Making (ADM) systems are preferred in the Dutch P&C insurance industry. Part of the research is this questionnaire which focus is to understand what ADM-systems are used in de Dutch P&C insurance industry.**

This survey will take approximately 15 minutes. It should be emphasized that your responses will be kept confidential and used only for research purposes. Your participation in this study is voluntary. In case you decide in the future that you do not want to participate anymore, you can drop out at any moment.

What's in it for YOU? By taking this survey you can gain new insights about algorithmic decision-making.

It is of utmost importance that you answer honestly and to the best of your knowledge.

Finally, the questions will be asked in English, but your answers can be in Dutch or English, whichever you prefer.

Thank you for your cooperation,

Eric Schotman

Page 1 of 13

On the following 2 pages, you first will be asked some questions about your educational level, current occupation and knowledge level of ADM-systems.

*** 1. What is your age? (Select one option)**

- Under 20
- 20 - 34
- 35 - 49
- 50 - 65
- 65 or older

*** 2.**
What is the highest level of education you have completed? (Select one option)

- Primary school
- Pre-vocational secondary education (VMBO)
- Senior general secondary education (HAVO)
- Pre-university education (VWO)
- Secondary vocational education (MBO)
- Higher professional education (HBO)
- Associate Degree
- Bachelor Degree
- Master's Degree
- Professional Doctorate Degree
- Academic Doctorate Degree

*** 3.**
What type of organization do you currently work for? (Select one option)

- P&C insurer
- Other insurer
- Regulator or supervisor
- Industry association
- Financial services consultancy
- Other (Please specify) ___________

*** 4. What is the size, in gross written premium (GWP), of the P&C insurer you work for? (Select one option)**

- ☐ Bigger than 500 million GWP
- ☐ Smaller than 500 million GWP
- ☐ Other (Please specify) ___________

*** 5.**
Which of the following best describes your current occupation? (Select one option)

- ☐ Policy maker
- ☐ Management
- ☐ Privacy-expert
- ☐ Data analyst/scientist
- ☐ Actuary
- ☐ Other (Please specify) ___________

*** 6. How would you assess your current knowledge level on the use of Algorithmic Decision Making systems in the insurance industry? (Select one option)**

0	1	2	3	4	5	6	7	8	9	10

*** 7. Which of the choices below describe a rule-based model: (Select one option)**

- Rule-based models refer to every model that generate statistical rules.
- Rule-based models mimic the reasoning of a human expert in solving a knowledge-intensive problem.
- Rule-based models mimic the functioning of the brain by generating rules.

*** 8. Which of the choices below describe a statistical probabilistic model: (Select one option)**

- Statistical probabilistic models are statistical models which mimic brain functions.
- Statistical probabilistic models refer to statistical models whose rules give a probability distribution as a solution.
- Statistical probabilistic models refer to probabilistic models whose links represent the conditional dependencies between a set of variables.

*** 9. Which of the choices below describe an artificial neural network: (Select one option)**

- Artificial neural networks try to simulate the atomic processes of the human brain.
- Artificial neural networks create rules to simulate the human brain.
- Artificial neural networks refer to statistical models that simulate the atomic processes of the human brain.

10. Which statements are true for a rule-based model:

	True	**False**	**Don't know**
*(a) Rule-based models are relatively interpretable models.			
*(b) Rule-based systems are also known as production systems or expert systems.			
*(c) Rule-based models are implicit.			

11. Which statements are true for a statistical probabilistic model:

	True	False	Don't know
*(a) Statistical probabilistic models present the relationships between features and target.	☐	☐	☐
*(b) Statistical probabilistic models are explicit.	☐	☐	☐
*(c) Statistical probabilistic models are implicit.	☐	☐	☐

12. Which statements are true for artificial neural networks:

	True	False	Don't know
*(a) Artificial neural networks can never be 100% predictable, error-free or explainable.	☐	☐	☐
*(b) Artificial neural networks are implicit.	☐	☐	☐
*(c) Artificial neural networks are explicit.	☐	☐	☐

On the following pages, you will be introduced to the definition of ADM-systems and the different types. Afterwards, a few questions will be asked on your perception of the use of ADM-systems within the P&C insurer that you work for. Of course, there are no right or wrong answers here.

Your perception will be asked for the following parts of the P&C insurance value-chain:

- **Product development;**
- **Pricing and underwriting;**
- **Sales and distribution;**
- **Post-sales services and assistance;**
- **Fraud and claims management.**

Please read the provided information carefully!

Much of the decisions which were historically made by humans are now made by algorithms. Decisions on prioritization, classification, association, and filtering are made by these so-called Algorithmic Decision Making systems. Insurance is one of the industries where ADM-systems are used throughout the value chain.

Three types of models* are distinguished:

1. **Deductive and rule-based systems (such as decision trees);**
2. **Statistical probabilistic models (such as Bayesian networks);**
3. **Artificial neural networks (such as multi-layer perceptrons).**

***A more comprehensive and detailed distinction between different models could be made. Because of the complex character of these models and the variety of definitions, these three types are followed.**

PRODUCT DEVELOPMENT

*** 13. Do you use ADM-system(s) in product development of P&C insurance products? (Select one option)**

- ☐ Yes
- ☐ No
- ☐ Unknown

14. What ADM-system(s) do you use in product development of P&C insurance products? Also, indicate whether the(se) system(s) are machine learning system(s) or traditional system(s).

