Calculating the relative importance of condition attributes based on the characteristics of decision rules and attribute reducts: application to crowdfunding

Chakhar, S., Ishizaka, A., Thorpe, A., Cox, J., Nguyen, T. & Ford, L.

Author post-print (accepted) deposited by Coventry University's Repository

Original citation & hyperlink:
Chakhar , S, Ishizaka, A, Thorpe, A, Cox, J, Nguyen, T & Ford, L 2020, 'Calculating the relative importance of condition attributes based on the characteristics of decision rules and attribute reducts: application to crowdfunding', European Journal of Operational Research, vol. 286, no. 2, pp. 689-712.
https://dx.doi.org/10.1016/j.ejor.2020.03.039

DOI 10.1016/j.ejor.2020.03.039
ISSN 0377-2217

Publisher: Elsevier

© 2020, Elsevier. Licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International http://creativecommons.org/licenses/by-nc-nd/4.0/

Copyright © and Moral Rights are retained by the author(s) and/ or other copyright owners. A copy can be downloaded for personal non-commercial research or study, without prior permission or charge. This item cannot be reproduced or quoted extensively from without first obtaining permission in writing from the copyright holder(s). The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the copyright holders.

This document is the author's post-print version, incorporating any revisions agreed during the peer-review process. Some differences between the published version and this version may remain and you are advised to consult the published version if you wish to cite from it.

Calculating the Relative Importance of Condition Attributes Based on the Characteristics of Decision Rules and Attribute Reducts: Application to Crowdfunding

Salem Chakhar[a,b], Alessio Ishizaka[c], Andy Thorpe[a,*], Joe Cox[a,d], Thang Nguyen[e], Liz Ford[a]

[a] *Portsmouth Business School, University of Portsmouth, Portsmouth, UK.*
[b] *Centre for Operational Research and Logistics, University of Portsmouth, Portsmouth, UK.*
[c] *NEOMA Business school, Mont-Saint-Aignan, France.*
[d] *Faculty of Business, Athabasca University, Athabasca, Canada.*
[e] *School of Economics, Finance and Accounting, Faculty of Business & Law, Coventry University, Coventry, UK.*

Abstract

Crowdfunding is the practice of funding a project or venture by raising monetary contributions from a large number of people, typically via the Internet. Lendwithcare is amongst the first crowdfunding platforms specifically dedicated to support individual and group entrepreneurs in developing countries through partner microfinance institutions. A key objective of Lendwithcare is to identify the attributes (i.e., the characteristics of crowdfunding projects in their online descriptions) that affect investors/potential investors when taking their investment decision. This paper proposes a decision rule-based approach to address this issue. This approach relies on the Dominance-Based Rough Approach (DRSA), which is a well-known multicriteria sorting method. The outputs of DRSA are a collection of if-then decision rules and a collection of attribute reducts. In this paper, new measures are proposed for calculating the relative importance of condition attributes based on the characteristics of decision rules and of attribute reducts. Decision rule-based measures are parameterised in order to consider the characteristics of decision rules using both learning and testing datasets. The proposed measures can be aggregated into a comprehensive measure indicating the overall importance of each condition attribute. Furthermore, the proposed measures are extended in order to compute the relative importance of a collection of condition attributes taken together. In addition, decision rule-based measures are extended to evaluate the relative importance of specific values of condition attributes. The proposed approach has been applied and validated using real-world data from Lendwithcare.

Keywords: Rough sets, Dominance-based Rough Set Approach, Relative importance, Decision rules, Attribute reducts, Crowdfunding, Lendwithcare.

1. Introduction

Crowdfunding is an emerging phenomenon involving large numbers of contributors coming together in order to collectively fund projects [8][20][83]. While group funding schemes are not a new phenomenon, the significant reductions in search and transaction costs resulting from a migration onto online platforms are vastly increasing both the scale and scope of projects that can be funded by contributors from around the world. Crowdfunding promises to revolutionise the way in which a wide variety of activities are funded, such as business start-ups, new product innovations, creative and cultural activities and social ventures. The democratic nature of crowdfunding helps to overcome barriers in access to finance among under-represented groups of entrepreneurs and investors, including females, ethnic minorities and those from poorer backgrounds. There are a wide variety of platforms, ranging from donation and rewards-based [11][36] models popularised by Kickstarter and Indigogo, to equity-based funding and peer-to-peer lending.

*Corresponding author.

Even though most attention has been paid to the rise of crowdfunding in the developed world, crowdfunding also has an important role to play in fostering entrepreneurial activity in developing economies. A report from the World Bank [119] clearly highlights this potential by pointing out the substantial amount of entrepreneurial talent and activity lying dormant in many developing economies. The report further suggests that "traditional attitudes towards risk, entrepreneurship and finance stifle potential economic growth and innovation" with a potential to overcome this in part through embracing new financial tools such as crowdfunding [119]. Entrepreneurs and community project leaders from across the developing world have a choice of three operational modalities in leveraging crowdfunding support for their endeavours. First, they can place their project on a local crowdfunding platform. However, there are few such local platforms in the developing world[1] with the World Bank [119] suggesting that such sites are currently limited to Brazil (17 platforms), India (10), South Africa (4), and China (2). Moreover, a cursory examination of such sites shows their scope, in terms of projects, funders and monies raised to date, is somewhat limited. Second, they can place their project on major world-leading rewards-based crowdfunding platforms such as Kickstarter (`www.kickstarter.com`) and Indigogo (`www.indigogo.com`). While such a strategy reaches a much wider audience, this benefit is diluted as the interests of potential investors who access the site are not necessarily congruent with development-oriented projects. An analysis of the two main reward-based crowdfunding platforms (namely, Kickstarter and Indigogo) in fact reveals very few current (or past) entrepreneurial or community-based projects based in the developing world. Third, they can target potential development-oriented investors by placing their project upon crowdfunding platforms explicitly oriented to crowdfunding projects in the developing world (such as `www.lendwithcare.org`, `www.kiva.org` or `www.babyloan.org`).

Lendwithcare (LWC) is a particularly interesting example as it is amongst the first developing countries-dedicated crowdfunding platforms. Created and run by the Cooperative for Assistance and Relief Everywhere (CARE) International,[2] it solicits donations which are linked to the fulfilment of its broader development objectives. LWC allows individuals and groups to make small loans to entrepreneurs in developing countries through partner microfinance institutions. The entrepreneurs themselves do not post their campaigns to the LWC website. Rather, the campaigns are prepared and launched by LWC staff on behalf of the entrepreneurs. LWC currently works in eleven countries, namely Cambodia, Ecuador, Malawi, Palestinian Territories, Pakistan, Peru, the Philippines, Rwanda, Vietnam, Zambia and Zimbabwe, and has two partners in both Ecuador and Zimbabwe. The research considered in this paper is being undertaken in partnership with LWC. LWC uses an online crowdfunding platform where a description of projects can be published and then accessed by different investors. A project description generally consists of the project presentation, a profile photo of the entrepreneur and other pieces of information related to the entrepreneur and the loan requested. The main objective of this study was to help LWC identify the relative importance of the key factors contained in the descriptions of projects that investors/potential investors employ when taking their investment decision. The response to this question is particularly useful to LWC in the sense that it will support its development objectives, enhance the design of its online crowdfunding platform and increase the contributions to crowdfunding campaigns.

To achieve this objective, a rule-based approach that relies on the Dominance-Based Rough Approach (DRSA) is introduced in this paper. The DRSA is a well-known multicriteria sorting (or ordinal classification) method proposed by [48][49][104][105] to overcome the shortcomings of conventional Rough Sets Theory [84][85] in multicriteria sorting by allowing preference-oriented attributes and decision classes. The input of DRSA is a decision table representing the description of a set of objects with respect to a set of condition and decision attributes. The entries of the decision table are attribute-value pairs. The main output of DRSA is a collection of decision rules. Each elementary condition in decision rules is built upon a single condition attribute, while a consequence is defined based on a decision attribute. Another important output of DRSA is a collection of reducts, which are subsets of condition attributes that characterise the knowledge in the decision table. The set of condition attributes that are common to all reducts is called the core.

The condition and decision attributes in DRSA are assumed to be preference ordered. It is possible then to exploit this monotonic property of condition and decision attributes in order to induce the relative importance of condition attributes through the analysis of the characteristics of decision rules and reducts/core subsets. The basic assumption is that attributes that are more important should appear more frequently in the condition parts of decision rules and in the reducts/core subsets than less important attributes.

[1] Examples include: `www.fundfind.co.za`, `www.thundafund.com`, `changa.co.ke`, and `www.startme.co.za`.
[2] See `https://www.careinternational.org.uk/`.

In this paper, new measures are proposed for calculating the relative importance of condition attributes using the characteristics of decision rules and of attribute reducts. The decision rule-based measures are parameterised in order to consider the characteristics of decision rules with respect to both learning and testing datasets. The proposed measures can be aggregated into a comprehensive measure indicating the overall importance of each condition attribute. Furthermore, the proposed measures are extended in order to compute the relative importance of a collection of condition attributes taken together. In addition, decision rule-based measures are extended to evaluate the relative importance of specific values of condition attributes.

The proposed analysis strategy and measures have been applied using a real-world case study that has been conducted in partnership with the pro-social lending-based crowdfunding platform LWC. The dataset contains information on observed real-world patterns of pro-social behaviour taken over several years. In respect to the role of condition attributes in the attractiveness of crowdfunding projects, we can establish the following facts: (i) the predominant role of the condition attribute Activity Type, as confirmed by the analysis of both decision rules and reducts; (ii) the condition attributes Loan Requested, Entrepreneur Gender and Average Loan Value are also significant, but less important than Activity Type; (iii) the condition attributes Country and Number of Dependents play a moderate role in the attractiveness of projects; and (iv) the other condition attributes, namely Largest Loan Value, Number of Participants, Entrepreneur Age and Project Type, play a marginal role in the attractiveness of crowdfunding projects. Results also indicate that the relative importance of pairs of condition attributes is higher for pairs with higher individual relative importance values. The analysis of specific values of Activity Type showed that the most attractive activities are: farming, green loan, raising poultry, sewings/tailoring, food market stall and clothes shops. We also found that projects initiated by females or by pairs and teams are more attractive than projects initiated by males or by individuals.

Although these findings are specific to LWC, they may be useful to other pro-social lending platforms. Furthermore, the proposed measures are generic and can be applied with no modification to any other dataset.

The remainder of the paper is organized as follows. Section 2 discusses related work. Section 3 provides a brief overview on the DRSA and enumerates the basic characteristics of decision rules. Section 4 details the proposed approach. Section 5 applies the proposed approach to crowdfunding projects within LWC. Section 6 presents an additional case study. Section 7 concludes the paper.

2. Related work

This section reviews research on the induction of relative importance of attributes in general (Section 2.1) and in relation to crowdfunding projects in particular (Section 2.2). Then, the novelty of the proposed measures is highlighted (Section 2.3).

2.1. Specification and extraction of relative importance of attributes

2.1.1. Relative importance of attributes within multicriteria analysis methods

Most existing multicriteria methods require the specification of a set of relative importance to the considered attributes. The definition of relative importance of condition attributes is a critical task since it largely determines the final output of the method. We distinguish two major approaches to specify relative importance of attributes [25][34][44]: direct or indirect. Within the direct approach (e.g., [1]), which is most often used in practice, decision makers explicitly specify the relative importance values. The indirect approach (e.g., [35]), sees the relative importance of attributes implicitly derived from the input data. Indirect induction is particularly interesting in practice since it largely reduces the cognitive effort required from the decision maker/expert. Most indirect approaches use mathematical programming formulations [103].

The authors in [27] enumerated and discussed 13 different interpretations of the concept of relative importance of attributes in multicriteria analysis. In particular, they note that most of the authors of relative importance estimation methods misunderstood and/or neglected the interpretation and implications of using relative importance values. More recently, the authors in [103] have enumerated several indirect methods. This paper extends the list given in [103] by adding several relevant and recent methods:

- Simos's card method [100][101] in which the decision maker is asked to rank attributes from the most important to the least important. Relative importance of condition attributes are then induced based on the obtained ranking.

There are several extensions to Simos's method such as the one proposed in [38]. A review of applications of the Simos's method is given in [103].

- The centralized weights method [106] requests that the decision maker makes a number of ordinal comparisons of condition attributes that are formulated as linear inequalities, in order to obtain the centroid of the vertices of a polyhedron.
- In the TACTIC method [118] the relative importance of condition attributes is depicted and assessed as a system of functional representations of relations.
- The DIVAPIME [75] approach, which has been adapted from ELECTRE methods, is implemented by making pairwise comparisons of fictitious alternatives in order to support the elicitation of importance variation intervals.
- The Analytic Hierarchy Process (AHP) [93] is probably the most well-known method for calculating relative importance values from the input data. In AHP, the decision maker is asked to provide pairwise comparisons over the priority of condition attributes on a prespecified numerical scale.
- MACBETH [6] infers the weights as values of attractiveness from pairwise comparisons of the condition attributes on a qualitative scale, thus measuring the magnitude of attractiveness.
- STAB1 and STAB2 methods proposed in [12] and [13], respectively, use a mixed integer linear programming model to infer the relative importance of condition attributes from overall outranking statements, maximizing the stability of the induced median-cut outranking digraph.
- The WAP [117] method assesses weights through prioritization. WAP is a specific integrated implementation of the Robust Simos Method with enriched preferential information, and leads to the estimation of more robust weighting vectors.
- In the FUCOM [81] approach, two groups of constraints that need to satisfy the optimal values of weight coefficients are specified. These constraints involve two conditions: (i) the ratio of the weight coefficients is equal to the comparative priority among the condition attributes; and (ii) weight coefficients should satisfy the condition of mathematical transitivity. The optimal weight values are then obtained by mathematical programming.

The relative importance of decision attributes can also be specified using other methods (e.g., [61][77][78][82]). A critical review of attribute weighting methods in multicriteria analysis is reported in [86].

2.1.2. Relative importance of attributes within rough approximation-based methods

Let us first mention that reducts and core, which are subsets of condition attributes, can be used to partially measure the relative importance of condition attributes and deduce their role in the attractiveness of crowdfunding projects. However, reducts and core subsets are not sufficient to fully and explicitly measure the relative importance of condition attributes. A basic argument in this respect is that one cannot differentiate between the condition attributes in the same reduct or those in the core because they are implicitly considered to have the same importance.

There is also a series of papers that use rough approximation to induce the relative importance of attributes, for example, [44][57][59][67][69][124]. For instance, the authors in [44] used rough sets to prioritise travel attributes based on their proportional impact on tourists' overall satisfaction of their travel experience. The main output of [44] is that improving tourism infrastructures of the country in addition to globally promoting the image of the country are of the highest priority for the country's tourism industry if it is to reach its full potential. These findings provide precious information for tourism policy makers by prioritizing those travel attributes that have the greatest impact on foreign tourists' overall satisfaction with their travel experience.

The authors in [49] investigated the interaction relationships between condition attributes through Shapley value [98] and Banzhaf [7] value. These measures originated in game theory, but can be interpreted, with respect to multicriteria analysis, as specific kinds of weighted average contributions of a given criterion alone to all subsets of condition attributes [49]. Within rough-based methods, the Shapley and Banzhaf values can be interpreted as measures of the contribution of attributes to the quality of approximation of the considered classification [49].

Furthermore, and as observed by [46], fuzzy measures constitute a useful tool for modeling the importance of coalitions. Fuzzy measures have been used in [47] to assess the relative value of information supplied by each condition attribute and to analyse the interactions among attributes, based on the quality of classification calculated from the rough set approach.

In addition to the indices concerning particular condition attributes, other indices have been proposed to measure the interaction between pairs of condition attributes. Interaction indices have been suggested by [76] and [92] with respect to Shapley value and Banzhaf value, respectively. These interaction indices can be interpreted as specific kinds of average-added values resulting from putting two condition attributes together in each possible coalition. Extensions of interaction indices from non-ordered pairs to any subset have been proposed by [45] and [92] with respect to Shapley index and Banzhaf index, respectively. Within rough-based methods, all these interaction indexes can be interpreted as the average conjoint contribution of the non-ordered pair of condition attributes to the quality of the classification.

2.1.3. Relative importance of attributes within other analysis techniques

Induction of relative importance of attributes has also been considered in other analysis techniques such as regression analysis and feature selection. In regression analysis, different authors (e.g., [22][58][62][64][115]) have been interested in identifying the relative importance of considered variables. The authors in [22] applied dominance analysis [21] in order to measure and interpret the relative importance of correlated predictors (variables) in regression models in the context of organizational research.

Feature selection is a very active research domain in data mining, artificial intelligence and also in rough set theory (see e.g., [59][69][107]). The objective of feature selection is to identify the most important attributes. In [59], for example, the authors used filter feature selection algorithms to select subsets of attributes from medical data. The medical relevance of the selected attributes have been checked with the help of domain experts. To eliminate irrelevant or redundant data and improve the performance of a classification system, the author in [107] used a simple wrapper model, based on a sequential backward search procedure, to establish a ranking of attributes and estimate their relevance for the constructed classifier.

2.2. Attractiveness of crowdfunding projects

While previous studies, including [28][41][42][116], have investigated the question of motivation in crowdfunding, others study the role of characteristics in crowdfunding projects and the characteristics of successful crowdfunding campaigns (e.g [74][91]). The authors in [56] presented an interdisciplinary review of investor decision-making in crowdfunding. In the work by [30], the authors studied the effect of profile photos on pro-social crowdfunding campaigns. The authors concluded that profile photos positively affect pro-social behaviours among contributors to online pro-social crowdfunding campaigns.

In the study by [74], the author advocated that personal networks and project quality are associated with the success of crowdfunding efforts, and that geography is related to both the type of projects proposed and successful fundraising. Furthermore, the author in [74] found that being featured on the front page is strongly associated with success. The authors in [91] showed that both social identification with the crowdfunding community and innovativeness have a positive effect on intention to participate. In addition, attitudes toward helping others and interpersonal connectivity indirectly affect intention to participate in crowdfunding through social identification with the crowdfunding community.

An investigation of project duration on a crowdfunder's investment choice is reported in [94] where the authors proposed a dynamic programming approach for computing the value of a platform's opportunity at a given time and deriving the optimal funding decision. They observed in particular the existence of a key interval where the remaining campaign duration can be set up for successful funding, and a corresponding dependence on the link between project target and utility. The authors in [122] confirmed [94]'s results for long duration projects although this is found to be uncertain for short and moderate project duration. Additionally, the author in [74] advocated that duration decreases the chances of success, and explained this as longer durations are seen as a sign of a lack of confidence.

According to [2], all crowdfunding contributors, to some degree, may be thought of as investors, making decisions about which projects to support based on their expectations for success and the underlying appeal of the project. The authors in [3] studied the success of equity crowdfunding platforms. They concluded that: (i) intellectual capital has no effect on the success of crowdfunding projects; (ii) social (alliance) capital has an insignificant effect; and (iii) information about risks and internal governance positively affect the success of crowdfunding projects.

In the study reported in [39], the authors concluded that projects with pairs and teams demonstrate much higher success rates than projects with individuals. They also observed that projects created by females experienced a higher success rate than males. The authors in [26] assessed the relative importance of project, product category, entrepreneur and location effects on reward-based crowdfunding success. They found that agency factors, specifically project and entrepreneur effects, explain the highest relative variance across three crowdfunding success outcomes, namely pledge amount, number of backers and funding success. They also observed that structural factors, specifically product category and location effects, have lower but still significant effects.

2.3. Novelty of proposed measures

A common characteristic of the papers discussed earlier is the fact that they relied on input data to indirectly deduce the relative importance of attributes. In this paper, the relative importance of attributes are deduced from the output data. The proposed measures rely on using existing decision rule-related and attribute reduct-related elementary measures. However, to the best knowledge of the authors, these elementary measures have never been used to calculate the relative importance of attributes.

From a practical point of view, calculating the relative importance of condition attributes using a posteriori information is highly advocated when the decision maker needs to explain and justify the decision to stakeholders in terms of considered relative importance values. Indeed, in several real-world decision problems, the decision maker may appreciate differently the role of each criterion [66][121].

The relative importance of condition attributes may also be useful in the classification of objects using decision rules in situations where it is covered by more than one rule (as discussed in [16]). In this case, it is possible to exploit the relative importance of condition attributes in order to select the decision rule to apply.