	Used (traditional)	Used (machine learning)	Used (combination of machine learning and traditional)	Not used	Unknown
*(a) Deductive and rule-based systems	☐	☐	☐	☐	☐
*(b) Statistical probabilistic models	☐	☐	☐	☐	☐
*(c) Artificial neural networks	☐	☐	☐	☐	☐

*** 15. Please tell us what type(s) of algorithm(s) are used in the(se) ADM-system(s) and with what goal?**

PRICING & UNDERWRITING

* **16. Do you use ADM-system(s) in pricing and underwriting of P&C insurance products? (Select one option)**

- ☐ Yes
- ☐ No
- ☐ Unknown

17. What ADM-system(s) do you use in pricing and underwriting of P&C insurance products? Also, indicate whether the(se) system(s) are machine learning system(s) or traditional system(s).

	Used (traditional)	Used (machine learning)	Used (combination of machine learning and traditional)	Not used	Unknown
*(a) Deductive and rule-based systems	☐	☐	☐	☐	☐
*(b) Statistical probabilistic models	☐	☐	☐	☐	☐
*(c) Artificial neural networks	☐	☐	☐	☐	☐

* **18. Please tell us what type(s) of algorithm(s) are used in the(se) ADM-system(s) and with what goal?**

SALES & DISTRIBUTION

* 19. Do you use ADM-system(s) in the sales and distribution of P&C insurance products? (Select one option)

- ☐ Yes
- ☐ No
- ☐ Unknown

20. What ADM-system(s) do you use in the sales and distribution of P&C insurance products? Also, indicate whether the(se) system(s) are machine learning system(s) or traditional system(s).

	Used (traditional)	Used (machine learning)	Used (combination of machine learning and traditional)	Not used	Unknown
*(a) Deductive and rule-based systems	☐	☐	☐	☐	☐
*(b) Statistical probabilistic models	☐	☐	☐	☐	☐
*(c) Artificial neural networks	☐	☐	☐	☐	☐

* **21. Please tell us what type(s) of algorithm(s) are used in the(se) ADM-system(s) and with what goal?**

POST-SALES SERVICES & ASSISTANCE

*** 22. Do you use ADM-system(s) in post-sales services and assistance for P&C insurance products? (Select one option)**

- ☐ Yes
- ☐ No
- ☐ Unknown

23. What ADM-system(s) do you use in post-sales services and assistance for P&C insurance products? Also, indicate whether the(se) system(s) are machine learning system(s) or traditional system(s).

	Used (traditional)	Used (machine learning)	Used (combination of machine learning and traditional)	Not used	Unknown
*(a) Deductive and rule-based systems	☐	☐	☐	☐	☐
*(b) Statistical probabilistic models	☐	☐	☐	☐	☐
*(c) Artificial neural networks	☐	☐	☐	☐	☐

*** 24. Please tell us what type(s) of algorithm(s) are used in the(se) ADM-system(s) and with what goal?**

FRAUD & CLAIMS MANAGEMENT

*** 25. Do you use ADM-system(s) in fraud and claims management of P&C insurance products? (Select one option)**

- ☐ Yes
- ☐ No
- ☐ Unknown

26. What ADM-system(s) do you use in fraud and claims management of P&C insurance products? Also, indicate whether the(se) system(s) are machine learning system(s) or traditional system(s).

	Used (traditional)	Used (machine learning)	Used (combination of machine learning and traditional)	Not used	Unknown
*(a) Deductive and rule-based systems	☐	☐	☐	☐	☐
*(b) Statistical probabilistic models	☐	☐	☐	☐	☐
*(c) Artificial neural networks	☐	☐	☐	☐	☐

*** 27. Please tell us what type(s) of algorithm(s) are used in the(se) ADM-system(s) and with what goal?**

* 28. Do you have any other thoughts about ADM-systems in the P&C insurance industry that you would like to share?

Part of the follow-up research is a 'Delphi-study' which is a method to submit a complex problem to a group of experts. If you are open to participating in this panel of experts please contact me on Linkedin or via email (schotman.eric@gmail.com).

If you are interested in the results of the thesis research, please also contact me on Linkedin or via email (schotman.eric@gmail.com).

We thank you for your time spent taking this survey.

Your response has been recorded.

Appendix IX – 1st round Delphi-study

*** Required Information**

Dear Participant,

Welcome to my MSc Thesis Research!

I highly appreciate your time and effort in participating in this expert panel. The focus of this research is to understand what explanations, towards customers, of Algorithmic Decision Making (ADM) systems, are preferred in the Dutch P&C insurance industry.

To do this we will use the Delphi-method. This is a method for structuring a group communication process so that the process is effective in allowing a group of individuals, as a whole, to deal with a complex problem. The Delphi-method consists of 2 to 3 iterative rounds of surveys within a time period of a few weeks, depending on the number of rounds required to reach consensus. This first round will take approximately 30 minutes to participate in. The following two rounds will take around 10 minutes.

It should be emphasized that your responses will be kept confidential and used only for research purposes. Your participation in this study is voluntary. In case you decide in the future that you do not want to participate anymore, you can drop out at any moment.

What's in it for YOU? By taking part in this study you can gain new insights about algorithmic decision-making and their explainability.

It is of utmost importance to me that you answer honestly and to the best of your knowledge.

Thank you for your cooperation,

Eric Schotman

Page 1 of 11

Much of the decisions which were historically made by humans are now made by algorithms. Decisions on prioritization, classification, association, and filtering are made by these so-called Algorithmic Decision Making systems. Insurance is one of the industries where ADM-systems are used throughout the value chain.

On the following 2 pages, you will be introduced to a few examples of ADM-systems used by P&C insurers and theory on explainability.

Afterward, you will be questioned on your perception of how these ADM-systems are best explained to customers from the perspective of the insurer.

Please read the provided information carefully!

Page 2 of 11

Based on the survey results, and a thematic review on big data analytics performed by EIOPA, the following ADM-systems are selected.

Pricing and underwriting

- **GLM and random forest - Generalized linear models (GLM) used for pricing. To enhance these GLM's, machine learning random forest models are used in the background. These random forest models combine different decision trees to obtain an aggregated prediction/regression.**
- **Price optimization - Machine learning churn models for price optimization purposes. Price optimization refers to the practice of adjusting the premiums, paid by different groups of consumers, based on the behaviours and economic characteristics of the consumer, in ways unrelated to their risk or cost.**

Sales and distribution

- **Recommendation engine - A recommendation engine used for the 'next best action'. This is used to evaluate the customer's past behaviour, recent actions and needs to deliver the right message, at the right time, and via the right channel.**

Fraud and claims management

- **Rule-based fraud detection - A rule-based model used to assess claims and evaluate whether they present anomalies and flag potentially fraudulent claims.**
- **Optical character recognition (OCR) - Deep learning networks (Artificial Neural Networks) used to extract information from scanned documents such as images from damaged cars to estimate repair costs.**

During this Delphi-study we will determine the preferred explanation towards customers, for these ADM-systems, from the perspective of the insurer.

When determining an explanation for an ADM-system there are different aspects to be considered. For this study, 'explainability' is considered to have 'content' and 'presentation' elements.

Regarding 'content' we consider four types of information that can be included in the explanation:

- **Input parameters refer to the inputs that were provided for a particular decision problem. Such as the car registration number for motor insurance.**
- **Knowledge base information resides in the knowledge base of an ADM-system. The provided information can be:**
  - **user knowledge (tailored to the specific user) or**
  - **background knowledge (selected independent of the current user).**
- **Decision inference process information is related to the internal process of the ADM-system. This can refer to the:**
  - **Specific decision, which can be:**
    - **Procedural (i.e. describe the steps taken to reach a decision).**
    - **Declarative (such as the confidence in the decision).**
  - **General idea behind the**
    - **Algorithm (e.g. recommendation of alternatives that similar users like);**
    - **Data (e.g. use of users' driving behavior to set the premium discount).**
- **Decision output refers to the decision reasoning outcome and, for example, describes the particular features and feature values of the recommended and non-recommended alternatives. An example of this is when the pros and cons of the alternatives are explained.**

Looking at the 'presentation' element we consider three facets:

- **Baselines: The baseline for comparison can be**
  - **a single alternative to that recommended or**
  - **a group of alternatives.**
- **Formats: Different output formats can be chosen such as:**
  - **Natural language or**
  - **Different types of multimedia (i.e. audio, images or film)**
- **Perspective: The perspective in which an explanation is presented can be either:**
  - **Positive (why an alternative is suitable for a user) or**
  - **Negative (why certain negative aspects of an alternative could be acceptable).**

On the following pages, a few questions will be asked on your perception of how these ADM-systems are best explained, to customers, from the perspective of the insurer.

Pricing and underwriting (GLM and random forest)

Please select which 'content' and 'presentation' elements should be included, in the explanation towards customers, for the following ADM-system within the process of pricing and underwriting:

<u>GLM and random forest</u> - Generalized linear models (GLM) used for pricing. To enhance these GLM's, machine learning random forest models are used in the background. These random forest models combine different decision trees to obtain an aggregated prediction/regression.

When you would like to review the previously described theory on the elements of an explanation, please click on the following <u>Link</u>

*** 1. Content [Please select at most 6 options.]**

- ☐ Input parameters
- ☐ Knowledge base (user knowledge)
- ☐ Knowledge base (background knowledge)
- ☐ Decision inference process (specific decision, procedural)
- ☐ Decision inference process (specific decision, declarative)
- ☐ Decision inference process (general idea, algorithm)
- ☐ Decision inference process (general idea, data)
- ☐ Decision output
- ☐ Other (Please specify) ________________

*** 2. Please elaborate on your 'content' choice of explanation-elements.**

*** 3. Presentation [Please select at most 4 options.]**

- [ ] Baseline (single alternative)
- [ ] Baseline (group of alternatives)
- [ ] Format (natural-language)
- [ ] Format (multimedia)
- [ ] Perspective (positive)
- [ ] Perspective (negative)
- [ ] Other (Please specify) _______________

* 4. Please elaborate on your 'presentation' choice of explanation-elements.

Page 6 of 11

Pricing and underwriting (Price optimization)

Please select which 'content' and 'presentation' elements should be included, in the explanation towards customers, for the following ADM-system within the process of pricing and underwriting:

<u>Price optimization</u> - Machine learning churn models for price optimization purposes. Price optimization refers to the practice of adjusting the premiums, paid by different groups of consumers, based on the behaviours and economic characteristics of the consumer, in ways unrelated to their risk or cost.