In addition, the proposed measures go beyond the preferential information contained in the attribute reduct and core. Indeed, one cannot differentiate between the condition attributes in the same reduct or those in the core because they are implicitly considered to have the same importance; the proposed measures will better discriminate between the condition attributes.

3. Background

3.1. Dominance-based Rough Set Approach

The Dominance-based Rough Set Approach (DRSA) [48][49][104][105] is an extension of conventional Rough Set Theory to multicriteria analysis. A brief overview of the DRSA is presented in the rest of this section. More details on this method are available in [23][24][49].

3.1.1. Basic definitions

In rough sets theory, information regarding the *decision objects* is often structured in a 4-tuple *information table* $\mathbf{S} = \langle U, Q, V, f \rangle$, where U is a non-empty finite set of objects and Q is a non-empty finite set of attributes such that $q : U \to V_q$ for every $q \in Q$. The V_q is the domain of attribute q, $V = \bigcap_{q \in Q} V_q$, and $f : U \times Q \to V$ is the *information function* defined such that $f(x, q) \in V_q$ for each attribute q and object $x \in U$. The set Q is often divided into a subset $C \neq \emptyset$ of *condition attributes* and a subset $D \neq \emptyset$ of *decision attributes*, such that $C \cup D = Q$ and $C \cap D = \emptyset$. In this case, **S** is called a *decision table*.

In multicriteria decision making, the domains of the condition attributes are ordered according to a decreasing or increasing preference. The proponents of DRSA assume that the preference is increasing with $f(\cdot, q)$ for every $q \in C$. They also assume that the set of decision attributes D is a singleton. The unique decision attribute makes a partition of U into a finite number of preference ordered decision classes $\mathbf{Cl} = \{Cl_t, t \in T\}$, $T = \{1, \ldots, |T|\}$, such that each $x \in U$ belongs to one and only one class.

3.1.2. Approximations

In DRSA the represented knowledge is a collection of *upward unions* $Cl_t^{\geq}$ and *downward unions* $Cl_t^{\leq}$ of classes defined as follows:

$$Cl_t^{\geq} = \bigcup_{s \geq t} Cl_s, Cl_t^{\leq} = \bigcup_{s \leq t} Cl_s. \tag{1}$$

The assertion "$x \in Cl_t^{\geq}$" means that "x belongs to at least class Cl_t" while assertion "$x \in Cl_t^{\leq}$" means that "x belongs to at most class Cl_t". The basic idea of DRSA is to replace the indiscernibility relation used in the conventional Rough Set Theory with a dominance relation. Let $P \subseteq C$ be a subset of condition attributes. The *dominance relation* Δ_P associated with P is defined for each pair of objects x and y as follows:

$$x\Delta_P y \Leftrightarrow f(x,q) \succeq f(y,q), \forall q \in P. \tag{2}$$

In the definition above, the symbol "$\succeq$" should be replaced with "$\preceq$" for condition attributes that are ordered according to decreasing preferences. To each object $x \in U$, we associate two sets: (i) the *P-dominating set* $\Delta_P^+(x) = \{y \in U : y\Delta_P x\}$ containing the objects that dominate x, and (ii) the *P-dominated set* $\Delta_P^-(x) = \{y \in U : x\Delta_P y\}$ containing the objects dominated by x.

Then, the P-lower and P-upper approximations of $Cl_t^{\geq}$ with respect to P are defined as follows:

- $\underline{P}(Cl_t^{\geq}) = \{x \in U : \Delta_P^+(x) \subseteq Cl_t^{\geq}\}$,
- $\bar{P}(Cl_t^{\geq}) = \{x \in U : \Delta_P^-(x) \cap Cl_t^{\geq} \neq \emptyset\}$.

Analogously, the P-lower and P-upper approximations of $Cl_t^{\leq}$ with respect to P are defined as follows:

- $\underline{P}(Cl_t^{\leq}) = \{x \in U : \Delta_P^-(x) \subseteq Cl_t^{\leq}\}$,
- $\bar{P}(Cl_t^{\leq}) = \{x \in U : \Delta_P^+(x) \cap Cl_t^{\leq} \neq \emptyset\}$.

The lower approximations group the objects which certainly belong to class unions $Cl_t^{\geq}$ (resp. $Cl_t^{\leq}$). The upper approximations group the objects which could belong to $Cl_t^{\geq}$ (resp. $Cl_t^{\leq}$).

The P-boundaries of $Cl_t^{\geq}$ and $Cl_t^{\leq}$ are defined as follows:

- $Bn_P(Cl_t^{\geq}) = \bar{P}(Cl_t^{\geq}) - \underline{P}(Cl_t^{\geq})$,
- $Bn_P(Cl_t^{\leq}) = \bar{P}(Cl_t^{\leq}) - \underline{P}(Cl_t^{\leq})$.

The boundaries group objects that can neither be ruled in nor out as members of class $Cl_t^{\geq}$ (resp. $Cl_t^{\leq}$).

3.1.3. Quality and accuracy of approximation

The *quality of approximation* of a partition **Cl** by means of a set of condition attributes P is defined as the ratio of all *P*-correctly classified objects to all objects in the system. Mathematically,

$$\gamma(\mathbf{Cl}) = \frac{|U - ((\bigcup_{t \in T} Bn_P(Cl_t^{\geq})) \bigcup (\bigcup_{t \in T} Bn_P(Cl_t^{\leq})))|}{|U|}. \tag{3}$$

The accuracy of the rough-set representation of classes is computed as the ratio between the number of objects in the lower approximation and the number of objects in the upper approximation. Mathematically,

$$\alpha(Cl_t^{\geq}) = \frac{\underline{P}(Cl_t^{\geq})}{\overline{P}(Cl_t^{\geq})}, \quad \text{and} \quad \alpha(Cl_t^{\leq}) = \frac{\underline{P}(Cl_t^{\leq})}{\overline{P}(Cl_t^{\leq})} \tag{4}$$

3.1.4. Reduct and core

The DRSA defines two concepts that indicate information about the importance of the condition attributes. These concepts are the reduct and the core. A reduct is a subset of condition attributes that can, by itself, fully characterise the knowledge in the decision table. The reduct of the decision table is not unique and there may be many subsets of attributes that preserve the equivalence-class structure. The set of attributes that is common to all reducts is called the core. Therefore, they are the condition attributes that cannot be removed from the decision table without causing a collapse of the equivalence-class structure.

Finding all the reducts is an NP-hard problem [80]. Different algorithms for generating reducts have been proposed in the literature, (e.g., [37][80]). For example reducts and core can be generated based on the discernibility matrix as follows [80]: (i) the reduct is the minimal element in the discernibility matrix that intersects all the element of the discernibility matrix; and (ii) the core is the set of all singleton entries in the discernibility matrix.

3.1.5. Decision rules

The decision attribute induces a partition of U in a way that is independent of the condition attributes. Hence, a decision table may be seen as a set of if-then decision rules. The condition part specifies the values assumed by one or more condition attributes, and the decision part specifies an assignment to one or more decision classes. Three types of decision rules may be considered: (i) certain rules generated from the lower approximations of unions of classes, (ii) possible rules generated from the upper approximations of unions of classes, and (iii) approximate rules generated from the boundary regions.

The general structures of certain decision rules are as follows:

If *condition(s)*, **then** *At Most* Cl_t
If *condition(s)*, **then** *At Least* Cl_t

The decision part of a certain rule takes the form of an assignment to at most class unions or at least class unions.

The general structures of possible decision rules are as follows:

If *condition(s)*, **then** *Possibly At Most* Cl_t
If *condition(s)*, **then** *Possibly At Least* Cl_t

In this case, the decision part specifies a possible assignment to class unions.

Finally, the general structure of approximate rules is as follows:

If *condition(s)*, **then** *Belongs to* $Cl_s \cup Cl_{s+1} \cup \ldots \cup Cl_t$

Here, the decision part is defined as the union of several decision classes.

3.2. Basic characteristics of decision rules

In this section, we provide a set of basic quantitative characteristics of decision rules. More details are available in [53][63][71][87][109][113]. In the rest of the paper, a decision rule is represented as a consequence relation $E \to H$ (i.e., read as *If* E, *then* H) where E is a condition premise (or evidence, antecedent) and H is a decision (or consequence, conclusion, hypothesis). We also denote by $[[E]]$ and $[[H]]$ the set of objects from U, having the property E and H, respectively. Additionally, the cardinality of a given set, let us say X, will be denoted as $card(X)$.

The following list provides the basic quality measures for decision rule $\rho : E \to H$:

Matching. An object $u \in U$ *matches* decision rule ρ in case $u \in [[E]]$, in other words, the object u verifies the premise of the rule.

Support. The *support* is the number of objects matching both the premise and the conclusion of the rule:

$$sup(\rho) = card([[E \wedge H]]). \tag{5}$$

Strength. The *strength* is defined as the number of positive objects covered by the rule, in other words, number of objects correctly classified by the rule in the training phase divided by the total number of objects:

$$str(\rho) = \frac{card([[E \wedge H]])}{card(U)}. \tag{6}$$

Accuracy. The *accuracy* is the number of positive objects covered by the rule divided by the number of objects covered by the rule:

$$acc(\rho) = \frac{card([[E \wedge H]])}{card([[E]])}. \tag{7}$$

We note that some authors (as [10][63]) defined accuracy through other terms such as *consistency level*, *certainty factor*, *confidence factor*, *discrimination level* or *precision*.

Coverage factor. The *coverage factor* is the number of positive objects covered by the rule divided by the number of all positive objects in the class:

$$cov(\rho) \quad = \quad \frac{card([[E \wedge H]])}{card([[H]])}. \tag{8}$$

We note that some authors, for example, [87][109], defined the coverage factor by *relative strength*.

Specificity. The specificity of the rule ρ, *spec*(ρ), is the number of descriptors in the premise of the rule. The term *length* is also often used (see e.g., [5][125]) to design the number of descriptors in the premise of the rule. The authors in [109] used the term *simplicity* instead of specificity while those in [40] call it *conciseness*.

4. Measures for calculating the relative importance of condition attributes

4.1. General schema of the analysis strategy

As underlined previously, the analysis strategy relies on the DRSA. The basic idea consists in using the characteristics of the outputs of dominance-based approximations in order to define a set of relative importance measures associated with the different condition attributes. The main outputs of DRSA are (i) a collection of decision rules, and (ii) a collection of reducts and a core. This will then lead to two different ways to induce the relative importance of condition attributes: (1) relative importance based on the characteristics of decision rules, and (2) relative importance based on the characteristics of reducts and core.

In this paper, new measures are proposed for determining the relative importance of condition attributes based on the characteristics of decision rules and of attribute reducts. The decision rule-based measures are parameterised in order to consider the characteristics of decision rules using both learning and testing datasets. The proposed measures can be aggregated into a comprehensive measure indicating the overall importance of each condition attribute. Additionally, measures based on the characteristics of decision rules are extended to evaluate the relative importance of specific values of condition attributes.

The main input of the analysis strategy is obtained from the approximations of the decision table using the DRSA. However, in the case study considered in this paper, only the information table is available. This is because the definition of an appropriate decision table (by assigning the projects to different classes) by the decision maker will be time-consuming due to the large dataset used in the case study. To overcome this issue, the assignment procedure that we have proposed in [29][31] has been used to assign the crowdfunding projects into different attractiveness classes, which are later used as decision classes for the application of the DRSA.

We should stress that the use of the assignment procedure is not necessary if the assignment of decision objects into different classes can be obtained from the decision maker. Furthermore, any other approach to assign the decision objects into classes can be used. In particular, multicriteria clustering methods (e.g., [19][32][33][72][89][90]) can be easily applied to group crowdfunding projects into different attractiveness decision classes.

The whole analysis strategy is depicted graphically in Figure 1. First, the additional attributes introduced in Section 5.3 are used as inputs to the assignment procedure so as to produce the attractiveness classes of crowdfunding projects. Then, the information table is transformed into a decision table by adding the specification of decision classes to which the crowdfunding projects have been assigned. The DRSA is then applied on the obtained decision table. Next, the relative importance of condition attributes are induced from the outputs of DRSA using the proposed measures. Finally, the results of proposed measures are combined to give the overall relative of importance condition attributes.

4.2. Assignment procedure

The assignment procedure relies on the use of a posteriori information collected from the LWC database. This information permits us to define three criteria: (i) the project advertising campaign duration, which is defined as the number of days from the publication of the project on the LWC platform until the full funding of the project is obtained;

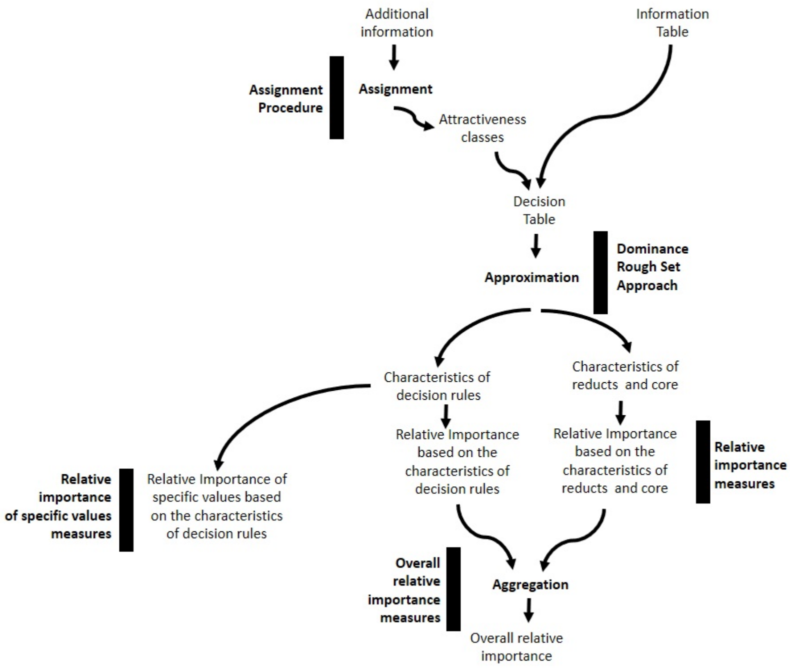


Figure 1: Analysis strategy

(ii) the number of lenders who effectively supported the project; and (iii) the number of lender groups who effectively supported the project. These criteria have different effects on the attractiveness of the crowdfunding projects. Indeed, an increase in the number of supporting lenders or group of lenders indicates a higher attractiveness while a decrease in the values of these criteria shows a lower attractiveness. The campaign duration varies inversely with project attractiveness, since an increase in the value of the campaign duration indicates a lower attractiveness while a decrease shows a higher attractiveness.

The rationale of the assignment procedure consists in using the three criteria defined above and then applying the dominance relation in order to group the crowdfunding projects in the learning set into different classes based on their description with respect to these criteria. Let U be the set of crowdfunding projects. Then, one can use the dominance relation Δ_P (where P is the set of the above-cited criteria) given in Equation (2) to assign the crowdfunding projects in U into different preference ordered decision classes. Note that Equation (2) implements the weak version of the dominance relation (since all the inequality in Equation (2) are large), which is reflexive (i.e., $x\Delta_P x$, $x \in U$), and transitive (i.e., if $x\Delta_P y$ and $y\Delta_P z$, then $x\Delta_P z$, $\forall x, y, z \in U$). The weak version of the dominance relation defines a partial preorder on the set U of crowdfunding projects. Any preorder can be represented by a directed graph, with elements of the set U corresponding to vertices, and the order relation between pairs of elements corresponding to the directed edges between vertices. Therefore, the dominance relation can be represented as a directed graph $G = (U, X)$ with elements of the crowdfunding projects set U corresponding to vertices, and the dominance relation between pairs of elements corresponding to the directed edges X between vertices, defined as $X = \{(x, y) \in U \times U : x\Delta y\}$. The graph G is constructed using a top-to-bottom order. This means that if a node x dominates a node y, x appears above y in the graph.

The set of decision classes can be induced from the directed graph $G = (U, X)$ using the following idea. First, we identify a minimal subset $N \subseteq U$ such that: (i) any crowdfunding project that is not in N is dominated by at least one crowdfunding project from N; and (ii) the crowdfunding projects in set N are incomparable (i.e., they do not dominate each other, excepting the self-dominance). The set N is called a *kernel* of graph G, the dominant subset or also the external stability. We note that if the graph G has no cycle, the kernel exists and is unique. We note also that each cycle can be replaced by a unique element (considering the crowdfunding projects in the cycle as tied). The elements of N are assigned to the most preferred decision class. Then, the same procedure is used to identify the kernel N' of the sub-graph $G' = (U \setminus N, X')$. The elements of set N' are then assigned to the second most preferred decision class. The same procedure is repeated until all the crowdfunding projects are assigned. An illustrative example of the assignment procedure is provided in Appendix A.

4.3. Relative importance based on the characteristics of decision rules

The process for evaluating the role played by each condition attribute in the attractiveness of crowdfunding projects based on the characteristics of decision rules is composed of two steps. First, the characteristics of decision rules are used to deduce the attractiveness of decision rules. Second, the importance of condition attributes are computed using the attractiveness of decision rules. The basic rules attractiveness (interestingness) measures have been introduced in Section 3.2. However, measuring the attractiveness of decision rules has been addressed in a large number of papers and other different measures have been proposed in the literature, see for example, [40][65][87][79][123]. The authors in [60] allocated rule evaluation measures into two main categories: (i) the first involves assessing the structure of the rule, based primarily on its length, and (ii) the second relates to rules performance such as rule coverage, certainty, and confirmation, etc.

In most of the existing literature, rule attractiveness measures have been investigated from the perspective of rules pruning, especially when there is a large number of decision rules, in order to eliminate irrelevant, weak or obvious ones; see for example [108]. A second application of rules attractiveness measures is feature extraction, where the characteristics of decision rules are used to identify the most important features; see for example, [70][95]. Another current application of rules attractiveness measures concerns the selection of the rule to apply in case a test object is covered by more than one decision rule; see for example, [16][111][112]. In this paper, rules attractiveness measures are used to induce importance of condition attributes. At this level, it is important to note the existence of some works that use attributes rankings as in [107][108], or the importance of elementary conditions as in [99] as basis for rule selection or redefinition.

The literature shows that most rules attractiveness measures consider the qualitative evaluation of rules with respect to the learning dataset. We however advocate that rules attractiveness measures should also take into account the behaviour of decision rules with respect to testing datasets. In this paper, we propose a comprehensive and parameterised measure that combines qualitative evaluation of decision rules with respect to both learning and testing datasets.

The attractiveness of decision rule ρ with respect to learning data relies on a cost-type consistency measure $\widehat{\epsilon}(\rho)$, the accuracy $acc(\rho)$ and coverage $cov(\rho)$ of the decision rule. The cost-type consistency measure $\widehat{\epsilon}(\rho)$, introduced in [17][18] (see also [111][112]), is defined as follows:

$$\widehat{\epsilon}(\rho) \quad = \quad \frac{card([[E]] \cap \neg H)}{card(\neg H)}. \tag{9}$$

This cost-type consistency measure has been used in [112] for defining the strength of rule ρ as $\tau(\rho) = (1 - \widehat{\epsilon}(\rho))cov(\rho)$ where $cov(\rho)$ is the coverage factor of rule ρ defined in Equation (8). In this paper, we propose to enhance the definition of $\tau(\rho)$ by replacing the coverage factor $cov(\rho)$ by a combination of accuracy and coverage, proposed by [73], and defined as follows:

$$\mu(\rho) \quad = \quad \frac{1}{2}acc(\rho) + \frac{1}{4}acc(\rho)^2 + \frac{1}{2}cov(\rho) - \frac{1}{4}acc(\rho)cov(\rho). \tag{10}$$

We then define the attractiveness of decision rule ρ with respect to learning data as follows:

$$a_l(\rho) \quad = \quad (1 - \widehat{\epsilon}(\rho))\mu(\rho). \tag{11}$$

The main argument of this new definition is that $\mu(\rho)$ better expresses the trade-off between accuracy and coverage, as advocated by [87]: one may have a high accuracy on a relatively small set of covered objects, or lesser accuracy on a larger set of covered objects.

Let us now define the attractiveness of a decision rule ρ with respect to the testing dataset. As remarked by [40], a set of classification rules as a whole is often used for the prediction of an unseen dataset. The most common measure used to evaluate the quality of a set of classification rules is *predictive accuracy*. The predictive accuracy of each decision rule ρ can be defined as the number of testing objects correctly classified by the rule divided by the total number of testing objects covered by the rule:

$$pacc(\rho) = \begin{cases} \frac{card([[E \wedge H]]_t)}{card([[E]]_t)}, & \text{if } [[E]]_t \neq \emptyset, \\ 0, & \text{otherwise.} \end{cases} \quad (12)$$

where $[[E \wedge H]]_t$ is the number of positive testing objects (i.e., satisfying the premise and conclusion of the rule ρ) and $[[E]]_t$ is the number of testing objects covered by the rule ρ. This conditional definition of the predictive accuracy precludes the division by zero when there are no testing objects covered by the rule.

Similarly, a rule *predictive coverage* for each decision rule ρ can be defined as the number of testing objects correctly classified divided by the number of all positive testing objects:

$$pcov(\rho) = \begin{cases} \frac{card([[E \wedge H]]_t)}{card([[H]]_t)}, & \text{if } [[H]]_t \neq \emptyset, \\ 0, & \text{otherwise.} \end{cases} \quad (13)$$

where $[[E \wedge H]]_t$ is as above and $[[H]]_t$ is the number of all positive testing objects. As previously mentioned, this conditional definition of the predictive coverage avoids the division by zero when there are no positive testing objects. We also need to point out that in the calculation of the predictive accuracy and coverage values, in particular when looking for supporting testing objects (i.e., the set $[[E \wedge H]]_t$) and positive testing objects (i.e., the set $[[H]]_t$), we should consider decision attribute values as specified by the decision rules, not the initial values.

The attractiveness $a_t(\rho)$ of decision rule ρ with respect to the testing dataset can then be defined by combining rule predictive accuracy and rule predictive coverage in a similar way to Equation (10):

$$a_t(r) = \frac{1}{2}pacc(\rho) + \frac{1}{4}pacc(\rho)^2 + \frac{1}{2}pcov(\rho) - \frac{1}{4}pacc(\rho)pcov(\rho). \quad (14)$$

The overall attractiveness $a(\rho)$ of decision rule ρ with respect to both learning and testing datasets can be defined as a linear combination of $a_l(\rho)$ and $a_t(\rho)$ as follows:

$$a(\rho) = \alpha a_l(\rho) + (1-\alpha)a_t(\rho). \quad (15)$$

where $\alpha \in [0, 1]$. This comprehensive and parameterised measure offers a flexible way for evaluating the attractiveness of decision rules. It is easy to see that a value of $\alpha = 1$ means that the overall attractiveness of the rule relies solely on the attractiveness of the decision rule with respect to the learning dataset, while a value of $\alpha = 0$ means that the overall attractiveness of the rule relies solely on the attractiveness of the decision rule with respect to the testing dataset. For other values of α such that $0 < \alpha < 1$, the overall attractiveness of rule ρ takes into account both learning and testing datasets.

The rules attractiveness measure given in Equation (15) can be exploited to compute the importance of condition attributes. The contribution of a decision rule ρ in the importance of condition attribute q can be measured by $a(\rho)/spec(\rho)$, in other words, the ratio of the attractiveness $a(\rho)$ of rule ρ and its specificity (i.e., the number of descriptors in the premise of the rule ρ). This means that the importance of condition attributes vary positively with respect to the attractiveness of decision rule ρ and negatively with the specificity of this rule. The arguments behind the use of the ratio $a(\rho)/spec(\rho)$ to measure the contribution of decision rules to the importance of condition attributes is that rules with less elementary conditions have a higher interestingness, as advocated by several authors, see for example, [60][109][125], and are seen as more important. The authors in [5] justified the use of rule length as a parameter through the Minimum Description Length principle, established by [88], in which shorter data descriptions are preferred to longer data descriptions.

Let R be the set of decision rules. Then, for each attribute $q \in C$, we define $F(q)$ as the set of rules including an elementary condition based on condition attribute q. The absolute $i(q)$ and relative $i'(q)$ versions of the importance of condition attribute q are then respectively defined as follows:

$$i(q) = \sum_{\rho \in F(q)} \frac{a(\rho)}{spec(\rho)}. \tag{16}$$

and

$$i'(q) = \frac{1}{card(R)} \sum_{\rho \in F(q)} \frac{a(\rho)}{spec(\rho)}. \tag{17}$$

Rule 1. *Let q_1 and $q_2 \in C$ be two condition attributes from C. Based on the absolute (resp. relative) importance measure of condition attributes with respect to decision rules, we can conclude that if $i(q_1) > i(q_2)$ (resp. $i'(q_1) > i'(q_2)$) then condition attribute q_1 is more important than condition attribute q_2.*

The definitions of absolute and relative importance measures in Equations (16) and (17) can be extended to compute the absolute and relative importance measures of a collection of condition attributes:

$$i_e(q_1, \ldots, q_p) = \sum_{\rho \in F_e(q_1, \ldots, q_p)} \frac{a(\rho)}{spec(\rho)}. \tag{18}$$

and

$$i'_e(q_1, \ldots, q_p) = \frac{1}{card(R)} \sum_{\rho \in F_e(q_1, \ldots, q_p)} \frac{a(\rho)}{spec(\rho)}. \tag{19}$$

where $F_e(q_1, \ldots, q_p)$ is the set of rules including (jointly) elementary conditions based on condition attributes $q_1, \ldots, q_p$.

Rule 2. *Let $(q_1^a, \ldots, q_p^a) \subseteq C$ and $(q_1^b, \ldots, q_p^b) \subseteq C$ be two collections of condition attributes from C. Based on the absolute (resp. relative) importance measures of condition attributes, we can conclude that if $i(q_1^a, \ldots, q_p^a) > i(q_1^b, \ldots, q_p^b)$ (resp. $i'(q_1^a, \ldots, q_p^a) > i'(q_1^b, \ldots, q_p^b)$) then the collection of condition attributes $q_1^a, \ldots, q_p^a$ is more important than the collection of condition attributes $q_1^b, \ldots, q_p^b$.*

4.4. Relative importance based on the characteristics of attribute reducts

In this section, we propose some attribute importance measures based on the characteristics of attribute reducts associated with the learning dataset. A reduct was previously defined as a subset of condition attributes that fully characterise the knowledge in the learning dataset. The set of condition attributes that is common to all reducts is called the core. Let RED be the set of reducts obtained for a given learning set. Each reduct $rd \in RED$ can be characterized by (i) the list of condition attributes included in the reduct; and (ii) its specificity $spec(rd)$ defined as the number of condition attributes composing the reduct.

The importance of condition attribute q with respect to a given reduct rd can be defined as follows:

$$j(q|rd) = \begin{cases} \frac{1}{spec(rd)}, & \text{if } q \in rd, \\ 0, & \text{otherwise.} \end{cases} \tag{20}$$

In this definition, we assumed that the importance of condition attribute q varies negatively with the cardinality of reduct rd. This relies on the fact that the discrimination level of a given condition attribute decreases with the number of the condition attributes in the reduct. This is also in concordance with the assumption that decision rules with shorter elementary conditions are more attractive in practice, as advocated by different authors, for example, [5][60][109][88][125].

The absolute and relative importance of a condition attribute $q \in C$ with respect to a set RED of reducts can then be computed as follows:

$$j(q) \quad = \sum_{rd \in K(q)} j(q|rd). \tag{21}$$

$$j'(q) \quad = \quad \frac{\sum_{rd \in K(q)} j(q|rd)}{card(RED)}. \tag{22}$$

where $K(q)$ is the set of reducts including condition attribute q.

Rule 3. *Let q_1 and $q_2 \in C$ be two condition attributes from C. Based on the absolute (resp. relative) importance measure of condition attributes with respect to reducts, we can conclude that if $j(q_1) > j(q_2)$ (resp. $j'(q_1) > j'(q_2)$) then condition attribute q_1 is more important than condition attribute q_2.*

As with decision rules, the absolute and relative importance of a condition attribute with respect to reducts can be extended to several condition attributes. Let us first extend Equation (20) to more than one condition attribute as follows:

$$j_e(q_1, \ldots, q_p|rd) \quad = \quad \begin{cases} \frac{1}{spec(rd)}, & \text{if } (\{q_1, \ldots, q_p\} \subseteq rd, \\ 0, & \text{otherwise.} \end{cases} \tag{23}$$

Then, the extended absolute and relative importance of a set of condition attributes with respect to reducts are computed as follows:

$$j_e(q_1, \ldots, q_p) \quad = \sum_{rd \in K_e(q_1, \ldots, q_p)} j_e(q_1, \ldots, q_p|rd). \tag{24}$$

$$j'_e(q_1, \ldots, q_p) \quad = \quad \frac{\sum_{rd \in K_e(q_1, \ldots, q_p)} j_e(q_1, \ldots, q_p|rd)}{card(RED)}. \tag{25}$$

where $K_e(q_1, \ldots, q_p)$ is the set of reducts containing the condition attributes $q_1, \ldots, q_p$.

Rule 4. *Let $(q_1^a, \ldots, q_p^a) \subseteq C$ and $(q_1^b, \ldots, q_p^b) \subseteq C$ be two collections of condition attributes from C. Based on the absolute (resp. relative) importance measure of condition attributes with respect to reducts, we can conclude that if $j(q_1^a, \ldots, q_p^a) > j(q_1^b, \ldots, q_p^b)$ (resp. $j'(q_1^a, \ldots, q_p^a) > j'(q_1^b, \ldots, q_p^b)$) then the collection of condition attributes $q_1^a, \ldots, q_p^a$ is more important than the collection of condition attributes $q_1^b, \ldots, q_p^b$.*

Finally, we observe that the relative importance of attributes can be deduced from the core, denoted COR in the rest of this paper. However, this will not be considered in this paper since the relative importance information that can be induced from the core is already included in the reduct-based measures. This is because for any pair of attributes $q_1 \in COR$ and $q_2 \notin COR$, we have $j(q_1) > j(q_2)$ (resp. $j'(q_1) > j'(q_2)$). This result holds also for any collection of attributes $q_1^a, \ldots, q_p^a \subseteq COR$ and $q_1^b, \ldots, q_p^b \nsubseteq COR$. These properties are proved in Appendix B.3 (Property 4.2) and Appendix B.4 (Property 6.2) , respectively.

4.5. Overall importance of condition attributes

The decision rule-based and attribute reducts-based importance measures of condition attributes can be combined to obtain an overall relative importance as follows:

$$o(q) \quad = \quad \beta i(q) + (1-\beta) j(q). \tag{26}$$

$$o'(q) \quad = \quad \beta i'(q) + (1-\beta) j'(q). \tag{27}$$

where $\beta \in [0,1]$. The parameter β permits us to parameterise the computing of the overall importance of condition attributes by giving more importance to decision rule-based importance measures than to attribute reducts-based importance measures (when $\beta > 0.5$) or the opposite (when $\beta < 0.5$).

Rule 5. *Let $q_1, q_2 \in C$ be two condition attributes from C. Based on the absolute (resp. relative) overall importance measure of condition attributes, we can conclude that if $o(q_1) > o(q_2)$ (resp. $o'(q_1) > o'(q_2)$), then condition attribute q_1 is more important than condition attribute q_2.*

The overall absolute and relative importance above can be extended to more than one condition attribute as follows:

$$o_e(q_1, \ldots, q_p) \quad = \quad \beta i_e(q_1, \ldots, q_p) + (1-\beta) j_e(q_1, \ldots, q_p). \tag{28}$$

$$o'_e(q_1, \ldots, q_p) \quad = \quad \beta i'_e(q_1, \ldots, q_p) + (1-\beta) j'_e(q_1, \ldots, q_p). \tag{29}$$

where $\{q_1, \ldots, q_p\} \subseteq C$ is a subset of condition attributes and $\beta \in [0,1]$.

Rule 6. *Let $\{q^a_1, \ldots, q^a_p\} \subseteq C$ and $\{q^b_1, \ldots, q^b_p\} \subseteq C$ be two collections of condition attributes from C. Based on the absolute (resp. relative) extended overall importance measure of condition attributes, we can conclude that if $o(q^a_1, \ldots, q^a_p) > o(q^b_1, \ldots, q^b_p)$ (resp. $o'(q^a_1, \ldots, q^a_p) > o'(q^b_1, \ldots, q^b_p)$) then the collection of condition attributes $q^a_1, \ldots, q^a_p$ is more important than the collection of condition attributes $q^b_1, \ldots, q^b_p$.*

4.6. Importance of specific values of condition attributes

In this section, we further analyse the condition attributes based on their specific values. The importance of specific values is first defined based on the characteristics of decision rules. The obtained measures can further be constrained by using the characteristics of attractiveness decision classes.

4.6.1. Importance of specific values based on characteristics of decision rules

The decision rule-based importance of condition attributes given earlier can be extended to compute the importance of specific values of these attributes by constraining Equations (16) and (17). First, focus on the case of condition attributes with discrete domains, in other words, symbolic, nominal or ordinal attributes. Then, Equation (16) can be constrained as follows

$$s(q|v) \quad = \sum_{\rho \in F(q) \wedge \rho.r_q = v} \frac{a(\rho)}{spec(\rho)}; \tag{30}$$

where $v \in V_q$ and $\rho.r_q$ is the right-hand side of the elementary condition relative to condition attribute q in decision rule ρ. The quantity $i(q,v)$ represents the absolute importance of specific value v of condition attributes q.

Similarly, the relative importance of condition attributes given in Equation (17) can be extended as follows:

$$s'(q|v) \quad = \quad \frac{1}{card(R)} \sum_{\rho \in F(q) \wedge \rho.r_q = v} \frac{a(\rho)}{spec(\rho)}. \tag{31}$$

Rule 7. *Let $q \in C$ be a condition attribute from C with discrete domain V_q. Let v_1 and v_2 be two attribute values from V_q. Based on the absolute (resp. relative) importance measure of condition attributes' specific values defined based on characteristics of decision rules, we can conclude that if $s(q|v_1) > s(q|v_2)$ (resp. $s'(q|v_1) > s'(q|v_2)$), then attribute value v_1 is more important than attribute value v_2.*

Equations (30) and (31) can be extended to a collection of condition attributes as follows:

$$s_e(q_1,\ldots,q_p|v_1,\ldots,v_p) \quad = \sum_{\rho \in F_e(q_1,\ldots,q_p)\ and \bigwedge_{k=1}^{p} \rho.r_{q_k}=v_k} \frac{a(\rho)}{spec(\rho)}, \tag{32}$$

and

$$s'_e(q_1,\ldots,q_p|v_1,\ldots,v_p) \quad = \quad \frac{1}{card(R)} \sum_{\rho \in F_e(q_1,\ldots,q_p)\ and \bigwedge_{k=1}^{p} \rho.r_{q_k}=v_k} \frac{a(\rho)}{spec(\rho)}; \tag{33}$$

where, for $k = 1,\ldots,p$, $v_k \in V_{q_k}$ and $\rho.r_{q_k}$ is the right-hand side of the elementary condition relative to condition attribute q_k in decision rule ρ.

Rule 8. *Let $\{q_1,\ldots,q_p\} \subseteq C$ be a collection of condition attributes from C with discrete domains $V_1,\ldots,V_p$. Let $v_1^a,\ldots,v_p^a$ and $v_1^b,\ldots,v_p^b$ be two collections of attribute values from $V_1,\ldots,V_p$. Based on the extended absolute (resp. relative) importance measure of collections of condition attributes' specific values defined based on characteristics of decision rules, we can conclude that if $s(q_1,\ldots,q_p|v_1^a,\ldots,v_p^a) > s(q|v_1^b,\ldots,v_p^b)$ (resp. $s'(q_1,\ldots,q_p|v_1^a,\ldots,v_p^a) > s'(q|v_1^b,\ldots,v_p^b)$), then the collection of attributes' values $v_1^a,\ldots,v_p^a$ is more important than the collection of attributes' values $v_1^b,\ldots,v_p^b$.*

The relative importance measures developed above apply to condition attributes with discrete domains. For condition attributes with non-discrete domains, we can use the following idea. It consists in discretizing the scales of numerical condition attributes into several ordered intervals $I_1,\ldots,I_s$ that can be mapped into a set of ordered categories labeled, for instance, $v_1,\ldots,v_s$. Different discretizing techniques can be used for this purpose, see for example, [4][14][102]. Then, each of these category labels can be considered as a specific value to the considered condition attributes and the importance measures given in Equations (30) to (33) can be applied with no modification.

4.6.2. Importance of specific values based on characteristics of decision rules and attractiveness of decision classes

First assume that there are $|T|$ decision classes $Cl_1, Cl_2,\ldots,Cl_{|T|}$ where $Cl_{|T|}$ is the most preferred one. Then, we need to assign a numerical weight w_t to every decision class Cl_t ($t = 1,\ldots,|T|$). In this paper, we assume that the weights are computed as follows:

$$w_1 \quad \geq \quad 0, \tag{34}$$

$$w_t \quad = \quad mw_{t-1} + 1, \quad t = 2,\ldots,|T|; \tag{35}$$

where $w > 1$. This computation ensures that a single assignment to decision class Cl_t ($t = 2,\ldots,|T|$) will be greater than a set of m assignments to less preferred decision classes taken together. Indeed, the weights verify the following condition:

$$w_t \quad > \quad mw_{t'}, \quad \forall t > t'. \tag{36}$$

Similar weighing systems have been used in [9] and [43] to combine different similarity degrees associated with a set of ordinal semantic preference relations into an overall score for web services ranking.

Equations (30) and (31) can then be further constrained as follows:

$$z(q|v) \quad = \sum_{\rho \in F(q) \wedge \rho.r_q = v} \frac{a(\rho)}{spec(\rho)} w_{cl(\rho)}, \tag{37}$$

and

$$z'(q|v) \quad = \quad \frac{1}{card(R)} \sum_{\rho \in F(q) \wedge r_q(\rho)=v} \frac{a(\rho)}{spec(\rho)} w_{cl(\rho);} \tag{38}$$

where $v \in V_q$, $\rho.r_q$ is the right-hand side of the elementary condition relative to condition attribute q in decision rule ρ and $cl(\rho)$ is the index of the decision class of decision rule ρ.

Rule 9. *Let $q \in C$ be a condition attribute from C with discrete domain V_q. Let v_1 and v_2 be two attribute values from V_q. Based on the absolute (resp. relative) importance measure of condition attributes' specific values defined based on characteristics of decision rules and attractiveness of decision classes, we can conclude that if $z(q|v_1) > z(q|v_2)$ (resp. $z'(q|v_1) > z'(q|v_2)$), then attribute value v_1 is more important than attribute value v_2.*

Equations (37) and (38) can be extended to a collection of condition attributes as follows:

$$z_e(q_1,\ldots,q_p|v_1,\ldots,v_p) \quad = \quad \sum_{\rho \in F_e(q_1,\ldots,q_p)\ and \bigwedge_{k=1}^{p} \rho.r_{q_k}=v_k} \frac{a(\rho)}{spec(\rho)} w_{cl(\rho)}, \tag{39}$$

and

$$z'_e(q_1,\ldots,q_p|v_1,\ldots,v_p) \quad = \quad \frac{1}{card(R)} \sum_{\rho \in F_e(q_1,\ldots,q_p)\ and \bigwedge_{k=1}^{p} \rho.r_{q_k}=v_k} \frac{a(\rho)}{spec(\rho)} w_{cl(\rho)}; \tag{40}$$

where, for $k = 1,\ldots,p$, $v_k \in V_{q_k}$, $\rho.r_{q_k}$ is the right-hand side of the elementary condition relative to condition attribute q_k in decision rule ρ, and $cl(\rho)$ is the index of the decision class of decision rule ρ.