When you would like to review the previously described theory on the elements of an explanation, please click on the following Link

*** 5. Content [Please select at most 6 options.]**

- ☐ Input parameters
- ☐ Knowledge base (user knowledge)
- ☐ Knowledge base (background knowledge)
- ☐ Decision inference process (specific decision, procedural)
- ☐ Decision inference process (specific decision, declarative)
- ☐ Decision inference process (general idea, algorithm)
- ☐ Decision inference process (general idea, data)
- ☐ Decision output
- ☐ Other (Please specify) ________________

*** 6. Please elaborate on your 'content' choice of explanation-elements.**

*** 7. Presentation [Please select at most 4 options.]**

- [ ] Baseline (single alternative)
- [ ] Baseline (group of alternatives)
- [ ] Format (natural-language)
- [ ] Format (multimedia)
- [ ] Perspective (positive)
- [ ] Perspective (negative)
- [ ] Other (Please specify) _______________

* 8. Please elaborate on your 'presentation' choice of explanation-elements.

Sales and distribution (Recommendation engine)

Please select which 'content' and 'presentation' elements should be included, in the explanation towards customers, for the following ADM-system within the process of sales and distribution:

Recommendation engine - A recommendation engine used for the 'next best action'. This is used to evaluate the customer's past behaviour, recent actions and needs to deliver the right message, at the right time, and via the right channel.

When you would like to review the previously described theory on the elements of an explanation, please click on the following Link

*** 9. Content [Please select at most 6 options.]**

- [ ] Input parameters
- [ ] Knowledge base (user knowledge)
- [ ] Knowledge base (background knowledge)
- [ ] Decision inference process (specific decision, procedural)
- [ ] Decision inference process (specific decision, declarative)
- [ ] Decision inference process (general idea, algorithm)
- [ ] Decision inference process (general idea, data)
- [ ] Decision output
- [ ] Other (Please specify) ________________

*** 10. Please elaborate on your 'content' choice of explanation-elements.**

*** 11. Presentation [Please select at most 4 options.]**

- ☐ Baseline (single alternative)
- ☐ Baseline (group of alternatives)
- ☐ Format (natural-language)
- ☐ Format (multimedia)
- ☐ Perspective (positive)
- ☐ Perspective (negative)
- ☐ Other (Please specify) ________________

* 12. Please elaborate on your 'presentation' choice of explanation-elements.

Page 8 of 11

Fraud and claims management (Rule-based fraud detection)

Please select which 'content' and 'presentation' elements should be included, in the explanation towards customers, for the following ADM-system within the process of fraud and claims management:

Rule-based fraud detection - A rule-based model used to assess claims and evaluate whether they present anomalies and flag potentially fraudulent claims

When you would like to review the previously described theory on the elements of an explanation, please click on the following Link

*** 13. Content [Please select at most 6 options.]**

- Input parameters
- Knowledge base (user knowledge)
- Knowledge base (background knowledge)
- Decision inference process (specific decision, procedural)
- Decision inference process (specific decision, declarative)
- Decision inference process (general idea, algorithm)
- Decision inference process (general idea, data)
- Decision output
- Other (Please specify) ________________

* 14. Please elaborate on your 'content' choice of explanation-elements.

* 15. Presentation [Please select at most 4 options.]

- ☐ Baseline (single alternative)
- ☐ Baseline (group of alternatives)
- ☐ Format (natural-language)
- ☐ Format (multimedia)
- ☐ Perspective (positive)
- ☐ Perspective (negative)
- ☐ Other (Please specify) ________________

* 16. Please elaborate on your 'presentation' choice of explanation-elements.

Page 9 of 11

Fraud and claims management (Optical character recognition)

Please select which 'content' and 'presentation' elements should be included, in the explanation towards customers, for the following ADM-system within the process of fraud and claims management:

Optical character recognition (OCR) - Deep learning networks (Artificial Neural Networks) used to extract information from scanned documents such as images from damaged cars to estimate repair costs.

When you would like to review the previously described theory on the elements of an explanation, please click on the following Link

*** 17. Content [Please select at most 6 options.]**

- ☐ Input parameters
- ☐ Knowledge base (user knowledge)
- ☐ Knowledge base (background knowledge)
- ☐ Decision inference process (specific decision, procedural)
- ☐ Decision inference process (specific decision, declarative)
- ☐ Decision inference process (general idea, algorithm)
- ☐ Decision inference process (general idea, data)
- ☐ Decision output
- ☐ Other (Please specify) ________________

*** 18. Please elaborate on your 'content' choice of explanation-elements.**

*** 19. Presentation [Please select at most 4 options.]**

- [ ] Baseline (single alternative)
- [ ] Baseline (group of alternatives)
- [ ] Format (natural-language)
- [ ] Format (multimedia)
- [ ] Perspective (positive)
- [ ] Perspective (negative)
- [ ] Other (Please specify) ________________

* 20. Please elaborate on your 'presentation' choice of explanation-elements.

We thank you for your time spent participating in this study.

Your response has been recorded. After the results have been analyzed you will be invited to participate in the second round.

Appendix X – 2nd round Delphi-study

*** Required Information**

Dear Participant,

Welcome to my MSc Thesis Research!

I highly appreciate your time and effort in participating in this expert panel. The focus of this research is to understand what explanations, towards customers, of Algorithmic Decision Making (ADM) systems, are preferred in the Dutch P&C insurance industry.

To do this we will use the Delphi-method. This is a method for structuring a group communication process so that the process is effective in allowing a group of individuals, as a whole, to deal with a complex problem. The Delphi-method consists of 2 to 3 iterative rounds of surveys within a time period of a few weeks, depending on the number of rounds required to reach consensus. This round will take approximately 15 minutes to participate in.