Rule 10. *Let $\{q_1,\ldots,q_p\} \subseteq C$ be a collection of condition attributes from C with discrete domains $V_1,\ldots,V_p$. Let $v_1^a,\ldots,v_p^a$ and $v_1^b,\ldots,v_p^b$ be two collections of attribute values from $V_1,\ldots,V_p$. Based on the extended absolute (resp. relative) importance measure of collections of condition attributes' specific values defined based on characteristics of decision rules and attractiveness of decision classes, we can conclude that if $z(q_1,\ldots,q_p|v_1^a,\ldots,v_p^a) > z(q_1,\ldots,q_p|v_1^b,\ldots,v_p^b)$ (resp. $z'(q_1,\ldots,q_p|v_1^a,\ldots,v_p^a) > z'(q_1,\ldots,q_p|v_1^b,\ldots,v_p^b)$), then collection of attributes' values $v_1^a,\ldots,v_p^a$ is more important than collection of attributes' values $v_1^b,\ldots,v_p^b$.*

4.7. Characterisation, generalisation and exploitation of importance measures

We first comment on the characterisation and generalisation of importance measures to more than one dataset (Section 4.7.1). Then, we discuss the practical exploitation of these measures (Section 4.7.2).

4.7.1. Characterisation and generalisation

The importance measures introduced in Sections 4.3 to 4.6 have some mathematical properties, which are detailed in Appendix B. Furthermore, the definitions of absolute and relative importance measures presented in Sections 4.3 to 4.6 have been designed for a single dataset. However, it is possible to use simple statistical operators (including minimum, maximum, arithmetic mean (average), or mid-range) to combine importance measures values obtained from different datasets. Let $S_1,\ldots,S_n$ with $n \geq 1$, be a collection of datasets. Let also $\phi(\cdot)$ be any absolute, relative or overall importance measures presented in Sections 4.3 to 4.5. Then, the absolute and relative importance measures obtained from different datasets can be combined as follows:

- minimum value:
$$\phi(q) = \min_t \phi_t(q). \tag{41}$$
- maximum value:
$$\phi(q) = \max_t \phi_t(q). \tag{42}$$

- arithmetic mean:

$$\phi(q) = \frac{1}{n}\sum_{t}\phi_t(q). \tag{43}$$

- geometric mean:

$$\phi(q) = \sqrt[n]{\phi_1(q)\cdot\ldots\cdot\phi_n(q)}. \tag{44}$$

- mid-range:

$$\phi(q) = \frac{\min_t \phi_t(q) + \max_t \phi_t(q)}{2}. \tag{45}$$

where $\phi_t(q)$ is an absolute, relative or overall importance measure obtained based on dataset S_t $(t = 1, \ldots, n)$.

The proposed absolute and relative importance measures can also be combined through weighted-sum aggregation rules as follows:

$$\phi(q) = \sum_{t}\gamma_t\phi_t(q). \tag{46}$$

where γ_t is the quality of approximation associated with dataset S_t $(t = 1, \ldots, n)$.

These generalisation operations can also be applied to the extended absolute and relative importance measures or those relative to condition attributes' specific values introduced in Section 4.6.

4.7.2. Exploitation of importance measures

The numbers $i(q)$, $j(q)$ and $o(q)$ given by Equations (16), (21) and (26), respectively, represent the absolute scores of condition attribute q based on the characteristics of decision rules, the characteristics of attribute reducts or on a combination of both. Absolute scores can be used to evaluate the differences in the importance of condition attributes or to rank them according to their scores. Absolute scores can also be converted, using an appropriate normalization technique, into absolute weights of condition attributes for further analysis.

The numbers $i'(q)$, $j'(q)$ and $o'(q)$ given by Equations (17), (22) and (27), respectively, represent the relative scores of condition attribute q based on the characteristics of decision rules, the characteristics of attribute reducts or on a combination of both. Relative scores are less informative than absolute scores but they can be directly used as weights for condition attributes.

The numbers $i_e(q_1, \ldots, q_p)$, $j_e(q_1, \ldots, q_p)$ and $o_e(q_1, \ldots, q_p)$, given by Equations (18), (24) and (28), respectively, and $i'_e(q_1, \ldots, q_p)$, $j'_e(q_1, \ldots, q_p)$ and $o'_e(q_1, \ldots, q_p)$, given by Equations (19), (25) and (29), respectively, are the extended versions of absolute and relative importance scores. They can be used to compare the role played by different collections of attributes. They can also be used as a validation tool since, as indicated by [54], if a particular condition attribute is highly important, then the other condition attributes, which are correlated with this attribute, are also likely to be highly important.

5. Case study

5.1. LWC crowdfunding platform

LWC is part of CARE International, an international development organisation (IDO) with a global turnover of 610 million euros in 2011/2012, which runs 927 poverty-fighting and humanitarian aid programmes benefitting 97 million people across 87 countries. The organisation has a long involvement in microfinance, developing a village saving and loan association (VSLA) methodology in 1991, which has subsequently been applied to set up over 54,000 VSLAs serving more than one million members in 21 African countries. In 2008, CARE launched an Access Africa programme offering savings-led microfinance to support "permanent, beneficial social change" which is expected to improve the lives of 30 million people (70% of them women) across 38 countries within 10 years [114]. According to the CARE USA office, in 2017 CARE worked in 93 countries, reaching 63 million people through 950 poverty-fighting development and humanitarian aid programmes. LWC is an innovative form of microfinance donations launched by the British wing of CARE International in 2008. Presently, LWC has 53,871 lenders, 95% from the UK, and total lending since the initiative begun is just over £23 million.

The LWC programme involves four steps. First, the LWC core team identifies new countries in which it wishes to establish a presence.[3] Second, a LWC microfinance expert visits microfinance institutions (MFI) in the country, generally those that are already working with CARE International on an ongoing development project, to review its policies and processes and assess its suitability as a LWC MFI partner. Regular visits to established MFI partners take place to ensure their continued compliance with LWC criteria. Third, approved MFI partners continue with their in-country lending activities, but are now able to forward a cross-selection of their funded entrepreneurs (with an accompanying photograph, business plan and biographic details) to LWC for approval. Fourth, the LWC core team review the proposed entrepreneurs and, if they approve, edit the materials supplied and then post the profiles to the LWC platform. Existing LWC funders or new visitors to the platform can then select the entrepreneur they wish to support and pledge funds using a simple online process. Most loans are pre-disbursed, that is they are given by LWC's local partners and then uploaded onto LWC platform to be "re-financed". However, on occasion, loans are post-disbursed. Here funding is sought from Lendwithcare first—all loans in Vietnam for example are post-disbursed. The funds pledged are transferred to the MFI promoting the entrepreneur at regular intervals, thus allowing the MFI to more rapidly recycle funds to prospective entrepreneurs. Although direct contact between the LWC funder and the funded is not possible, LWC encourage, through their MFI partners, LWC-funded entrepreneurs to provide regular updates on their progress. Capital repayment, without interest or LWC administration charges, occurs generally anywhere between 6 to 48 months, most typically 12 to 18 months, with many LWC funders using the repaid capital to invest in further LWC projects. As funders bear the downside exchange-rate risk, in some instances the capital repaid can be less than the originally invested sum. Following the same argument, the amount of capital repaid can also be larger than the originally invested sum.

5.2. LWC objectives

Crowdfunding is generally defined as the practice of funding a project or venture by raising monetary contributions from a large number of people, typically via the Internet. Hence, computer-human interaction and social technologies are crucial to online crowdfunding. A typical online project description within the LWC platform contains the project, a profile photo of the entrepreneur and other pieces of information related to the entrepreneur as well as the loan requested. The aim of LWC is to extend its activity to new developing countries and increase its effectiveness in reducing poverty across these countries. More specifically, to improve efficiency, LWC aims at:

(1) understanding the motivations of lenders who use online crowdfunding to support entrepreneurs, specifically in developing countries,

(2) analysing the attractiveness of crowdfunding projects,

(3) identifying the relative importance of key factors that investors/potential investors employ when taking their investment decision.

The responses to these aims represent useful operational intelligence to LWC, particularly in supporting the design of marketing advertisements.

While a selection of studies (e.g., [28][41][42][116]) have already provided some general insights into the motivations of contributors to crowdfunding and to the study of the attractiveness of crowdfunding projects, only a few papers (e.g., [29][56]) have addressed the third objective. In a previous study [30], we find empirical evidence that profile photos positively affect pro-social behaviours among contributors to online pro-social crowdfunding campaigns. This paper extends the work in [30] by examining the role played by other characteristics of crowdfunding projects in online pro-social crowdfunding campaigns.

5.3. Considered attributes

The considered condition attributes represent the characteristics of crowdfunding projects. These condition attributes are organized into three groups:

[3] LWC currently has a presence in Cambodia, Ecuador, Malawi, Palestinian Territories, Pakistan, Peru, the Philippines, Rwanda, Vietnam, Zambia and Zimbabwe, and they have two partners in both Ecuador and Zimbabwe, and is currently contemplating expanding its activities to other countries.

- Entrepreneur-related condition attributes:
  - **Entrepreneur Gender** (female or male).
  - **Entrepreneur Age**.[4]
  - **Number of Dependents**.
- Project-related condition attributes:
  - **Project Type** (individual or group).
  - **Number of Participants** (number of individuals participating in the project).
  - **Activity Type** (nature of the activity to be funded such as farming, clothes shop and green loan).
  - **Country** (Country where the project has been launched).
- Loan-related condition attributes:
  - **Loan Requested**.
  - **Largest Loan Value**.
  - **Average Loan Value**.

We note that the Largest Loan Value and Average Loan Value condition attributes relate to the funds actually raised by the crowdfunding campaign at any given time. The description of condition attributes is straightforward. Values of nominal attributes Activity Type and Country are given in Appendix C. Furthermore, some of the attributes (namely Entrepreneur Type, Entrepreneur Gender, Number of Dependents, Activity Type and all the attributes relative to the loan) are directly and easily accessible on the profile description of the project on the LWC platform. The remaining attributes are not directly available and need to be extracted from the textual description of the project.

In terms of the condition attribute Average Loan Value, funders can derive this information using the proportion of funding achieved towards the target, as well as the number of funders that have contributed to the campaign up to that point. The condition attribute Largest Loan Value is included for control purposes.

In addition to condition attributes, we also collected some additional attributes:

- **Number of Lenders**: Number of lenders that supported the project.
- **Number of Supporting Groups**: Number of groups of lenders that supported the project.
- **Added Date**: Date when the project was advertised on the platform.
- **Completion Date**: Date when the project was completely funded.
- **Campaign Duration**: Number of days from the starting of the advertising campaign until full funding of the project.

The attribute "Added Date" is visible on the campaign page. The attributes Number of Lenders and Number of Supporting Groups are also visible on the LWC platform but these are running totals rather than a final tally. The attributes Completion Date and Campaign Duration are not visible on the LWC platform. The values of these attributes have been collected after the full funding of the project. These additional attributes are used to construct the attractiveness classes of crowdfunding projects.

---

[4] In the case of group-type projects, the attribute Entrepreneur Age corresponds to the age of the principal entrepreneur (responsible for the project).

5.4. Specification of condition attributes

The proposed analysis method requires the identification of a set of condition attributes that will be used to evaluate and compare the different projects. The characteristics of the condition attributes used in this case study are summarized in Table 1. These condition attributes can be organized into three groups: personal, project and loan. In addition to the description of each attribute, Table 1 provides also some additional parameters, specifically the preference direction and the scale type. The preference direction indicates the effect of the condition attributes on the assignment of projects to different ordered attractiveness classes. Three cases are possible: (i) gain: an increase on this attribute, will lead to a higher attractiveness; (ii) cost: an increase on this attribute, will lead to a lower attractiveness; or (iii) none: this is for nominal or symbolic condition attributes with no preference structure. The scale type may be nominal, symbolic, ordinal, integer or continuous. For nominal and symbolic condition attributes, we need to also indicate the set of possible values.

Table 1: List of condition attributes

Attribute	Preference	Scale type	Possible values
Entrepreneur Gender	none	nominal	f (female), m (male)
Entrepreneur Age	none	continuous	
Number of Dependents	gain	integer	
Project Type	none	nominal	i (individual), g (group)
Number of Participants	none	integer	
Activity Type	none	nominal	See Appendix C
Country	none	nominal	See Appendix C
Loan Requested	cost	continuous	
Largest Loan Value	gain	continuous	
Average Loan Value	gain	continuous	

It is worth noting that the preference direction of condition attribute Number of Participants has been set to "none". The general assumption is that a high number of individuals can increase the attractiveness of the project. However, one can also see that a larger group may be seen by lenders as more vulnerable than a smaller group, and might therefore be less attractive to lenders.

5.5. Datasets specification

The dataset used in this study relies on the pro-social lending-based crowdfunding platform LWC. The original dataset contains 18,688 entries; each corresponds to a funded project. The original dataset has been randomly divided into exclusive datasets. The assignment procedure is then applied to each dataset, leading to the construction of decision tables. Each of these decision tables is then split into a learning set and a testing set.

5.6. Application and results

For the purpose of illustration, we assume in the rest of this section that $\alpha = 0.75$ and $\beta = 0.5$. The values of these parameters have been set by the authors with the help of a Lendwithcare expert and based on some preliminary analysis. The first step of our approach is to approximate the different learning datasets by using the DRSA. The characteristics of rules and obtained attribute reducts are shown in Table 2. Results from different learning/testing datasets will be aggregated using the arithmetic mean (average). The software jMAF [15] has been used for rough approximation and for computing most of the decision rules basic evaluation measures.

Table 2: Characteristics of rules and reducts generated

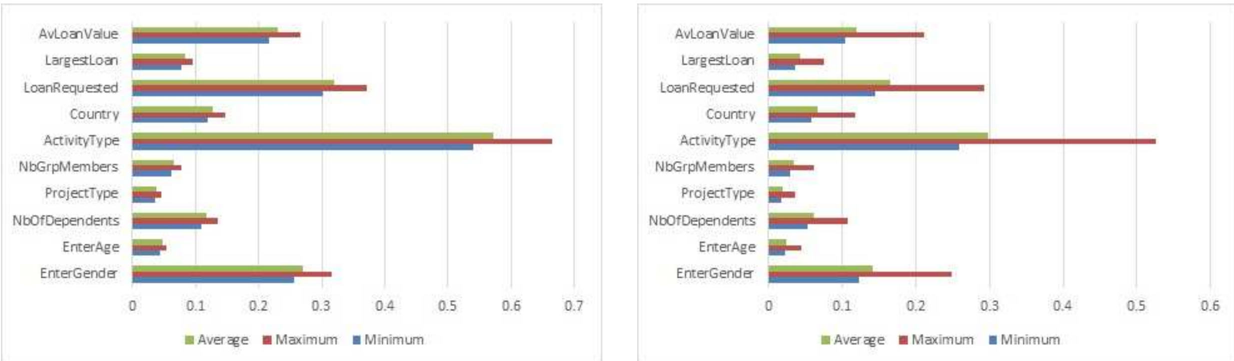
	Quality of Approx.(%)	Rules				Reducts	
		Minimal cover		All rules		Number of reducts	Number of attributes per reduct
		Certain	Possible	Certain	Possible		
Minimum	50	121	215	204	363	2	2
Maximum	99	223	424	618	849	22	10
Average	62	396	713	542	576	13.05	5.56

5.6.1. Relative importance based on the characteristics of decision rules

Equations (16) and (17) have been used to compute the absolute and relative importance measures for all considered condition attributes. These measures have been computed using certain decision rules in the minimal cover and in the set of all decision rules. The results concerning the relative importance measures are summarized in Table 3 and represented graphically in Figure 2. In Figure 2(a), the relative importance measures rely on the decision rules in the minimal cover, while those in Figure 2(b) are computed based on the description of all decision rules. Table 3 and Figure 2 show that Activity Type is the most important attribute, followed by the attributes Loan Requested, Entrepreneur Gender and Average Loan Value with moderate to high relative importance. The attributes Country and Number of Dependents are found to be of low importance. The remaining attributes, namely Largest Loan Value, Number of Participants, Entrepreneur Age and Project Type, show a very low relative importance.

Table 3: Relative importance measures based on certain decision rules

Condition Attribute	Minimal cover			All rules		
	Minimum	Maximum	Average	Minimum	Maximum	Average
Entrepreneur Gender	0.256	0.315	0.271	0.123	0.249	0.141
Entrepreneur Age	0.045	0.055	0.048	0.022	0.044	0.025
Number of Dependents	0.110	0.135	0.117	0.053	0.107	0.061
Project Type	0.037	0.046	0.039	0.018	0.036	0.020
Number of Participants	0.063	0.077	0.067	0.030	0.061	0.035
Activity Type	0.540	0.664	0.572	0.259	0.525	0.298
Country	0.120	0.148	0.127	0.058	0.117	0.066
Loan Requested	0.302	0.371	0.320	0.145	0.293	0.166
Largest Loan Value	0.078	0.096	0.083	0.037	0.076	0.043
Average Loan Value	0.217	0.267	0.230	0.104	0.211	0.120



(**a**) (**b**)

Figure 2: Relative importance measures based on certain decision rules

We also used Equations (18) and (19) to compute the relative importance for pairs of condition attributes based on the characteristics of decision rules. The results are given in Tables 4 and 5 for certain decision rules in the minimal cover and in the set of all decision rules. A careful examination of these tables shows the relative importance of pairs of condition attributes is higher for pairs with higher individual relative importance.

Table 4: Average values of extended relative importance measures for pairs of condition attributes based on certain decision rules in the minimal cover

	Entrepreneur Age	Number of Dependents	Project Type	Number of Participants	Activity Type	Country	Loan Requested	Largest Loan Value	Average Loan Value
Entrepreneur Gender	0.034	0.084	0.028	0.048	0.166	0.092	0.166	0.060	0.166
Entrepreneur Age	-	0.020	0.005	0.011	0.046	0.026	0.040	0.228	0.038
Number of Dependents		-	0.261	0.221	0.122	0.112	0.079	0.064	0.033
Project Type			-	0.193	0.107	0.098	0.069	0.056	0.028
Number of Participants				-	0.092	0.084	0.060	0.048	0.024
Activity Type					-	0.071	0.050	0.041	0.022
Country						-	0.045	0.037	0.016
Loan Requested							-	0.027	0.014
Largest Loan Value								-	0.009

Table 5: Average values of extended relative importance measures for pairs of condition attributes based on all certain decision rules

	Entrepreneur Age	Number of Dependents	Project Type	Number of Participants	Activity Type	Country	Loan Requested	Largest Loan Value	Average Loan Value
Entrepreneur Gender	0.025	0.060	0.020	0.034	0.118	0.065	0.118	0.043	0.118
Entrepreneur Age	-	0.014	0.003	0.008	0.033	0.019	0.029	0.163	0.027
Number of Dependents		-	0.186	0.158	0.087	0.080	0.057	0.046	0.024
Project Type			-	0.138	0.076	0.070	0.050	0.040	0.020
Number of Participants				-	0.065	0.060	0.043	0.034	0.017
Activity Type					-	0.051	0.036	0.029	0.015
Country						-	0.033	0.026	0.011
Loan Requested							-	0.020	0.010
Largest Loan Value								-	0.006

5.6.2. Relative importance based on the characteristics of the reducts and core

Equations (21) and (22) have been used to compute the absolute and relative importance measures for all considered condition attributes based on the characteristics of the reducts. The results concerning the relative importance measures are summarized in Table 6 and illustrated graphically in Figure 3. This result is similar to the one obtained using the rules except for the attribute Largest Loan Value, which is relatively better ranked than in the previous case.

Table 6: Relative importance measures based on attribute reducts

Attribute	Minimum	Maximum	Average
Entrepreneur Gender	0.371	0.472	0.399
Entrepreneur Age	0.065	0.083	0.070
Number of Dependents	0.160	0.203	0.171
Project Type	0.054	0.068	0.058
Number of Participants	0.091	0.116	0.098
Activity Type	0.783	0.996	0.841
Country	0.174	0.221	0.187
Loan Requested	0.438	0.557	0.471
Largest Loan Value	0.113	0.144	0.122
Average Loan Value	0.315	0.400	0.338

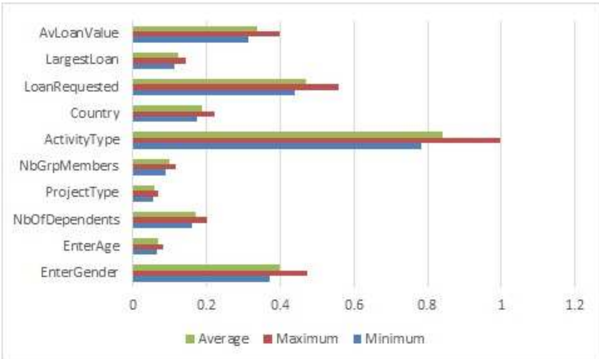


Figure 3: Relative importance measures based on attribute reducts

We also used Equations (24) and (25) to compute the relative importance for pairs of condition attributes based on the characteristics of reducts. The results are given in Table 7. The figures in this table confirm that the relative importance of pairs of condition attributes is higher for pairs with higher individual relative importance values.

Table 7: Average values of extended relative importance measures for pairs of condition attributes based on reducts

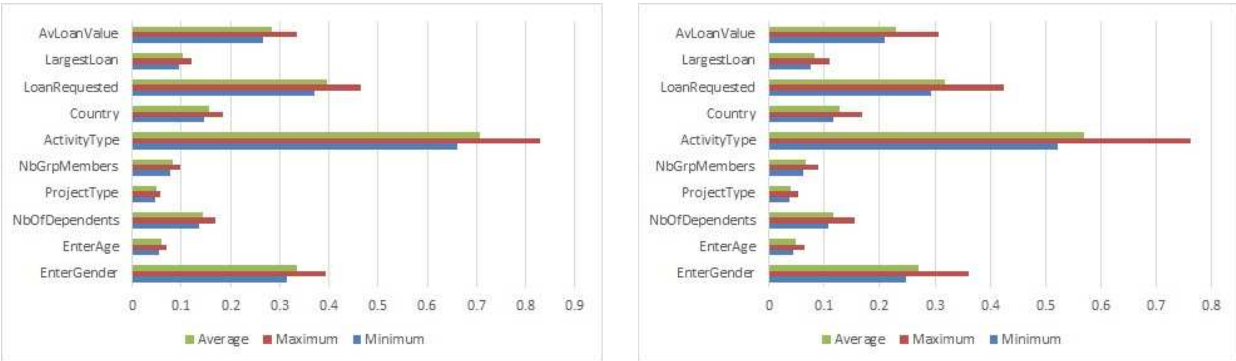
	Entrepreneur Age	Number of Dependents	Project Type	Number of Participants	Activity Type	Country	Loan Requested	Largest Loan Value	Average Loan Value
Entrepreneur Gender	0.012	0.022	0.005	0.012	0.051	0.029	0.045	0.255	0.042
Entrepreneur Age	-	0.249	0.292	0.247	0.137	0.125	0.089	0.072	0.037
Number of Dependents		-	0.012	0.216	0.12	0.11	0.078	0.063	0.032
Project Type			-	0.027	0.103	0.094	0.067	0.054	0.027
Number of Participants				-	0.072	0.08	0.056	0.046	0.024
Activity Type					-	0.051	0.051	0.041	0.018
Country						-	0.137	0.031	0.015
Loan Requested							-	0.089	0.01
Largest Loan Value								-	0.038

5.6.3. Overall relative importance of condition attributes

Equations (26) and (27) have been used to compute the absolute and relative overall importance measures for all considered condition attributes. The results concerning the relative importance measures are summarized in Table 8 and presented as histograms in Figure 4. This result is the same as the one obtained using the characteristics of decision rules.

Table 8: Overall relative importance measures

Condition Attribute	Minimal cover			All rules		
	Minimum	Maximum	Average	Minimum	Maximum	Average
Entrepreneur Gender	0.314	0.394	0.335	0.247	0.361	0.270
Entrepreneur Age	0.055	0.069	0.059	0.044	0.064	0.048
Number of Dependents	0.135	0.169	0.144	0.107	0.155	0.116
Project Type	0.046	0.057	0.049	0.036	0.052	0.039
Number of Participants	0.077	0.097	0.083	0.061	0.089	0.067
Activity Type	0.662	0.830	0.707	0.521	0.761	0.570
Country	0.147	0.185	0.157	0.116	0.169	0.127
Loan Requested	0.370	0.464	0.396	0.292	0.425	0.319
Largest Loan Value	0.096	0.120	0.103	0.05	0.110	0.083
Average Loan Value	0.266	0.334	0.284	0.210	0.306	0.229



(**a**) (**b**)

Figure 4: Overall relative importance measures

We also used Equations (28) and (29) to compute the overall relative importance for pairs of condition attributes. The results are given in Tables 9 and 10. The figures in Table 9 rely on the characteristics of decision rules in the minimal cover for the computation of decision rule-based relative importance measures while those in Table 10 are based on all certain decision rules.

Table 9: Average values of extended overall relative importance measures using the characteristics of decision rules in the minimal cover

	Entrepreneur Age	Number of Dependents	Project Type	Number of Participants	Activity Type	Country	Loan Requested	Largest Loan Value	Average Loan Value
Entrepreneur Gender	0.023	0.053	0.017	0.030	0.109	0.061	0.106	0.158	0.104
Entrepreneur Age	-	0.135	0.149	0.129	0.092	0.076	0.065	0.150	0.038
Number of Dependents		-	0.137	0.219	0.121	0.111	0.079	0.064	0.033
Project Type			-	0.110	0.105	0.096	0.068	0.055	0.028
Number of Participants				-	0.082	0.082	0.058	0.047	0.024
Activity Type					-	0.061	0.051	0.041	0.020
Country						-	0.091	0.034	0.016
Loan Requested							-	0.058	0.012
Largest Loan Value								-	0.024

5.6.4. Importance of specific values of condition attributes

We used the measures introduced in Section 4.6 to evaluate the importance of condition attributes values. First, consider the condition attribute Activity Type, which can take 53 different values (See Appendix C). Table 11 provides the top 6 activity types obtained using (i) characteristics of decision rules only; and (ii) the characteristics of decision rules and attractiveness decision classes. In the first case, the ranking of activity types is very close to their rank based on their respective frequencies in the database (while they are slightly different in the second case).

Table 10: Average values of extended overall relative importance measures using the characteristics of all decision rules

	Entrepreneur Age	Number of Dependents	Project Type	Number of Participants	Activity Type	Country	Loan Requested	Largest Loan Value	Average Loan Value
Entrepreneur Gender	0.019	0.041	0.013	0.023	0.085	0.047	0.082	0.149	0.080
Entrepreneur Age	-	0.132	0.148	0.128	0.085	0.072	0.059	0.118	0.032
Number of Dependents		-	0.099	0.187	0.104	0.095	0.068	0.055	0.028
Project Type			-	0.083	0.090	0.082	0.059	0.047	0.024
Number of Participants				-	0.069	0.070	0.050	0.040	0.021
Activity Type					-	0.051	0.044	0.035	0.017
Country						-	0.085	0.029	0.013
Loan Requested							-	0.055	0.010
Largest Loan Value								-	0.022

Table 11: Analysis of Activity Type

Rank	Case 1	Case 2
1	farming	farming
2	food market stall	green loan
3	sewings/tailoring	raising poultry
4	green loan	sewings/tailoring
5	raising poultry	food market stall
6	clothes shops	clothes shops

We applied the same analysis above to the other three ordinal and symbolic condition attributes, namely, Entrepreneur Gender, Entrepreneur Type and Country. Concerning Entrepreneur Gender, the obtained results show that projects initiated by females experienced a relatively higher attractiveness level than projects initiated by males. For Entrepreneur Type, we observed that projects with pairs and teams are more attractive than projects with individuals. Concerning condition attribute Country, the analysis shows that some countries (namely Pakistan, Cambodia, Philippines, Ecuador and Vietnam) are more attractive than others, but this seems to be more related to their frequency in the database rather than to a clear preference for a given country.

The investigation of specific values of condition attributes with continuous or integer domains showed that: (i) higher values are preferred to lower or moderate values for the condition attributes Number of Dependents, Number of Participants, Average Loan Value and Largest Loan Value; (ii) lower values are preferred to higher values for condition attribute Loan Requested; and (iii) no possible values for condition attribute Entrepreneur Age are clearly ranked first.

5.7. Summary and discussion

Based on these results, we can state the following facts concerning the role of condition attributes on the attractiveness of crowdfunding projects:

- the predominant role of the condition attribute Activity Type;
- the condition attributes Loan Requested, Entrepreneur Gender and Average Loan Value are also important, but no so important as Activity Type;
- the condition attributes Country and Number of Dependents play a moderate role in the attractiveness of projects; and
- the other condition attributes, namely Largest Loan Value, Number of Participants, Entrepreneur Age and Project Type are unimportant with regards to the attractiveness of crowdfunding projects.

The analysis of activity types showed that farming, green loan, food market stall, sewing/tailoring, raising poultry and clothes shops are the most attractive. We also found that projects initiated by females or by pairs and teams are more attractive than projects initiated by males or by individuals. These results confirm the findings of [39]. We also found that projects located in some countries are more attractive than others. Although this result is concordant with the findings reported in [74][26], the obtained ranking seems to be related to the frequency of the considered countries in the database.

6. Additional case study

The objective of this additional case study is to apply the proposed measures so as to identify the role played by different condition attributes that are based on the creditworthiness of a set of European countries. The data used in this case study are extracted from [23].

6.1. Condition attributes and datasets

In this case study and as described in Table 12, a set of seven condition attributes are considered. Table 12 also shows the decision attribute corresponding to rating of the considered countries given by the rating agency Moody.

Table 12: List of condition and decision attributes

Code	Description	Scale type	Preference	Attribute type
GDP	Gross Domestic Product	Continuous	Gain	Condition
InflRate	Inflation Rate	Continuous	Cost	Condition
PublicDebt	Public Debt	Continuous	Cost	Condition
ExternDebt	External Debt	Continuous	Cost	Condition
FDI	Foreign Direct Investment (as % of GDP)	Continuous	Gain	Condition
CA	Current Account Balance (as % of GDP)	Continuous	Gain	Condition
UnempRate	Unemployment Rate	Continuous	Cost	Condition
Rank	Ranking by Moody	Ordinal	Gain	Decision

The learning and testing sets used in this case study are given in Tables 13 and 14, respectively. These tables provide the assessment of the considered 28 European countries with respect to the seven condition attributes given in Table 12 and to the ranking of these countries as established by Moody.

Table 13: Learning set

#	Country	GDP	InfRate	PublicDebt	ExternDebt	FDI	CA	UnempRate	Rank
1	Austria	13067	3	74.6	200	3.2	2.63	4.9	Aaa
2	Belgium	382692	0.34	99.6	266	-0.6	-1.84	8.5	Aa3
3	Bulgaria	39940	-0.8	17.9	90	3.5	1.77	13	Baa2
4	Croatia	43128	-0.1	52.1	99	1	1.24	16	Baa3
5	Cyprus	16504	-0.58	80.9	129	2.8	-1.93	16.9	B3
6	Denmark	248975	0.8	45.3	180	0.5	7.12	6.6	Aaa
7	Estonia	18613	-0.4	6	87	3.9	-1.21	7.7	A1
8	Finland	193443	0.8	53.5	155	-2	-0.92	8.4	Aaa
9	Germany	2737600	0.85	79.9	142	1.4	6.86	5.2	Aaa
10	Greece	182054	-0.7	161.3	174	1.2	0.58	26.8	Caa
11	Italy	1560024	0.09	126.1	108	0.6	0.97	12.6	Baa2
12	Lithuania	34631	0.3	40.2	80	1.6	1.47	11.9	Baa1
13	Luxembourg	45478	1	18.4	3443	50	5.26	6.1	Aaa
14	Malta	7263	0.6	77	72	-19.4	0.88	6.9	A3
15	Poland	38965	2.2	53.8	72.6	0.5	11.2	9.8	Aaa
16	Portugal	165690	-0.2	129	223	-0.9	-1.35	14.8	A2
17	Romania	142245	-0.9	37.2	67	3.5	0.51	7.2	Ba2
18	Slovenia	35275	0	53.2	47.6	-0.9	6.14	9.6	Ba1
19	Spain	1022988	-0.85	85.3	167	3.2	0.77	25.2	Baa2
20	UK	1899098	1.2	90	406	1.8	-4.26	6.6	Aa1

Table 14: Testing set

#	Country	GDP	InfRate	PublicDebt	ExternDebt	FDI	CA	UnempRate	Rank
21	Netherlands	602658	0.89	68.7	73	3.8	10.2	7.2	Aaa
22	France	2059852	0.5	89.9	182	0.2	-1.43	10.4	Aa1
23	Hungary	97948	0.1	78.6	115	-3.2	4.12	7.8	Ba1
24	Sweden	420849	0	38.6	47	-0.9	5.96	8	Aaa
25	Czech Republic	149491	0.5	43.9	45	2.4	-1.37	6.6	A1
26	Latvia	23372	0.6	39.2	146	2.8	-0.81	11.5	Baa1
27	Ireland	164050	0.3	118	1008.2	21.5	6.22	12	Baa1
28	Slovakia	72134	1	48.6	68	2.2	-0.94	14	Baa3

6.2. Rough approximation and induction of decision rules

We applied the DRSA method to the learning set using the software jMAF [15]. The result of approximation is given in Table 15. The quality of the approximations and the accuracy of classes are all equal to 1. The obtained attributes and core are shown in Table 16.

We then applied the VC-DomLEM algorithm [18] to infer the decision rules. The obtained rules are given in Table 17. The applied rules have then been used to classify the countries in the testing set. Only 37% of the countries have been correctly classified.

Table 15: Approximation results

Class	Lower approximations	
Cl	At most Cl	At least Cl
Caab	10	-
B3	5, 10	1, 2, 3, 4, 5, 6, 7, 8, 9, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20
B2	5, 10	1, 2, 3, 4, 6, 7, 8, 9, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20
B1	5, 10	1, 2, 3, 4, 6, 7, 8, 9, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20
Ba3	5, 10	1, 2, 3, 4, 6, 7, 8, 9, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20
Ba2	5, 10, 17	1, 2, 3, 4, 6, 7, 8, 9, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20
Ba1	5, 10, 17, 18	1, 2, 3, 4, 6, 7, 8, 9, 11, 12, 13, 14, 15, 16, 18, 19, 20
Baa3	4, 5, 10, 17, 18	1, 2, 3, 4, 6, 7, 8, 9, 11, 12, 13, 14, 15, 16, 19, 20
Baa2	3, 4, 5, 10, 11, 17, 18, 19	1, 2, 3, 6, 7, 8, 9, 11, 12, 13, 14, 15, 16, 19, 20
Baa1	3, 4, 5, 10, 11, 12, 17, 18, 19	1, 2, 6, 7, 8, 9, 12, 13, 14, 15, 16, 20
A3	3, 4, 5, 10, 11, 12, 14, 17, 18, 19	1, 2, 6, 7, 8, 9, 13, 14, 15, 16, 20
A2	3, 4, 5, 10, 11, 12, 14, 16, 17, 18, 19	1, 2, 6, 7, 8, 9, 13, 15, 16, 20
A1	3, 4, 5, 7, 10, 11, 12, 14, 16, 17, 18, 19	1, 2, 6, 7, 8, 9, 13, 15, 20
Aa3	2, 3, 4, 5, 7, 10, 11, 12, 14, 16, 17, 18, 19	1, 2, 6, 8, 9, 13, 15, 20
Aa2	2, 3, 4, 5, 7, 10, 11, 12, 14, 16, 17, 18, 19	1, 6, 8, 9, 13, 15, 20
Aa1	2, 3, 4, 5, 7, 10, 11, 12, 14, 16, 17, 18, 19, 20	1, 6, 8, 9, 13, 15, 20
Aaa	-	1, 6, 8, 9, 13, 15

Table 16: Reducts and core

Reducts	1.	{GDP,InfRate,FDI,CA,UnempRate}
	2.	{GDP,InfRate,PublicDebt,CA,UnempRate}
Core	1.	{GDP,InfRate,CA,UnempRate}

Table 17: Decision rules

#	Rule description
1	**If** (CA $\geq$ 6.86) **then** (Rank $\geq$ Aaa)
2	**If** (UnempRate $\leq$ 6.1) **then** (Rank $\geq$ Aaa)
3	**If** (GDP $\geq$ 193443.0) & (CA $\geq$ -0.92) & (UnempRate $\leq$ 8.4) **then** (Rank $\geq$ Aaa)
4	**If** (UnempRate $\leq$ 6.6) **then** (Rank $\geq$ Aa1)
5	**If** (GDP $\geq$ 193443.0) & (UnempRate $\leq$ 8.4) **then** (Rank $\geq$ Aa1)
6	**If** (GDP $\geq$ 193443.0) & (UnempRate $\leq$ 8.5) **then** (Rank $\geq$ Aa3)
7	**If** (FDI $\geq$ 3.9) **then** (Rank $\geq$ A1)
8	**If** (GDP $\geq$ 165690.0) & (InfRate $\leq$ -0.2) & (UnempRate $\leq$ 14.8) **then** (Rank $\geq$ A2)
9	**If** (UnempRate $\leq$ 6.9) **then** (Rank $\geq$ A3)
10	**If** (FDI $\geq$ 1.6) & (CA $\geq$ 1.47) & (UnempRate $\leq$ 11.9) **then** (Rank $\geq$ Baa1)
11	**If** (PublicDebt $\leq$ 17.9) **then** (Rank $\geq$ Baa2)
12	**If** (GDP $\geq$ 165690.0) & (PublicDebt $\leq$ 129.0) **then** (Rank $\geq$ Baa2)
13	**If** (FDI $\geq$ 1.6) & (CA $\geq$ 1.47) **then** (Rank $\geq$ Baa2)
14	**If** (FDI $\geq$ 1.0) & (CA $\geq$ 1.24) **then** (Rank $\geq$ Baa3)
15	**If** (CA $\geq$ 0.77) **then** (Rank $\geq$ Ba1)
16	**If** (GDP $\geq$ 165690.0) & (UnempRate $\leq$ 14.8) **then** (Rank $\geq$ Ba1)
17	**If** (UnempRate $\leq$ 16.0) **then** (Rank $\geq$ Ba2)
18	**If** (PublicDebt $\leq$ 129.0) **then** (Rank $\geq$ B3)
19	**If** (PublicDebt $\geq$ 161.3) **then** (Rank $\leq$ Caa)
20	**If** (FDI $\leq$ 2.8) & (UnempRate $\geq$ 16.9) **then** (Rank $\leq$ B3)
21	**If** (GDP $\leq$ 142245.0) & (PublicDebt $\geq$ 37.2) & (CA $\leq$ 0.51) **then** (Rank $\leq$ Ba2)
22	**If** (GDP $\leq$ 35275.0) & (FDI $\leq$ -0.9) & (UnempRate $\geq$ 9.6) **then** (Rank $\leq$ Ba1)
23	**If** (FDI $\leq$ 2.8) & (UnempRate $\geq$ 16.0) **then** (Rank $\leq$ Baa3)
24	**If** (UnempRate $\geq$ 16.0) **then** (Rank $\leq$ Baa2)
25	**If** (GDP $\leq$ 39940.0) & (UnempRate $\geq$ 12.6) **then** (Rank $\leq$ Baa2)
26	**If** (InfRate $\geq$ 0.09) & (UnempRate $\geq$ 12.6) **then** (Rank $\leq$ Baa2)
27	**If** (GDP $\leq$ 35275.0) & (PublicDebt $\geq$ 53.2) & (UnempRate $\geq$ 9.6) **then** (Rank $\leq$ Baa2)
28	**If** (GDP $\leq$ 39940.0) & (UnempRate $\geq$ 11.9) **then** (Rank $\leq$ Baa1)
29	**If** (GDP $\leq$ 35275.0) & (UnempRate $\geq$ 9.6) **then** (Rank $\leq$ Baa1)
30	**If** (GDP $\leq$ 142245.0) & (FDI $\leq$ 3.5) & (CA $\leq$ 0.51) **then** (Rank $\leq$ Baa1)
31	**If** (GDP $\leq$ 7263.0) **then** (Rank $\leq$ A3)
32	**If** (CA $\leq$ 6.14) & (UnempRate $\geq$ 9.6) **then** (Rank $\leq$ A2)
33	**If** (GDP $\leq$ 142245.0) & (CA $\leq$ 0.51) **then** (Rank $\leq$ A1)
34	**If** (CA $\leq$ -1.21) & (UnempRate $\geq$ 7.7) **then** (Rank $\leq$ Aa3)
35	**If** (CA $\leq$ -1.21) **then** (Rank $\leq$ Aa1)

6.3. Application of proposed measures

6.3.1. Relative importance with respect to decision rules

To calculate the relative importance of condition attributes with respect to decision rules, we need first to compute the attractiveness of decision rules by using learning and testing datasets. The calculation details of the attractiveness of the decision rules are summarized in Table 18. The attractiveness values for the decision rules are given in the last column of Table 18. We note that these values are obtained using $\alpha = 0.8$. The value of this parameter procures more power to the accuracy and coverage of decision rules with respect to the learning set than those relative to the testing set. This is due to the low accuracy obtained through the testing set.