It should be emphasized that your responses will be kept confidential and used only for research purposes. Your participation in this study is voluntary. In case you decide in the future that you do not want to participate anymore, you can drop out at any moment.

What's in it for YOU? By taking part in this study you can gain new insights about algorithmic decision-making and their explainability.

It is of utmost importance to me that you answer honestly and to the best of your knowledge.

Thank you for your cooperation,

Eric Schotman

Page 1 of 8

In the first round of the Delphi-study, each respondent assessed which 'content' and 'presentation' elements should be included in the explanation, towards customers from the perspective of the insurer. This was assessed for 5 ADM-systems. Based on the results of the first round the most cited explanation-elements are combined to form different scenarios of explanations towards customers.

In this second round, you will be presented with these different scenarios and will be asked to rank the scenarios based on preferability from the perspective of the insurer.

The presented substantiations for explanation elements are a summary outcome of the first round. These are neither facts nor the researcher's opinion, but rather the opinions of the participants.

Pricing and underwriting (GLM and random forest)

Please rank the different scenarios based on which you prefer as an explanation towards customers, from the perspective of the insurer, for the following ADM-system within the process of pricing and underwriting:

GLM and random forest - Generalized linear models (GLM) used for pricing. To enhance these GLM's, machine learning random forest models are used in the background. These random forest models combine different decision trees to obtain an aggregated prediction/regression.

When you would like to review the theory on the elements of an explanation, please click on the following Link

Before you are asked to rank the different scenarios, please read the substantiations from the first round carefully.

Substantiation of content elements:

- **Input parameters are required by GDPR/AVG regulation. Also, customers need to understand the rationale behind there premium; this includes the input parameters.**
- **Background knowledge base includes which market knowledge is used as a starting point. This should be included based on the GDPR/AVG regulation. Also, this is important to make customers understand the rationale behind there premiums.**
- **General idea behind the data includes the substantiation of the used pricing variables. It should explain why the data is used and why this is objectively justifiable.**

Substantiation of why a baseline should or should not be included for comparison:

- **It should be included because it is required to present a baseline from a model perspective. So present the preferred solution for the customer and possible alternatives. Other elements are less important to explain the model.**
- **It should be included so that the customer has options to choose from.**
- **It should not be included, because this will lead to more questions.**
- **It should not be included, because it is of less importance than that the output is explainable.**

*** 1. Please rank the different scenarios based on which you prefer as an explanation towards customers, from the perspective of the insurer.**

Scenario 1: The input parameters and background knowledge base are included in the explanation. : ________

Scenario 2: The input parameters and general idea behind the data (decision inference process) are included in the explanation. : ________

Scenario 3: The input parameters and decision output are included in the explanation. : ________

Scenario 4: The background knowledge base and the general idea behind the data (decision inference process) are included in the explanation. : ________

Scenario 5: The background knowledge base and decision output are included in the explanation. : ________

Scenario 6: The general idea behind the data (decision inference process) and decision output are included in the explanation. : ________

2. All scenarios include a natural-language format and are presented from a positive perspective. Should there also be a baseline (single or group of alternatives) included for comparison?

(Select one option)

☐ Yes

☐ No

*** 3. Please elaborate on your chosen order of scenarios.**

____

____

Page 3 of 8

Pricing and underwriting (Price optimization)

Please rank the different scenarios based on which you prefer as an explanation towards customers, from the perspective of the insurer, for the following ADM-system within the process of pricing and underwriting:

Price optimization - Machine learning churn models for price optimization purposes. Price optimization refers to the practice of adjusting the premiums, paid by different groups of consumers, based on the behaviors and economic characteristics of the consumer, in ways unrelated to their risk or cost.

When you would like to review the theory on the elements of an explanation, please click on the following Link

Before you are asked to rank the different scenarios, please read the substantiations from the first round carefully.

Substantiation of the general idea behind the algorithm:

- **An explanation should be clear on what determines the premium. But for this ADM-system, it is more pragmatic to mention it not explicitly.**
- **Transparency on the internal processes is no content for customers (The baker does not share his secret baking recipe). We can, however, be transparent on that we use price optimization.**
- **A more procedural explanation seems to be more fitting than an explanation on an individual basis. An example explanation could be 'we kijken naar een passende prijs voor verschillende klantgroepen op basis van consumentengedrag' instead of 'vanwege uw salaris valt u in de categorie die bereid is een hogere premie te betalen dus rekenen we die'.**

Substantiation of whether price optimization should be used:

- **No opinion on whether it should be used or not, but if this ADM-system will be made even a little transparent it loses its function with rational customer behavior. Besides, it conflicts with the explainability of risk-based-pricing. A non-risk related premium surcharge invalidates the explainability of risk-based pricing.**
- **It should not be used, because it is legally complex to differentiate based on factors other than risk variables. Also, it is socially not acceptable.**
- **Should not be used, simply because we should not want to use is.**

* **4. Please rank the different scenarios based on which you prefer as an explanation towards customers, from the perspective of the insurer.**

Scenario 1: The input parameters and decision output are included in the explanation. This is explained from a positive perspective. : ________

Scenario 2: The general idea behind the algorithm (decision inference process) is included in the explanation. This is explained from a positive perspective. : ________

Scenario 3: Price optimization should not be used and therefore not explained. : ________

Scenario 4: Price optimization can be used but should not be explained towards customers. : ________

* **5. Please elaborate on your chosen order of scenarios.**

Sales and distribution (Recommendation engine)

Please rank the different scenarios based on which you prefer as an explanation towards customers, from the perspective of the insurer, for the following ADM-system within the process of sales and distribution:

Recommendation engine - A recommendation engine used for the 'next best action'. This is used to evaluate the customer's past behavior, recent actions and needs to deliver the right message, at the right time, and via the right channel.