Table 18: Calculation of the attractiveness of the decision rules

Rule	Learning					Testing			
(ρ)	$acc(\rho)$	$cov(\rho)$	$\mu(\rho)$	$\widehat{\epsilon}(\rho)$	$a_l(\rho)$	$pacc(\rho)$	$pcov(\rho)$	$a_t(\rho)$	$a(\rho)$
1	1	0.500	0.875	0	0.875	1	1	1	0.900
2	1	0.500	0.875	0	0.875	0	0	0	0.700
3	1	0.500	0.875	0	0.875	1	1	1	0.900
4	1	0.714	0.929	0	0.929	1	0.333	0.833	0.910
5	1	0.571	0.893	0	0.893	1	0.667	0.917	0.898
6	1	0.625	0.906	0	0.906	1	0.667	0.917	0.908
7	1	0.222	0.806	0	0.806	1	0.250	0.813	0.807
8	1	0.100	0.775	0	0.775	0	0	0	0.620
9	1	0.545	0.886	0	0.886	1	0.200	0.800	0.869
10	1	0.250	0.813	0	0.813	0	0	0	0.650
11	1	0.133	0.783	0	0.783	0	0	0	0.627
12	1	0.533	0.883	0	0.883	1	0.600	0.900	0.887
13	1	0.267	0.817	0	0.817	1	0.400	0.850	0.823
14	1	0.375	0.844	0	0.844	1	0.400	0.850	0.845
15	1	0.706	0.926	0	0.926	1	0.667	0.917	0.925
16	1	0.412	0.853	0	0.853	1	0.500	0.875	0.857
17	1	0.944	0.986	0	0.986	1	1	1	0.989
18	1	1	1	0	1	1	1	1	1
19	1	1	1	0	1	0	0	0	0.800
20	1	1	1	0	1	0	0	0	0.800
21	1	0.667	0.917	0	0.917	1	1	1	0.933
22	1	0.250	0.813	0	0.813	0	0	0	0.650
23	1	0.600	0.900	0	0.900	0.167	1	0.549	0.830
24	1	0.500	0.875	0	0.875	0	0	0	0.700
25	1	0.250	0.813	0	0.813	0	0	0	0.650
26	1	0.125	0.781	0	0.781	0.500	0.250	0.406	0.706
27	1	0.250	0.813	0	0.813	0	0	0	0.650
28	1	0.333	0.833	0	0.833	0	0	0	0.667
29	1	0.333	0.833	0	0.833	1	0.250	0.813	0.829
30	1	0.222	0.806	0	0.806	1	0.500	0.875	0.819
31	1	0.100	0.775	0	0.775	0	0	0	0.620
32	1	0.818	0.955	0	0.955	1	0.750	0.938	0.951
33	1	0.250	0.813	0	0.813	1	0.400	0.850	0.820
34	1	0.308	0.827	0	0.827	1	0.200	0.800	0.822
35	1	0.357	0.839	0	0.839	1	0.333	0.833	0.838

The calculation details of the relative importance values of condition attributes based on certain decision rules are summarized in Table 19. Set $F(q)$ in this table indicates the collection of decision rules that have condition attribute q in its premise. A value of 1 in this table means that decision rule ρ belongs to $F(q)$ while a value of 0 indicates that $\rho \notin F(q)$. The number $i(q|\rho)$ in Table 19 refers to the value of $i(q)$ with respect to decision rule ρ. The importance values are then computed as $i(q) = \sum_{\rho} i(q|\rho)$ and $i(q) = \frac{1}{|R|} \sum_{\rho} i(q|\rho)$. On the basis of the figures in this table, we can conclude that the most important condition attribute is UnempRate, followed by attributes CA and then GDP. The condition attribute PublicDebt comes next, followed by condition attribute FDI. The attribute InfRate is slightly important while attribute ExternalDebt is not important at all.

6.3.2. Relative importance with respect to attribute reducts

Table 20 provides the calculation details of the relative importance values of condition attributes with respect to attribute reducts. On the basis of the figures in Table 20, we can conclude that the condition attributes GDP, InfRate, CA and UnempRate have the same relative importance value. The condition attributes PublicDebt and FDI are moderately important. Finally, the condition ExternalDebt is not important at all.

6.3.3. Overall relative importance

The overall relative importance values of all condition attributes for $\beta = 0.6$ are given in Table 21. According to this table, the most important condition attribute is UnempRate, followed by relatively less important condition attributes CA and then GDP right after. The condition attribute PublicDebt follows. Condition attributes FDI and InfRate have very low relative importance values. ExternalDebt is not important at all.

6.3.4. Relative importance values using extended measures

We also used the extended measures to compute the relative importance for pairs of condition attributes based on the characteristics of decision rules and attribute reducts. The results are given in Tables 22 and 23, respectively. The figures in Table 22 indicate that (GDP,UnempRate) is the most important pair of condition attributes, followed by the pairs (GDP,CA), (FDI,CA) and (CA,UnempRate). The next important pairs of condition attributes are (InfRate,UnempRate)

Table 19: Relative importance values based on certain decision rules

Rule			F(q)							$i(q \mid \rho)$						
			q							q						
ρ	$a(\rho)$	$spec(\rho)$	GDP	InfRate	PublicDebt	ExternalDebt	FDI	CA	UnempRate	GDP	InfRate	PublicDebt	ExternalDebt	FDI	CA	UnempRate
1	0.9	1	0	0	0	0	0	1	0	0	0	0	0	0	0.9	0
2	0.7	1	1	0	0	0	0	1	1	0.7	0	0	0	0	0.7	0.7
3	0.9	3	0	0	0	0	0	0	1	0	0	0	0	0	0	0.3
4	0.910	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0.910
5	0.898	2	1	0	0	0	0	0	1	0.449	0	0	0	0	0	0.449
6	0.908	2	1	0	0	0	0	0	1	0.454	0	0	0	0	0	0.454
7	0.807	1	0	0	0	0	1	0	0	0	0	0	0	0.807	0	0
8	0.620	3	1	1	0	0	0	0	1	0.207	0.207	0	0	0	0	0.207
9	0.869	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0.869
10	0.650	3	0	0	0	0	1	1	1	0	0	0	0	0.217	0.217	0.217
11	0.627	1	0	0	1	0	0	0	0	0	0	0.627	0	0	0	0
12	0.887	2	1	0	1	0	0	0	0	0.443	0	0.443	0	0	0	0
13	0.823	2	0	0	0	0	1	1	0	0	0	0	0	0.412	0.412	0
14	0.845	2	0	0	0	0	1	1	0	0	0	0	0	0.423	0.423	0
15	0.925	1	0	0	0	0	0	1	0	0	0	0	0	0	0.925	0
16	0.857	2	1	0	0	0	0	0	1	0.429	0	0	0	0	0	0.429
17	0.989	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0.989
18	1	1	0	0	1	0	0	0	0	0	0	1	0	0	0	0
19	0.8	1	0	0	1	0	0	0	0	0	0	0.8	0	0	0	0
20	0.8	2	0	0	0	0	1	0	1	0	0	0	0	0.4	0	0.4
21	0.933	3	1	0	1	0	0	1	0	0.311	0	0.311	0	0	0.311	0
22	0.650	3	1	0	0	0	1	0	1	0.217	0	0	0	0.217	0	0.217
23	0.830	2	0	0	0	0	1	0	1	0	0	0	0	0.415	0	0.415
24	0.7	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0.7
25	0.650	2	1	0	0	0	0	0	1	0.325	0	0	0	0	0	0.325
26	0.706	2	0	1	0	0	0	0	1	0	0.353	0	0	0	0	0.353
27	0.650	3	1	0	1	0	0	0	1	0.217	0	0.217	0	0	0	0.217
28	0.667	2	1	0	0	0	0	0	1	0.333	0	0	0	0	0	0.333
29	0.829	2	1	0	0	0	0	0	1	0.415	0	0	0	0	0	0.415
30	0.819	3	1	0	0	0	0	1	0	0.273	0	0	0	0	0.273	0
31	0.620	1	1	0	0	0	0	0	0	0.620	0	0	0	0	0	0
32	0.951	2	0	0	0	0	0	1	1	0	0	0	0	0	0.476	0.476
33	0.820	2	1	0	0	0	0	1	0	0.410	0	0	0	0	0.410	0
34	0.822	2	0	0	0	0	0	1	1	0	0	0	0	0	0.411	0.411
35	0.838	1	0	0	0	0	0	1	0	0	0	0	0	0	0.838	0
									$i(q)$	5.802	0.560	3.398	0	2.889	6.294	9.783
									$i'(q)$	0.166	0.016	0.097	0	0.083	0.180	0.280

Table 20: Relative importance values based on attribute reducts

Reduct		$j(q \mid rd)$						
		q						
rd	$spec(rd)$	GDP	InfRate	PublicDebt	ExternalDebt	FDI	CA	UnempRate
1	5	0.2	0.2	0	0	0.2	0.2	0.2
2	5	0.2	0.2	0.2	0	0	0.2	0.2
	$j(q)$	0.2	0.2	0.1	0	0.1	0.2	0.2
	$j'(q)$	0.2	0.2	0.1	0	0.1	0.2	0.2

Table 21: Overall relative importance values

Attribute q	$i(q)$	$i'(q)$	$j(q)$	$j'(q)$	$o(q)$	$o'(q)$
GDP	5.802	0.166	0.4	0.2	3.641	0.179
InfRate	0.56	0.016	0.4	0.2	0.496	0.09
PublicDebt	3.398	0.097	0.2	0.1	2.119	0.098
ExternalDebt	0	0	0	0	0	0
FDI	2.889	0.083	0.2	0.1	1.814	0.09
CA	6.294	0.180	0.4	0.2	3.936	0.188
UnempRate	9.783	0.28	0.4	0.2	6.03	0.248

and (GDP,CA). The other pairs of condition attributes have very low relative importance. Table 23 is less informative about the relative importance of condition attributes since we have a reduced number of attribute reducts.

Table 22: Extended overall relative importance measures by using the characteristics of decision rules

	InfRate	PublicDebt	ExternalDebt	FDI	CA	UnempRate
GDP	0.006	0.028	0	0.014	0.037	0.284
InfRate	-	0	0	0	0	0.016
PublicDebt		-	0	0	0	0.006
ExternalDebt			-	0	0	0
FDI				-	0.038	0.029
CA					-	0.034

Table 23: Extended overall relative importance measures by using the characteristics of attribute reducts

	InfRate	PublicDebt	ExternalDebt	FDI	CA	UnempRate
GDP	0.2	0.1	0	0.1	0.2	0.2
InfRate	-	0.1	0	0.1	0.2	0.2
PublicDebt		-	0	0	0.1	0.1
ExternalDebt			-	0	0	0
FDI				-	0.1	0.1
CA						0.2

7. Conclusion and future work

In this paper, we proposed an analysis strategy and a collection of measures for calculating the relative importance of attributes based on a posteriori information of DRSA [48][49][104][105], namely a collection of if-then decision rules and a collection of reducts and a core. The condition and decision attributes in DRSA are assumed to be preference ordered. It is possible then to exploit this monotonic property of condition and decision attributes in order to induce the relative importance of condition attributes through the analysis of the characteristics of decision rules and reducts/core subsets. The basic assumption beyond this is that attributes that are more important should appear more frequently in the condition parts of decision rules and also in the reducts/core subsets than do less important attributes. Relying on this assumption, several measures have been proposed for calculating the relative importance of condition attributes using the characteristics of decision rules and of attribute reducts. These measures are parameterised in order to consider the characteristics of decision rules and attribute reducts using both learning and testing datasets. The proposed measures can be aggregated into a comprehensive measure indicating the overall importance of each condition attribute. Furthermore, the proposed measures are extended in order to compute the relative importance of a collection of condition attributes taken together.

The proposed analysis strategy and measures have been applied using a real-world case study that has been conducted in partnership with the pro-social lending-based crowdfunding platform LWC. The dataset contains information on observed real-world patterns of pro-social behaviour taken over several years. Data have been collected through the online crowdfunding platform of LWC, where a description of projects can be published and then accessed by different investors. A project description generally consists of a profile photo of the entrepreneur and additional pieces of information concerning the entrepreneur, the project and the loan requested. The main objective of the study was to help LWC identify the relative importance of the factors contained in the descriptions of projects that investors/potential investors employ when taking their investment decision. Based on the induced importance values, it has been shown that the condition attribute Activity Type plays a predominant role in the attractiveness of crowdfunding projets. The condition attributes Loan Requested, Entrepreneur Gender and Average Loan Value are also important but substantially less important than Activity Type. It has also been shown that the condition attributes Country and Number of Dependents play a moderate role in the attractiveness of projects while condition attributes Largest Loan Value, Number of Participants, Entrepreneur Age and Project Type play a marginal role. Finally, it has been shown that relative importance of pairs of condition attributes is higher for pairs with higher individual relative importance values. This information is particularly useful to LWC in the sense that it can support the development objectives of LWC, enhancing the design of its online crowdfunding platform, and so may increase the contributions derived from crowdfunding campaigns.

Although the proposed measures have been validated using a dataset from LWC, the findings of the case study may be useful to other pro-social lending platforms operating in developing counties and covering similar project types. In particular, in this case study we found that projects initiated by females or by pairs and teams are more attractive than projects initiated by males or by individuals. These findings, which confirm the findings of previous studies about crowdfunding projects, as in [39], might suggest priorities for pro-social funders and the way they choose to allocate scarce loan capital. Furthermore, the proposed measures are generic and can be applied with no modification to any other dataset. In particular, we have also applied these measures using the creditworthiness of a set of European countries dataset.

Several topics need to be investigated in the future. The first topic concerns both the application and validation of the proposed measures to other case studies. In particular, applications that require the consideration of a larger set of condition attributes are particularly worth exploring. More importantly, it would be very interesting to use learning datasets where the decision classes are explicitly defined by experts. The second avenue for future research concerns the use of multicriteria clustering techniques to deduce the decision classes. Although the assignment procedure is based on the most elementary preference information, it may lead to a high number of decision classes. Some simple solutions have been proposed at the end of Section 4.2 in order to reduce the number of classes. However, a more advanced solution is to replace the assignment procedure by multicriteria clustering techniques like those proposed in [19][32][33][72][89][90]. The third research area concerns the use of some recent measures of rule interestingness [55][96], especially those based on Bayesian confirmation properties as proposed in [50][52][110]. The last topic concerns the study of the properties of proposed measures such as monotonicity. At this level, one can build on existing work, including [51][68][97][110][113][120], which employ the properties of rule interestingness measures to design new properties more appropriate to relative importance measures as we have proposed in this paper.

Acknowledgements

The authors would like to acknowledge the extensive cooperation and assistance provided by the Lendwithcare team, without whom this study would not have been possible. This work was partially supported by funding from the EPSRC and New Economic Models in the Digital Economy (NEMODE) network as part of the Volunteer and Crowdsourcing Economics project (EP/K039784/1). Special thanks go also to Ajaz Ahmed Khan, Senior Microfinance Advisor at CARE International UK, for his constructive comments on initial versions of this paper. Finally, the authors would like to thank the reviewers for their thoughtful comments and efforts towards improving the manuscript.

References

[1] Abalo, J., Varela, J., Manzano, V., 2007. Importance values for importance—performance analysis: A formula for spreading out values derived from preference rankings. Journal of Business Research 60, 115 – 121.

[2] Agrawal, A., Catalini, C., Goldfarb, A., 2015. Crowdfunding: Geography, social networks, and the timing of investment decisions. Journal of Economics & Management Strategy 24, 253–274.

[3] Ahlers, G., Cumming, D., G unther, C., Schweizer, D., 2015. Signaling in equity crowdfunding. Entrepreneurship Theory and Practice 39, 955–980.

[4] Ali, R., Siddiqi, M., Lee, S., 2015. Rough set-based approaches for discretization: a compact review. Artificial Intelligence Review 44, 235–263.

[5] Amin, T., Chikalov, I., Moshkov, M., Zielosko, B., 2014. Relationships between length and coverage of decision rules. Fundamenta Informaticae 129, 1–13.

[6] Bana e Costa, C., De Corte, J.M., Vansnick, J.C., 2012. MACBETH. International Journal of Information Technology & Decision Making 11, 359–387.

[7] Banzhaf, J., 1965. Weighted Voting Doesn't Work: A Mathematical Analysis. Rutgers Law Review 19, 317–343.

[8] Beaulieu, T., Sarker, S., Sarker, S., 2015. A conceptual framework for understanding crowdfunding. Communications of the Association for Information Systems 37, 1–31.

[9] Bellur, U., Kulkarni, R., 2007. Improved matchmaking algorithm for semantic web services based on bipartite graph matching, in: IEEE International Conference on Web Services (ICWS 2007), pp. 86–93.

[10] Berzal, F., Blanco, I., Sánchez, D., Vila, M., 2002. Measuring the accuracy and interest of association rules: A new framework. Intelligent Data Analysis 6, 221–235.

[11] Bi, G., Geng, B., Liu, L., 2019. On the fixed and flexible funding mechanisms in reward-based crowdfunding. European Journal of Operational Research 279, 168–183.

[12] Bisdorff, R., Meyer, P., Veneziano, T., 2009. Inverse analysis from a condorcet robustness denotation of valued outranking relations, in: Rossi, R., Tsoukias, A. (Eds.), Algorithmic Decision Theory, Springer Berlin Heidelberg, Berlin, Heidelberg. pp. 180–191.

[13] Bisdorff, R., Meyer, P., Veneziano, T., 2014. Elicitation of criteria weights maximising the stability of pairwise outranking statements. Journal of Multi-Criteria Decision Analysis 21, 113–124.

[14] Blajdo, P., Grzymala-Busse, J., Zdzislaw, H., Knap, M., Mroczek, T., Piatek, L., 2008. A comparison of six approaches to discretization—a rough set perspective, in: Wang, G., Li, T., Grzymala-Busse, J., Miao, D., Skowron, A., Yao, Y. (Eds.), Rough Sets and Knowledge Technology, Springer Berlin Heidelberg, Berlin, Heidelberg. pp. 31–38.

[15] Błaszczyński, J., Greco, S., Matarazzo, B., Słowiński, R., Szeląg, M., 2013. jMAF - Dominance-Based Rough Set Data Analysis Framework. Springer Berlin Heidelberg, Berlin, Heidelberg. pp. 185–209.

[16] Blaszczyński, J., Greco, S., Slowiński, R., 2007. Multi-criteria classification - a new scheme for application of dominance-based decision rules. European Journal of Operational Research 181, 1030–1044.

[17] Błaszczyński, J., Słowiński, R., Szeląg, M., 2010. Probabilistic rough set approaches to ordinal classification with monotonicity constraints, in: Proceedings of the Computational Intelligence for Knowledge-based Systems Design, and 13th International Conference on Information Processing and Management of Uncertainty, Springer-Verlag, Berlin, Heidelberg. pp. 99–108.

[18] Błaszczyński, J., Słowiński, R., Szeląg, M., 2011. Sequential covering rule induction algorithm for variable consistency rough set approaches. Information Sciences 181, 987–1002.

[19] Boujelben, M., De Smet, Y., 2016. A multicriteria ordered clustering algorithm to determine precise or disjunctive partitions. International Journal of Multicriteria Decision Making 6, 157–187.

[20] Bouncken, R., Komorek, M., Kraus, S., 2015. Crowdfunding: The current state of research. International Business & Economics Research Journal 14, 407–416.

[21] Budescu, D., 1993. Dominance analysis: A new approach to the problem of relative importance of predictors in multiple regression. Psychological Bulletin 114, 542–551.

[22] Budescu, D., Azen, R., 2004. Beyond global measures of relative importance: Some insights from dominance analysis. Organizational Research Methods 7, 341–350.

[23] Chakhar, S., Ishizaka, A., Labib, A., Saad, I., 2016. Dominance-based rough set approach for group decisions. European Journal of Operational Research 251, 206–224.

[24] Chakhar, S., Saad, I., 2012. Dominance-based rough set approach for groups in multicriteria classification. Decision Support Systems 54, 372–380.

[25] Chakhar, S., Saad, I., 2014. Incorporating stakeholders' knowledge in group decision-making. Journal of Decision Systems 23, 113–126.

[26] Chan, C., Park, H.D., Patel, P., Gomulya, D., 2018. Reward-based crowdfunding success: decomposition of the project, product category, entrepreneur, and location effects. Venture Capital 20, 285–307.

[27] Choo, E., Schoner, B., Wedley, W., 1999. Interpretation of criteria weights in multicriteria decision making. Computers & Industrial Engineering 37, 527–541.

[28] Cox, J., Nguyen, T., Kang, S., 2018a. The kindness of strangers? an investigation into the interaction of funder motivations in online crowdfunding campaigns. Kyklos 71, 187–212.

[29] Cox, J., Nguyen, T., Thorpe, A., Ishizaka, A., Chakhar, S., Meech, L., 2016. A decision rule approach for analysing the attractiveness of crowdfunding projects, in: The OR58 Annual Conference. The OR Society, Portsmouth, UK, pp. 1–20.

[30] Cox, J., Nguyen, T., Thorpe, A., Ishizaka, A., Chakhar, S., Meech, L., 2018b. Being seen to care: The relationship between self-presentation and contributions to online pro-social crowdfunding campaigns. Computers in Human Behavior 83, 45–55.

[31] Dau, H.N., Chakhar, S., Ouelhadj, D., Abubahia, A.M., 2020. Construction and refinement of preference ordered decision classes, in: Ju, Z., Yang, L., Yang, C., Gegov, A., Zhou, D. (Eds.), Advances in Computational Intelligence Systems, Springer International Publishing, Cham. pp. 248–261.

[32] De Smet, Y., Guzmán, L., 2004. Towards multicriteria clustering: An extension of the k-means algorithm. European Journal of Operational Research 158, 390–398.

[33] De Smet, Y., Nemery, P., Selvaraj, R., 2012. An exact algorithm for the multicriteria ordered clustering problem. Omega 40, 861–869.

[34] Dias, L., Clímaco, J., 2000. ELECTRE TRI for groups with imprecise information on parameter values. Group Decision and Negotiation 9, 355–377.

[35] Dias, L., Mousseau, V., Figueira, J., Clímaco, J., 2002. An aggregation/disaggregation approach to obtain robust conclusions with ELECTRE TRI. European Journal of Operational Research 138, 332–348.

[36] Du, S., Peng, J., Nie, T., Yu, Y., 2020. Pricing strategies and mechanism choice in reward-based crowdfunding. European Journal of Operational Research URL: `http://www.sciencedirect.com/science/article/pii/S0377221720300424`, doi:`https://doi.org/10.1016/j.ejor.2020.01.021`. (in press).

[37] Du, W.S., Hu, B.Q., 2018. A fast heuristic attribute reduction approach to ordered decision systems. European Journal of Operational Research 264, 440–452.

[38] Figueira, J., Roy, B., 2002. Determining the weights of criteria in the ELECTRE type methods with a revised Simos' procedure. European Journal of Operational Research 139, 317–326.

[39] Frydrych, D., Bock, A., Kinder, T., Koeck, B., 2014. Exploring entrepreneurial legitimacy in reward-based crowdfunding. Venture Capital 16, 247–269.

[40] Geng, L., Hamilton, H., 2006. Interestingness measures for data mining: A survey. ACM Computing Surveys 38. Article No. 9.

[41] Gerber, E., Hui, J., 2013. Crowdfunding: Motivations and deterrents for participation. ACM Transactions on Computer-Human Interaction 20, 1–32.

[42] Gerber, E., Hui, J., Kuo, P., 2012. Crowdfunding: Why people are motivated to post and fund projects on crowdfunding platforms, in: Szczuka, M., Kryszkiewicz, M., Ramanna, S., Jensen, R., Hu, Q. (Eds.), International Workshop on Design Influence and Social Technologies: Techniques, Impacts and Ethics, ., ACM, Seattle, WA.

[43] Gmati, F.G., Yacoubi Ayadi, N., Chakhar, S., 2014. Parameterized algorithms for matching and ranking web services, in: Meersman, R., Panetto, H., Dillon, T., Missikoff, M., Liu, L., Pastor, O., Cuzzocrea, A., Sellis, T. (Eds.), On the Move to Meaningful Internet Systems: OTM 2014 Conferences, Springer Berlin Heidelberg, Berlin, Heidelberg. pp. 784–791.

[44] Golmohammadi, A., Ghareneh, N., Keramati, A., Jahandideh, B., 2011. Importance analysis of travel attributes using a rough set-based neural network: The case of Iranian tourism industry. Journal of Hospitality and Tourism Technology 2, 155–171.

[45] Grabisch, M., 1996. The application of fuzzy integrals in multicriteria decision making. European Journal of Operational Research 89, 445–456.

[46] Grabisch, M., 1997. k-order additive discrete fuzzy measures and their representation. Fuzzy Sets and Systems 92, 167–189.

[47] Greco, S., Matarazzo, B., Slowinski, R., 1998. Fuzzy measures technique for rough set analysis, in: Proceedings of the Sixth European Congress on Intelligent Techniques and Soft Computing, Aachen, Germany. p. 99–103.

[48] Greco, S., Matarazzo, B., Słowiński, R., 1999. The use of rough sets and fuzzy sets in MCDM, in: Gal, T., Hanne, T., Stewart, T. (Eds.), Multicriteria decision making. Springer US. volume 21 of *International Series in Operations Research & Management Science*, pp. 397–455.

[49] Greco, S., Matarazzo, B., Słowiński, R., 2001. Rough sets theory for multicriteria decision analysis. European Journal of Operational Research 129, 1–47.

[50] Greco, S., Pawlak, Z., Słowiński, R., 2004. Can bayesian confirmation measures be useful for rough set decision rules? Engineering Applications of Artificial Intelligence 17, 345–361.

[51] Greco, S., Słowiński, R., Szczęch, I., 2012. Properties of rule interestingness measures and alternative approaches to normalization of measures. Information Sciences 216, 1–16.

[52] Greco, S., Słowiński, R., Szczęch, I., 2016. Measures of rule interestingness in various perspectives of confirmation. Information Sciences

346-347, 216–235.

[53] Grzymala-Busse, J., Hu, M., 2001. A comparison of several approaches to missing attribute values in data mining, in: Ziarko, W., Yao, Y. (Eds.), Rough Sets and Current Trends in Computing: Second International Conference, RSCTC 2000 Banff, Canada, October 16–19, 2000 Revised Papers. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 378–385.

[54] Hanczar, B., Courtine, M., Benis, A., Hennegar, C., Clément, K., Zucker, J., 2003. Improving classification of microarray data using prototype-based feature selection. ACM SIGKDD Explorations Newsletter 5, 23–30.

[55] Hébert, C., Crémilleux, B., 2007. A unified view of objective interestingness measures, in: Proceedings of the 5th International Conference on Machine Learning and Data Mining in Pattern Recognition, Springer-Verlag, Berlin, Heidelberg. pp. 533–547.

[56] Hoegen, A., Steininger, D., Veit, D., 2018. How do investors decide? an interdisciplinary review of decision-making in crowdfunding. Electronic Markets 28, 339–365.

[57] Hongjiu, L., Yanrong, H., 2015. An evaluating method with combined assigning-weight based on maximizing variance. Scientific Programming , 527–541Article ID 290379.

[58] Hyun, J., Ha, H., 2013. The dominance analysis approach for weighting the dimensions of customer satisfaction. Journal of Korea Service Management Society 14, 53–72.

[59] Ilczuk, G., Wakulicz-Deja, A., 2007. Selection of important attributes for medical diagnosis systems, in: Peters, J., Skowron, A., Marek, V., Orłowska, E., Słowiński, R., Ziarko, W. (Eds.), Transactions on Rough Sets VII: Commemorating the Life and Work of Zdzisław Pawlak, Part II. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 70–84.

[60] Klement, W., Wilk, S., Michalowski, M., Farion, K., 2010. Experienced physicians and automatic generation of decision rules from clinical data, in: Szczuka, M., Kryszkiewicz, M., Ramanna, S., Jensen, R., Hu, Q. (Eds.), Rough Sets and Current Trends in Computing, Springer Berlin Heidelberg, Berlin, Heidelberg. pp. 207–216.

[61] Klimova, O., Noghin, V., 2006. Using interdependent information on the relative importance of criteria in decision making. Computational Mathematics and Mathematical Physics 46, 2080–2091.

[62] Kumar, M., Kee, F., Manshor, A., 2009. Determining the relative importance of critical factors in delivering service quality of banks: An application of dominance analysis in SERVQUAL model. Managing Service Quality: An International Journal 19, 211–228.

[63] Lavrac, N., Flach, P., Zupan, B., 1999. Rule evaluation measures: A unifying view, in: Proceedings of the 9th International Workshop on Inductive Logic Programming, Springer-Verlag, Berlin, Heidelberg. pp. 174–185.

[64] LeBreton, J., Tonidandel, S., 2008. Multivariate relative importance: Extending relative weight analysis to multivariate criterion spaces. Journal of Applied Psychology 93, 329–345.

[65] Lenca, P., Meyer, P., Vaillant, B., Lallich, S., 2008. On selecting interestingness measures for association rules: User oriented description and multiple criteria decision aid. European Journal of Operational Research 184, 610–626.

[66] Liu, W., Zhai, Y., 2013. A new attribute decision making model based on attribute importance. Technology and Investment 4, 224–228.

[67] Liu, Y., Esseghir, M., Boulahia, L., 2016. Evaluation of parameters importance in cloud service selection using rough sets. Applied Mathematics 7, 527–541.

[68] Maddouri, M., Gammoudi, J., 2007. On semantic properties of interestingness measures for extracting rules from data, in: Proceedings of the 8th International Conference on Adaptive and Natural Computing Algorithms, Part I, Springer-Verlag, Berlin, Heidelberg. pp. 148–158.

[69] Maji, P., Paul, S., 2011. Rough set based maximum relevance-maximum significance criterion and gene selection from microarray data. International Journal of Approximate Reasoning 52, 408–426.

[70] Mallik, S., Zhongming, Z., 2017. ConGEMs: Condensed gene co-expression module discovery through rule-based clustering and its application to carcinogenesis. Genes 9, 7. Doi:10.3390/genes9010007.

[71] McGarry, K., 2005. A survey of interestingness measures for knowledge discovery. The Knowledge Engineering Review 20, 39–61.

[72] Meyer, P., Olteanu, A.L., 2013. Formalizing and solving the problem of clustering in MCDA. European Journal of Operational Research 227, 494 – 502.

[73] Michalski, R., 1980. Pattern recognition as rule-guided inductive inference. IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-2, 349–361.

[74] Mollick, E., 2014. The dynamics of crowdfunding: An exploratory study. Journal of Business Venturing 29, 1–16.

[75] Mousseau, V., 1995. Eliciting Information Concerning the Relative Importance of Criteria. Springer US, Boston, MA. pp. 17–43.

[76] Murofushi, T., Soneda, S., 1993. Techniques for reading fuzzy measures (iii): interaction index, in: Proceedings of the Ninth Fuzzy Systems Symposium, Sapporo, Japan. pp. 693–696. (in Japanese).

[77] Noghin, V., 1997. Relative importance of criteria: a quantitative approach. Journal of Multi-Criteria Decision Analysis 6.

[78] Noghin, V., 2001. What is the relative importance of criteria and how to use it in MCDM, in: Köksalan, M., Zionts, S. (Eds.), Multiple Criteria Decision Making in the New Millennium: Proceedings of the Fifteenth International Conference on Multiple Criteria Decision Making (MCDM) Ankara, Turkey, July 10–14, 2000. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 59–68.

[79] Ochin, S., Kumar, S., Joshi, N., 2016. Rule power factor: A new interest measure in associative classification. Procedia Computer Science 93, 12–18.

[80] Pal, S., Skowron, A. (Eds.), 1999. Rough-Fuzzy Hybridization: A New Trend in Decision Making. 1st ed., Springer-Verlag, Berlin, Heidelberg.

[81] Pamučar, D., Stević, Ž., Sremac, S., 2018. A new model for determiningweight coefficients of criteria in MCDM models: Full consistency method (FUCOM). Symmetry 10. `https://www.mdpi.com/2073-8994/10/9/393`.

[82] Pani, A., Kar, A., 2011. A study to compare relative importance of criteria for supplier evaluation in e-procurement, in: 2011 44th Hawaii International Conference on System Sciences, pp. 1–8.

[83] Parhankangas, A., Mason, C., Landstr om, H. (Eds.), 2019. Handbook of Research on Crowdfunding. Edward Elgar Publishing Ltd, Cheltenham. `https://doi.org/10.4337/9781788117210`.

[84] Pawlak, Z., 1982. Rough sets. International Journal of Information & Computer Sciences 11, 341–356.

[85] Pawlak, Z., 1991. Rough set. Theoretical aspects of reasoning about data. Kluwer Academic Publishers, Dordrecht.

[86] Pena, J., g. Nápoles, Salgueiro, Y., 2019. Explicit methods for attribute weighting in multi-attribute decision-making: a review study. Artificial Intelligence Review `https://doi.org/10.1007/s10462-019-09757-w`.

[87] Polkowski, L., Artiemjew, P., 2015. Granular Computing in Decision Approximation: An Application of Rough Mereology. Springer International Publishing, Cham, Switzerland. volume 77 of *Intelligent Systems Reference Library*. chapter Learning Data Classification: Classifiers in General and in Decision Systems. pp. 33–62.

[88] Rissanen, J., 1978. Modeling by shortest data description. Automatica 14, 465–471.

[89] Rocha, C., Dias, L., 2013. MPOC: an agglomerative algorithm for multicriteria partially ordered clustering. 4OR 11, 253–273.

[90] Rocha, C., Dias, L., Dimas, I., 2013. Multicriteria classification with unknown categories: A clustering–sorting approach and an application to conflict management. Journal of Multi-Criteria Decision Analysis 20, 13–27.

[91] Rodriguez-Ricardo, Y., Sicilia, M., López, M., 2018. What drives crowdfunding participation? the influence of personal and social traits. Spanish Journal of Marketing 22, 163–182.

[92] Roubens, M., 1996. Interaction between criteria through the use of fuzzy measures. Technical Report 96.007. Institut de Mathèmatique, Université de Liège,. Liège, Belgium.

[93] Saaty, T., 1980. The Analytic Hierarchy Process. McGraw-Hill, New York.

[94] Salahaldin, L., Angerer, M., Kraus, S., Trabelsi, D., 2019. A duration-based model of crowdfunding project choice. Finance Research Letters 29, 404–410.

[95] Sethi, P., Sathya, A., 2010. Association rule based similarity measures for the clustering of gene expression data. Open Medical Informatics Journal 4, 63–73.

[96] Shaharanee, I., Hadzic, H., Dillon, T., 2011. Interestingness measures for association rules based on statistical validity. Knowledge-Based Systems 24, 386–392.

[97] Shaikh, M., McNicholas, P., Antonie, M., Murphy, T., 2018. Standardizing interestingness measures for association rules. Statistical Analysis and Data Mining: The ASA Data Science Journal 11, 282–295.

[98] Shapley, L., 1953. A value for n-person games, in: Kuhn, H., Tucker, A. (Eds.), Contributions to the Theory of Games II. Princeton University Press, Princeton, pp. 307–317.

[99] Sikora, M., 2013. Redefinition of decision rules based on the importance of elementary conditions evaluation. Fundamenta Informaticae 123, 171–197.

[100] Simos, J., 1990a. Evaluer l'impact sur l'environnement: Une approche originale par l'analyse multicritère et la négociation. Presses Polytechniques et Universitaires Romandes, Lausanne.

[101] Simos, J., 1990b. L'évaluation environnementale: Un processus cognitif négocié. Thèse de doctorat. DGF EPFL, Lausanne.

[102] Singh, G., Minz, S., 2007. Discretization using clustering and rough set theory, in: International Conference on Computing: Theory and Applications (ICCTA'07), pp. 330–336.

[103] Siskos, E., Tsotsolas, N., 2015. Elicitation of criteria importance weights through the Simos method: A robustness concern. European Journal of Operational Research 246, 543–553.

[104] Slowiński, R., Greco, S., Matarazzo, B., 2002. Rough set analysis of preference-ordered data, in: Alpigini, J., Peters, J., Skowron, A., Zhong, N. (Eds.), Rough Sets and Current Trends in Computing. Springer Berlin Heidelberg. volume 2475 of *Lecture Notes in Artificial Intelligence*, pp. 44–59.

[105] Słowiński, R., Greco, S., Matarazzo, B., 2009. Rough sets in decision making, in: Meyers, R. (Ed.), Encyclopedia of Complexity and Systems Science. Springer, New York, pp. 7753–7787.

[106] Solymosi, T., Dombi, J., 1986. A method for determining the weights of criteria: The centralized weights. European Journal of Operational Research 26, 35–41.

[107] Stanczyk, U., 2015. Selection of decision rules based on attribute ranking. Journal of Intelligent & Fuzzy Systems 29, 899–915.

[108] Stańczyk, U., 2016. Weighting and pruning of decision rules by attributes and attribute rankings, in: Czachórski, T., Gelenbe, E., Grochla, K., Lent, R. (Eds.), Computer and Information Sciences, Springer International Publishing, Cham. pp. 106–114.

[109] Stefanowski, J., Vanderpooten, D., 2001. Induction of decision rules in classification and discovery-oriented perspectives. International Journal of Intelligent Systems 16, 13–27.

[110] Susmaga, R., Szczęch, I., 2018. Can confirmation measures reflect statistically sound dependencies in data? the concordance-based assessment. Foundations of Computing and Decision Sciences 43, 41–66.

[111] Szeląg, M., 2015. Application of the Dominance-based Rough Set Approach to Ranking and Similarity-based Classification Problems. Ph.D. thesis. Poznań University of Technology. `http://www.cs.put.poznan.pl/mszelag/Research/MSzPhD.pdf`.

[112] Szeląg, M., Greco, S., Słowiński, R., 2014. Variable consistency dominance-based rough set approach to preference learning in multicriteria ranking. Information Sciences 277, 525 – 552.

[113] Tew, C., Giraud-Carrier, C., Tanner, K., Burton, S., 2014. Behavior-based clustering and analysis of interestingness measures for association rule mining. Data Mining and Knowledge Discovery 28, 1004–1045.

[114] Thomas, A., 2013. Care international annual report. CARE International, Geneva, Switzerland. `https://www.care-international.org/files/files/publications/ CARE-International-Annual-Report-2013.pdf`.

[115] Tonidandel, S., LeBreton, J., 2011. Relative importance analysis: A useful supplement to regression analysis. Journal of Business and Psychology 26, 1–9.

[116] Tosatto, J., Cox, J., Nguyen, T., 2019. Handbook of research on crowdfunding, Edward Elgar Publishing Ltd, Cheltenham. chapter An Overview of Crowdfunding in the Creative and Cultural Industries, pp. 269–302. `https://doi.org/10.4337/9781788117210.00018`.

[117] Tsotsolas, N., Spyridakos, A., Siskos, E., Salmon, I., 2019. Criteria weights assessment through prioritizations (WAP) using linear programming techniques and visualizations. Operational Research 19.

[118] Vansnick, J.C., 1986. On the problem of weights in multiple criteria decision making (the noncompensatory approach). European Journal of Operational Research 24, 288–294.

[119] World Bank, 2013. Crowdfunding's potential for the developing world. *info*Dev, Finance and Private Sector Development Department. Washington, DC: World Bank. `https://www.infodev.org/infodev-files/wb_crowdfundingreport-v12.pdf`.

[120] Wróbel, L., Sikora, M., Michalak, M., 2016. Rule quality measures settings in classification, regression and survival rule induction : an

empirical approach. Fundamenta Informaticae 149, 419–449.

[121] Wu, Z., Ruixia, Y., 2013. Multi-attribute decision making based on attribute importance degree and case-based reasoning. Cybernetics and Information Technologies 13, 62–74.

[122] Yang, J., Liu, L., Yin, C., 2019. A non-liner decision model for green crowdfunding project success: Evidence from China. International Journal of Environmental Research and Public Health 16. URL: `http://www.mdpi.com/1660-4601/16/2/187`. Article #187.