When you would like to review the theory on the elements of an explanation, please click on the following Link

Before you are asked to rank the different scenarios, please read the substantiations from the first round carefully.

Substantiation of Content elements:

- **User knowledge base describes and communicates which personal information is used. This is a useful rationale to communicate the effectiveness of the recommendation.**
- **General idea behind the data explains the importance of the customer's past behavior on the recommendation. Also, transparency on the general idea behind the data, and how it results in the recommendation, will lead to more comfort for customers with the recommendation. It will also help them to understand why the recommendation is the right choice. This enables them to make a well-founded choice.**

All scenarios include a group of alternatives as a baseline, are presented in a multimedia format, and from a positive perspective.

* **6. Please rank the different scenarios based on which you prefer as an explanation towards customers, from the perspective of the insurer.**

Scenario 1: The input parameters and user knowledge base are included in the explanation. : ________

Scenario 2: The input parameters and general idea behind the data (decision inference process) are included in the explanation. : ________

Scenario 3: The input parameters and decision output are included in the explanation. : ________

Scenario 4: The user knowledge base and the general idea behind the data (decision inference process) are included in the explanation. : ________

Scenario 5: The user knowledge base and decision output are included in the explanation. : ________

Scenario 6: The general idea behind the data (decision inference process) and decision output are included in the explanation. : ________

* **7. Please elaborate on your chosen order of scenarios.**

Fraud and claims management (Rule-based fraud detection)

Please rank the different scenarios based on which you prefer as an explanation towards customers, from the perspective of the insurer, for the following ADM-system within the process of fraud and claims management:

Rule-based fraud detection - A rule-based model used to assess claims and evaluate whether they present anomalies and flag potentially fraudulent claims.

When you would like to review the theory on the elements of an explanation, please click on the following Link

Before you are asked to rank the different scenarios, please read the substantiations from the first round carefully.

Substantiation of Content elements:

- **Input parameters regarding the identification of anomalies should be included, because of the privacy perspective.**
- **Knowledge base (both user and background) information regarding the identification of anomalies should be included, because of the privacy perspective.**
- **Specific procedural decision information is included because of the sensitivity of the application. The social acceptance of differentiation in treatment is very low (e.g. the 'toeslagenaffaire' at the Tax Authority), which could lead to reputational damage. Therefore, it is important to be objective and transparent in how the specific decision is made.**
- **Decision output includes that the output is indicative because when a customer is marked as a fraudster by an algorithm, it will be perceived negatively. It also includes the consequences of being identified as a fraudster.**

All scenarios include a natural language format and are presented from a positive perspective

* **8. Please rank the different scenarios based on which you prefer as an explanation towards customers, from the perspective of the insurer.**

Scenario 1: The input parameters and knowledge base (both user and background) are included in the explanation. : ________

Scenario 2: The input parameters and specific procedural decision information (decision inference process) are included in the explanation. : ________

Scenario 3: The input parameters and decision output are included in the explanation. : ________

Scenario 4: The knowledge base (both user and background) and specific procedural decision information (decision inference process) are included in the explanation. : ________

Scenario 5: The knowledge base (both user and background) and decision output are included in the explanation. : ________

Scenario 6: The specific procedural decision information (decision inference process) and decision output are included in the explanation. : ________

* **9. Please elaborate on your chosen order of scenarios.**

Fraud and claims management (Optical character recognition)

Please rank the different scenarios based on which you prefer as an explanation towards customers, from the perspective of the insurer, for the following ADM-system within the process of fraud and claims management:

Optical character recognition (OCR) - Deep learning networks (Artificial Neural Networks) used to extract information from scanned documents such as images from damaged cars to estimate repair costs.

When you would like to review the theory on the elements of an explanation, please click on the following Link

Before you are asked to rank the different scenarios, please read the substantiations from the first round carefully.

Substantiation of Content elements:

- **User knowledge base and background knowledge base information is included for the substantiation of nonstandard claims.**
- **General idea behind the algorithm explains generally how the technique works because this ADM-system is just a smart way to use technology. When explaining 'what the machine does' it is important to consider the differences in knowledge level.**

* **10. Please rank the different scenarios based on which you prefer as an explanation towards customers, from the perspective of the insurer.**

Scenario 1: The input parameters and background knowledge base are included in the explanation. This is explained through a multimedia format. : ________

Scenario 2: The input parameters and decision output are included in the explanation. This is explained through a multimedia format. : ________

Scenario 3: The input parameters and general idea behind the algorithm (decision inference process) are included in the explanation. This is explained through a multimedia format. : ________

Scenario 4: The background knowledge base and decision output are included in the explanation. This is explained through a multimedia format. : ________

* **11. Please elaborate on your chosen order of scenarios.**

We thank you for your time spent participating in this study.

Your response has been recorded. After the results have been analyzed you will be invited to participate in the third and final round.

Appendix XI – 3rd round Delphi-study

*** Required Information**

Dear Participant,

Welcome to my MSc Thesis Research!

I highly appreciate your time and effort in participating in this expert panel. The focus of this research is to understand what explanations, towards customers, of Algorithmic Decision Making (ADM) systems, are preferred in the Dutch P&C insurance industry.

To do this we will use the Delphi-method. This is a method for structuring a group communication process so that the process is effective in allowing a group of individuals, as a whole, to deal with a complex problem. The Delphi-method consists of 3 iterative rounds of surveys within a time period of a few weeks, depending on the number of rounds required to reach consensus. This round will take approximately 15 minutes to participate in.

It should be emphasized that your responses will be kept confidential and used only for research purposes. Your participation in this study is voluntary. In case you decide in the future that you do not want to participate anymore, you can drop out at any moment.

What's in it for YOU? By taking part in this study you can gain new insights about algorithmic decision-making and their explainability.

It is of utmost importance to me that you answer honestly and to the best of your knowledge.

Thank you for your cooperation,

Eric Schotman

Page 1 of 9

In the first round of the Delphi-study, each respondent assessed which 'content' and 'presentation' elements should be included in the explanation, towards customers from the perspective of the insurer. This was assessed for 5 ADM-systems.

In the second round, multiple scenarios of explanations towards customers were ranked based on preferability from the perspective of the insurer.

In this third and final round, the rankings and minority opinions are presented for the 'Algorithmic Decision Making Systems' on which no consensus was reached. This will provide a final opportunity to revise your judgment.

Page 2 of 9

Pricing and underwriting (GLM and random forest)

ADM-System
GLM and random forest - Generalized linear models (GLM) used for pricing. To enhance these GLM's, machine learning random forest models are used in the background. These random forest models combine different decision trees to obtain an aggregated prediction/regression.

Ranking
In round 2, the scenarios (with an explanation towards customers) were ranked, based on preferability from the perspective of the insurer, as follows :

1. **Scenario b: The input parameters and general idea behind the data (decision inference process) are included in the explanation.**
2. **Scenario a: The input parameters and background knowledge base are included in the explanation.**
3. **Scenario c: The input parameters and decision output are included in the explanation.**
4. **Scenario d: The background knowledge base and the general idea behind the data (decision inference process) are included in the explanation.**
5. **Scenario f: The general idea behind the data (decision inference process) and decision output are included in the explanation.**
6. **Scenario e: The background knowledge base and decision output are included in the explanation.**

The consensus is reached, so no further questions for this ADM-system are required, please proceed to the next page.

Pricing and underwriting (Price optimization)

ADM-System
Price optimization - Machine learning churn models for price optimization purposes. Price optimization refers to the practice of adjusting the premiums, paid by different groups of consumers, based on the behaviors and economic characteristics of the consumer, in ways unrelated to their risk or cost.

Ranking
In round 2, the scenarios (with an explanation towards customers) were ranked, based on preferability from the perspective of the insurer, as follows :

1. **Scenario c: Price optimization should not be used and therefore not explained.**
2. **Scenario b: The general idea behind the algorithm (decision inference process) is included in the explanation. This is explained from a positive perspective.**
3. **Scenario a: The input parameters and decision output are included in the explanation. This is explained from a positive perspective.**
4. **Scenario d: Price optimization can be used but should not be explained towards customers.**

Minority opinions
The opinion outside the consensus is:

- **Scenario d should be ranked higher because it is an internal process, which does not need to be shared with customers. For that reason, Scenario c should be ranked lower.**

Finally, when you would like to review the theory on the elements of an explanation, please click on the following Link

*** 1. Do you concur with the ranking as described above? (Select one option)**

☐ Yes, (Please continue to the next page)

☐ No, (Please rank the different scenarios again in the next question)

2. Please rank the different scenarios based on which you prefer as an explanation towards customers, from the perspective of the insurer.

Scenario c: Price optimization should not be used and therefore not explained. : ________

Scenario b: The general idea behind the algorithm (decision inference process) is included in the explanation. This is explained from a positive perspective. : ________

Scenario a: The input parameters and decision output are included in the explanation. This is explained from a positive perspective. : ________

Scenario d: Price optimization can be used but should not be explained towards customers. : ________

3. Please elaborate on your chosen order of scenarios.

Sales and distribution (Recommendation engine)

ADM-System
Recommendation engine - A recommendation engine used for the 'next best action'. This is used to evaluate the customer's past behavior, recent actions and needs to deliver the right message, at the right time, and via the right channel.

Ranking
In round 2, the scenarios (with an explanation towards customers) were ranked, based on preferability from the perspective of the insurer, as follows :

1. **Scenario a: The input parameters and user knowledge base are included in the explanation.**
2. **Scenario d: The user knowledge base and the general idea behind the data (decision inference process) are included in the explanation.**
3. **Scenario b: The input parameters and general idea behind the data (decision inference process) are included in the explanation.**
4. **Scenario c: The input parameters and decision output are included in the explanation.**
5. **Scenario e: The user knowledge base and decision output are included in the explanation.**
6. **Scenario f: The general idea behind the data (decision inference process) and decision output are included in the explanation.**

Minority opinions
The opinions outside the consensus are:

- **Scenario a, b, and/or c should be ranked lower because the input parameters are less important for comfort to the client since the customer already knows this.**
- **Scenario c, e, and/or f should be ranked higher because the decision output (pros and cons compared to alternatives) are important to communicate.**
- **Scenario b, d, and/or f should be ranked higher because the general idea behind the data is information on how you came up with the next best action.**

Finally, when you would like to review the theory on the elements of an explanation, please click on the following Link

*** 4. Do you concur with the ranking as described above? (Select one option)**

- [ ] Yes, (Please continue to the next page)
- [ ] No, (Please rank the different scenarios again in the next question)

5. Please rank the different scenarios based on which you prefer as an explanation towards customers, from the perspective of the insurer.

Scenario a: The input parameters and user knowledge base are included in the explanation. : ________

Scenario d: The user knowledge base and the general idea behind the data (decision inference process) are included in the explanation. : ________

Scenario b: The input parameters and general idea behind the data (decision inference process) are included in the explanation. : ________

Scenario c: The input parameters and decision output are included in the explanation. : ________

Scenario e: The user knowledge base and decision output are included in the explanation. : ________

Scenario f: The general idea behind the data (decision inference process) and decision output are included in the explanation. : ________

6. Please elaborate on your chosen order of scenarios.

Fraud and claims management (Rule-based fraud detection)

ADM-System
Rule-based fraud detection - A rule-based model used to assess claims and evaluate whether they present anomalies and flag potentially fraudulent claims.

Ranking
In round 2, the scenarios (with an explanation towards customers) were ranked, based on preferability from the perspective of the insurer, as follows :

1. **Scenario b: The input parameters and specific procedural decision information (decision inference process) are included in the explanation.**
2. **Scenario a: The input parameters and knowledge base (both user and background) are included in the explanation.**
3. **Scenario d: The knowledge base (both user and background) and specific procedural decision information (decision inference process) are included in the explanation.**
4. **Scenario c: The input parameters and decision output are included in the explanation.**
5. **Scenario e: The knowledge base (both user and background) and decision output are included in the explanation.**
6. **Scenario f: The specific procedural decision information (decision inference process) and decision output are included in the explanation.**

Minority opinions
The opinions outside the consensus are:

- **Scenario c, e, and/or f should be ranked higher because the decision output is important. It should be clear on the interpretation of the output and state that it is just an indication.**
- **Scenario b, d, and/or f should be ranked higher because the specific procedural decision information should be included; being transparent about the decision inference process is of the most importance.**

Finally, when you would like to review the theory on the elements of an explanation, please click on the following Link

*** 7. Do you concur with the ranking as described above? (Select one option)**

☐ Yes, (Please continue to the next page)

☐ No, (Please rank the different scenarios again in the next question)

8. Please rank the different scenarios based on which you prefer as an explanation towards customers, from the perspective of the insurer.

Scenario b: The input parameters and specific procedural decision information (decision inference process) are included in the explanation. : ________

Scenario a: The input parameters and knowledge base (both user and background) are included in the explanation. : ________

Scenario d: The knowledge base (both user and background) and specific procedural decision information (decision inference process) are included in the explanation. : ________

Scenario c: The input parameters and decision output are included in the explanation. : ________

Scenario e: The knowledge base (both user and background) and decision output are included in the explanation. : ________

Scenario f: The specific procedural decision information (decision inference process) and decision output are included in the explanation. : ________

9. Please elaborate on your chosen order of scenarios.

Fraud and claims management (Optical character recognition)

ADM-System
Optical character recognition (OCR) - Deep learning networks (Artificial Neural Networks) used to extract information from scanned documents such as images from damaged cars to estimate repair costs.

Ranking
In round 2, the scenarios (with an explanation towards customers) were ranked, based on preferability from the perspective of the insurer, as follows :

1. **Scenario c: The input parameters and general idea behind the algorithm (decision inference process) are included in the explanation. This is explained through a multimedia format.**
2. **Scenario a: The input parameters and background knowledge base are included in the explanation. This is explained through a multimedia format.**
3. **Scenario d: The background knowledge base and decision output are included in the explanation. This is explained through a multimedia format.**
4. **Scenario b: The input parameters and decision output are included in the explanation. This is explained through a multimedia format.**

Minority opinions
The opinion outside the consensus is:

- **Scenario b and/or d should be ranked higher because the decision output is more important than explaining the input parameters.**

Additional scenario
For this specific ADM-system, based on the following substantiation from some of the participants, a fifth scenario has been added.

Substantiation: For the acceptance of this ADM-system, it is necessary to give and communicate the general idea behind the algorithm and what the decision outputs are for specific cases (where expert opinions have the function of a feedback system).

Scenario e: The general idea behind the algorithm (decision inference process) and decision output are included in the explanation. This is explained through a multimedia format.

Finally, when you would like to review the theory on the elements of an explanation, please click on the following Link

*** 10. Please rank the different scenarios based on which you prefer as an explanation towards customers, from the perspective of the insurer.**

Scenario c: The input parameters and general idea behind the algorithm (decision inference process) are included in the explanation. This is explained through a multimedia format. : ________

Scenario a: The input parameters and background knowledge base are included in the explanation. This is explained through a multimedia format. : ________

Scenario d: The background knowledge base and decision output are included in the explanation. This is explained through a multimedia format. : ________

Scenario b: The input parameters and decision output are included in the explanation. This is explained through a multimedia format. : ________

Scenario e: The general idea behind the algorithm (decision inference process) and decision output are included in the explanation. This is explained through a multimedia format. : ________

*** 11. Please elaborate on your chosen order of scenarios.**

* 12. **Do you have any other thoughts about the explainability of ADM-systems in the P&C insurance industry that you would like to share?**

Page 8 of 9

We thank you for your time spent participating in this study.

Your response has been recorded. This was the final round and therefore the end of this study. If you are interested in the results of the thesis research, please contact me on LinkedIn or via email (schotman.eric@gmail.com).