[123] Yao, Y., Zhong, N., 1999. An analysis of quantitative measures associated with rules, in: Zhong, N., Zhou, L. (Eds.), Methodologies for Knowledge Discovery and Data Mining, Springer Berlin Heidelberg, Berlin, Heidelberg. pp. 479–488.

[124] Zhou, L., Jiang, F., 2011. A rough set approach to feature selection based on relative decision entropy, in: Yao, J., Ramanna, S., Wang, G., Suraj, Z. (Eds.), Rough Sets and Knowledge Technology: 6th International Conference, RSKT 2011, Banff, Canada, October 9-12, 2011. Proceedings. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 110–119.

[125] Zielosko, B., Robaszkiewicz, M., 2016. Greedy algorithm for optimization of association rules relative to length, in: Czarnowski, I., Caballero, A., Howlett, R., Jain, J. (Eds.), Intelligent Decision Technologies 2016, Springer International Publishing, Cham. pp. 281–292.

Appendix A. Illustrative example of attractiveness classes construction

This appendix illustrates, through a reduced dataset, the assignment procedure introduced in Section 4.2.

Appendix A.1. An extract from the database

The values for the considered condition attributes have been assessed by LWC and automatically saved in their database. An extract from the database is given Table A.24. This provides the description of 25 randomly selected crowdfunding projects.

Table A.24: Information table

Project	Entrepreneur Gender	Entrepreneur Age	Number of Dependents	Project Type	Number of Participants	Activity Type	Country	Loan Requested	Largest Loan Value	Average Loan Value
1	m	31	2	g	15	Farming	Cambo	755.333	200	22.22
2	f	32	5	i	1	Farming	Vietn	503.268	73	25.16
3	f	45	1	i	1	Raising livestock	Phili	562.857	75	23.45
4	m	40	2	i	1	Trader	Phili	1758.483	150	26.25
5	m	43	11	i	1	Market Stall (Food)	Pakis	142.566	45	17.82
6	m	28	3	g	7	Market Stall	Benin	796.053	300	46.83
7	m	30	4	i	1	Market Stall	Rwand	246.601	75	30.83
8	f	70	3	g	22	Food Production	Benin	714.286	99	28.57
9	m	32	8	i	1	Market Stall	Benin	509.934	90	34.00
10	f	49	5	i	1	Farming	Cambo	631.274	141	25.25

The values of the additional attributes corresponding to the crowdfunding projects in Table A.24 are given in Table A.25.

Table A.25: Values of additional attributes

Project	Added Date	Completion Date	Campaign Duration	Number of Lenders	Number of Supporting Groups
1	26/06/2013	29/06/2013	3	34	0
2	15/11/2014	19/11/2014	4	20	0
3	24/01/2013	31/01/2013	7	24	3
4	24/12/2015	02/01/2016	9	67	0
5	06/08/2015	11/08/2015	5	8	0
6	02/08/2012	13/08/2012	11	17	0
7	21/08/2015	24/08/2015	3	8	0
8	27/03/2012	02/04/2012	6	25	4
9	24/07/2012	01/08/2012	8	15	0
10	01/10/2014	04/10/2014	3	25	2

Appendix A.2. Illustration of assignment procedure

As underlined in Section 4.2, decision classes representing the attractiveness of crowdfunding projects have been computed based on the additional attributes (namely, project advertising campaign duration, number of supporting individuals, and number of supporting groups) using an iterative assignment procedure. For each learning dataset, we first used the assignment procedure introduced in Section 4.2 to assign the different crowdfunding projects into different decision classes. For illustration purposes, let us consider the data given in Section Appendix A.1. The data in Table A.25, which shows the dates when the project is added to LWC platform, the date when the project is fully funded and the total number of supporting lenders, has been used to compute the decision classes of the crowdfunding projects shown in Table A.24. First, the data in Table A.25 has been used to compute the dominance relation between the different crowdfunding projects. The result is summarized in the dominance matrix given in Table A.26 and the corresponding dominance graph is shown in Figure A.5.

Table A.26: The P-dominance relation

	1	2	3	4	5	6	7	8	9	10
1	1	0	1	0	0	1	1	1	0	1
2	1	0	0	0	0	0	0	0	0	0
3	0	0	1	0	0	1	1	0	0	1
4	0	0	0	1	0	0	1	0	0	1
5	0	1	0	0	1	0	1	0	0	0
6	0	0	0	0	0	1	0	0	0	0
7	0	0	0	0	0	0	1	0	0	0
8	0	0	0	0	1	0	0	0	0	0
9	0	0	0	1	0	0	1	0	0	1
10	0	0	1	0	0	1	1	1	0	1

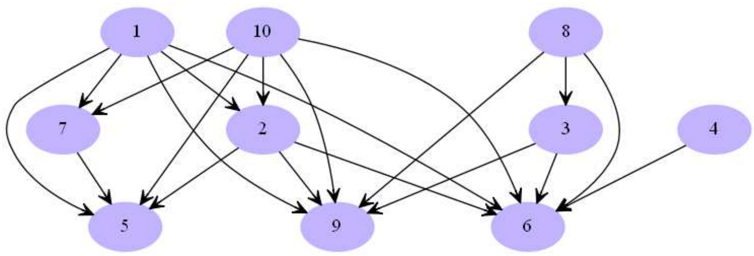


Figure A.5: Dominance graph

The application of the assignment procedure using the data in Table A.25 leads to the dominance graph shown in Figure A.5. The application of the assignment procedure using the obtained graph is illustrated in Figure A.6. In each iteration, a directed graph is constructed by exploiting the dominance relation between the crowdfunding projects. Then, the projects in the kernel of the graph are assigned to the current attractiveness class. The procedure is repeated until the assignment of all crowdfunding projects. The application of the assignment procedure leads to the following decision classes: $Cl_3 = \{1, 4, 8, 10\}$, $Cl_2 = \{2, 3, 7\}$ $Cl_1 = \{5, 6, 9\}$, where Cl_3 is the most preferred decision class and Cl_1 the least preferred decision class.

Appendix B. Characterisation of relative importance measures

Appendix B.1. Properties of Equation (16) and Equation (17)

Properties (1.1—1.5)**.** *Let $q, q_1, q_2 \in C$ and m is the total number of condition attributes. Then, the following properties hold for the absolute and relative importance measures of condition attributes:*

(1.1) $0 \leq i(q) \leq card(R)$ *and* $0 \leq i'(q) \leq 1$, $\forall q \in C$.

(1.2) $F(q) = \emptyset \Rightarrow i(q) = 0 \wedge i'(q) = 0$, $\forall q \in C$.

(1.3) $F(q) \neq \emptyset \Rightarrow i(q) > 0 \wedge i'(q) > 0$, $\forall q \in C$.

(1.4) $F(q_1) \subseteq F(q_2) \Rightarrow i(q_1) \leq i(q_2) \wedge i'(q_1) \leq i'(q_2)$, $\forall q_1, q_2 \in C$.

(1.5) $i(q_1) \geq i(q_2) \Rightarrow i'(q_1) \geq i'(q_2)$, $\forall q_1, q_2 \in C$.

Proof. Properties (1.2), (1.3) and (1.5) are trivial. We prove properties (1.1) and (1.4) for absolute importance measure $i(q)$. Similar proofs apply for the relative importance measure $i'(q)$.

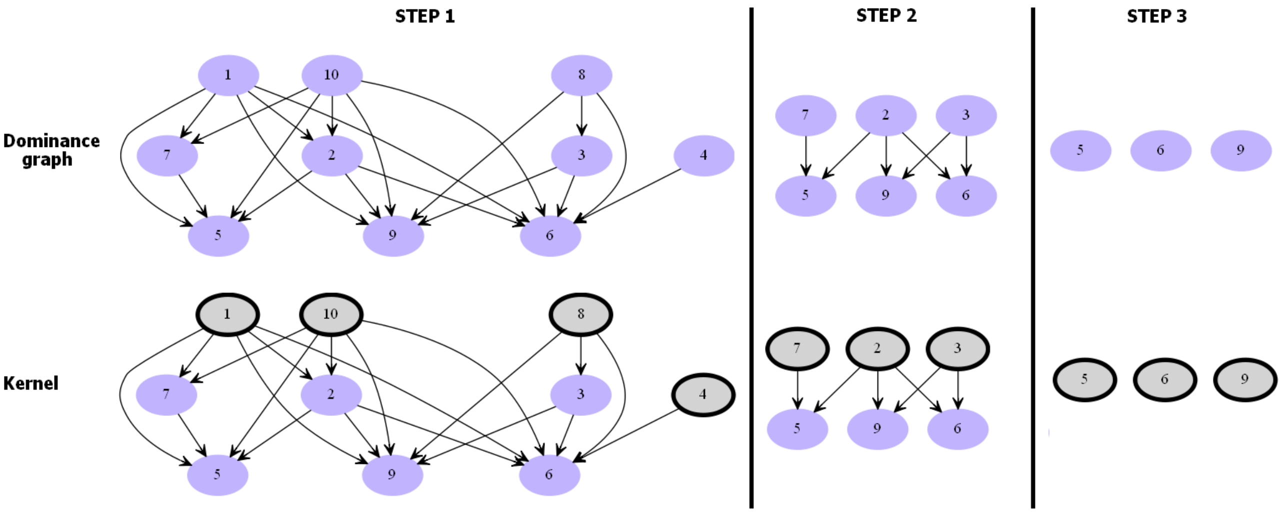


Figure A.6: Graphical illustration of the assignment procedure

(1.1) By definition, we have $a(\rho) \geq 0$ and $spec(\rho) > 0$, which leads to $i(q) \geq 0$. Based on Equation (15), we can establish that $0 \leq a_t(\rho) \leq 1$. Since we have $1 \leq spec(\rho) \leq m$, we get then: $0 \leq \frac{a_t(\rho)}{spec(\rho)} \leq \frac{1}{m}$. The rule attractiveness measure $a(\rho)$ is maximum when $F(q) = card(R)$ and $spec(\rho) = 1, \forall \rho \in R$. This leads to $0 \leq \sum_{\rho \in F(q)} \frac{a(\rho)}{spec(\rho)} \leq card(R)$ and consequently $0 \leq i(q) \leq card(R)$.

(1.4) We have: $F(q_1) \subseteq F(q_2)$. This leads to: $\sum_{\rho \in F(q_1)} \frac{a(\rho)}{spec(\rho)} \leq \sum_{\rho \in F(q_2)} \frac{a(\rho)}{spec(\rho)}$. Consequently, we get $i(q_1) \leq i(q_2)$.

□

Appendix B.2. Properties of Equation (18) and Equation (19)

Properties (2.1—2.5). *Let $q_i, q_i^a, q_i^b \in C$, $1 \leq i \leq p$ with $p \leq m$ and m is the total number of condition attributes. Then, the following properties hold for the extended absolute $i(q_1, \cdots, q_p)$ and relative $i'(q_1, \cdots, q_p)$ importance measures of condition attributes:*

(2.1) $0 \leq i(q_1, \cdots, q_p) \leq card(R)|$ *and* $0 \leq i'(q_1, \cdots, q_p) \leq 1, \forall \{q_1, \cdots, q_p\} \subseteq C$.

(2.2) $F_e(q_1, \cdots, q_p) = \emptyset \Rightarrow i(q_1, \cdots, q_p) = 0 \wedge i'(q_1, \cdots, q_p) = 0, \forall \{q_1, \cdots, q_p\} \subseteq C$.

(2.3) $F_e(q_1, \cdots, q_p) \neq \emptyset \Rightarrow i(q_1, \cdots, q_p) > 0 \wedge i'(q_1, \cdots, q_p) > 0, \forall \{q_1, \cdots, q_p\} \subseteq C$.

(2.4) $F_e(q_1^a, \cdots, q_p^a) \subseteq F(q_1^b, \cdots, q_p^b) \Rightarrow i(q_1^a, \cdots, q_p^a) \leq i(q_1^b, \cdots, q_p^b) \wedge i'(q_1^a, \cdots, q_p^a) \leq i'(q_1^b, \cdots, q_p^b)$, $\forall \{q_1^a, \cdots, q_p^a\} \subseteq C, \{q_1^b, \cdots, q_p^b\} \subseteq C$.

(2.5) $i(q_1^a, \cdots, q_p^a) \geq i(^b_1, \cdots, q_p^b) \Rightarrow i'(q_1^a, \cdots, q_p^a) \geq i'(^b_1, \cdots, q_p^b), \forall \{q_1^a, \cdots, q_p^a\} \subseteq C, \{q_1^b, \cdots, q_p^b\} \subseteq C$.

Proof. Properties (2.2), (2.3) and (2.5) are trivial. We prove properties (2.1) and (2.4) for extended absolute importance measure $i(q)$. Similar proofs apply for the extended relative importance measure $i'(q)$.

(2.1) Similar to the proof of property (1.1) where q is replaced by $q_1, \cdots, q_p$.

(2.4) Similar to the proof of property (1.4), where q_1 and q_2 are replaced by $q_1^a, \cdots, q_p^a$ and $q_1^b, \cdots, q_p^b$, respectively.

□

Appendix B.3. Properties of Equation (21) and Equation (22)

Properties (3.1—3.5)**.** *Let $q, q_1, q_2 \in C$. Then, the following properties hold for the extended absolute $j(q)$ and relative $j'(q)$ importance measures of condition attributes:*

(3.1) $0 \leq j(q) \leq card(RED)$ *and* $0 \leq j'(q) \leq 1$, $\forall q \in C$.

(3.2) $K(q) = \emptyset \Rightarrow j(q) = 0 \land j'(q) = 0$, $\forall q \in C$.

(3.3) $K(q) \neq \emptyset \Rightarrow j(q) > 0 \land j'(q) > 0$, $\forall q \in C$.

(3.4) $K(q_1) \subseteq K(q_2) \Rightarrow j(q_1) \leq j(q_2) \land j'(q_1) \leq j'(q_2)$, $\forall q_1, q_2 \in C$.

(3.5) $j(q_1) \geq j(q_2) \Rightarrow j'(q_1) \geq j'(q_2)$, $\forall q_1, q_2 \in C$.

Proof. Properties (3.2), (3.3) and (3.5) are trivial. We prove properties (3.1) and (3.4) for absolute importance measure $j(q)$. Similar proofs apply for the relative importance measure $j'(q)$.

(3.1) By definition, we have $j(q, rd) \geq 0$, $\forall q \in C, rd \in RED$, which leads to $j(q) \geq 0$. Based on Equation (21), we can establish that $j(q)$ is maximum when $j(q, rd)$ is maximum for all $rd \in RED$. The number $j(q, rd)$ is maximal when $spec(rd) = 1$. This leads to $0 \leq \sum_{rd \in K(q)} j(q, rd) \leq card(RED)$.

(3.4) We have: $K(q_1) \subseteq K(q_2)$. Hence, $\sum_{rd \in K(q_1)} j(q_1, rd) \leq \sum_{rd \in K(q_2)} j(q_2, rd)$, which means $j(q_1) \leq j(q_2)$.

□

Properties (4.1—4.3)**.** *Let $q, q' \in C$. Then, the following properties hold:*

(4.1) $q, q' \in COR \neq \emptyset \Rightarrow j(q) = j(q') \land j'(q) = j'(q')$, $\forall q, q' \in C$.

(4.2) $q \in COR \neq \emptyset \Rightarrow j(q) > j(q') \land j'(q) > j'(q')$, $\forall q, q' \in C$ *and* $q' \notin COR$.

(4.3) $COR = \emptyset \Rightarrow j(q) < card(RED) \land j'(q) < 1$, $\forall q, q' \in C$.

Proof. We provide the proof for $j(\cdot)$. The proof for $j'(\cdot)$ is equivalent.

(4.1) Since both q and q' are COR, then they appear in all the reducts, which mean that $j(q)$ and $j(q')$ are equal.

(4.2) If $q \in COR$ and $q' \notin COR$, then $\sum_{rd \in K(q)} j(q, rd) > \sum_{rd \in K(q')} j(q', rd)$ since $card(K(q)) > card(K(q))$. Consequently, $j(q) > j(q')$.

(4.3) If $COR = \emptyset$, then $K(q) < card(RED)$. And since $j(q, rd) \leq 1$, $\forall q \in C$, $rd \in RED$. Then, we get $\sum_{rd \in K(q)} j(q, rd) < card(RED)$. Consequently, $j(q) < card(RED)$.

□

Appendix B.4. Properties of Equation (24) and Equation (25)

Properties (5.1—5.5)**.** *Let $q_i, q_i^a, q_i^b \in C$, $1 \leq i \leq p$ with $p \leq m$ and m is the total number of condition attributes. Then, the following properties hold for the extended absolute $j(q_1, \cdots, q_p)$ and relative $j'(q_1, \cdots, q_p)$ importance measures of condition attributes:*

(5.1) $0 \leq j(q_1, \cdots, q_p) \leq card(RED)$ *and* $0 \leq j'(q_1, \cdots, q_p) \leq 1$, $\forall \{q_1, \cdots, q_p\} \subseteq C$.

(5.2) $K_e(q_1, \cdots, q_p) = \emptyset \Rightarrow j(q_1, \cdots, q_p) = 0 \wedge j'(q_1, \cdots, q_p) = 0$, $\forall \{q_1, \cdots, q_p\} \subseteq C$.

(5.3) $K_e(q_1, \cdots, q_p) \neq \emptyset \Rightarrow j(q_1, \cdots, q_p) > 0 \wedge j'(q_1, \cdots, q_p) > 0$, $\forall \{q_1, \cdots, q_p\} \subseteq C$.

(5.4) $K_e(q_1^a, \cdots, q_p^a) \subseteq K(q_1^b, \cdots, q_p^b) \Rightarrow j(q_1^a, \cdots, q_p^a) \leq j(q_1^b, \cdots, q_p^b) \wedge j'(q_1^a, \cdots, q_p^a) \leq j'(q_1^b, \cdots, q_p^b)$, $\forall \{q_1^a, \cdots, q_p^a\} \subseteq C, \{q_1^b, \cdots, q_p^b\} \subseteq C$.

(5.5) $j(q_1^a, \cdots, q_p^a) \geq j(_1^b, \cdots, q_p^b) \Rightarrow j'(q_1^a, \cdots, q_p^a) \geq j'(_1^b, \cdots, q_p^b), \forall \{q_1^a, \cdots, q_p^a\} \subseteq C, \{q_1^b, \cdots, q_p^b\} \subseteq C$.

Proof. Properties (5.2), (5.3) and (5.5) are trivial. We prove properties (5.1) and (5.4) for extended absolute importance measure $i(q)$. Similar proofs apply for the extended relative importance measure $i'(q)$.

(5.1) Similar to the proof of property (3.1) where q is replaced by $q_1, \cdots, q_p$.

(5.4) Similar to the proof of property (3.4), where q_1 and q_2 are replaced by $q_1^a, \cdots, q_p^a$ and $q_1^b, \cdots, q_p^b$, respectively.

□

Properties (6.1—6.5)**.** *Let $q_i^a, q_i^b \in C$, $1 \leq i \leq p$ with $p \leq m$ and m is the total number of condition attributes. Then, the following properties hold:*

(6.1) $\{q_1^a, \cdots, q_p^a, q'\}, \{q_1^b, \cdots, q_p^b\} \subseteq COR \neq \emptyset \Rightarrow j(q_1^a, \cdots, q_p^a) = j(q_1^b, \cdots, q_p^b) \wedge j'(q_1^a, \cdots, q_p^a) = j'(q')$, $\forall \{q_1^a, \cdots, q_p^a\} \subseteq C, \{q_1^b, \cdots, q_p^b\} \subseteq C$.

(6.2) $q \in COR \neq \emptyset \Rightarrow j(q_1^a, \cdots, q_p^a) > j(q_1^b, \cdots, q_p^b) \wedge j'(_1^a, \cdots, q_p^a) > j'(q_1^b, \cdots, q_p^b)$, $\forall \{q_1^a, \cdots, q_p^a\} \subseteq C, \{q_1^b, \cdots, q_p^b\} \subseteq C$ *and* $q' \notin COR$.

(6.3) $COR = \emptyset \Rightarrow j(q_1^a, \cdots, q_p^a) < card(RED) \wedge j'(q_1^a, \cdots, q_p^a) < 1$, $\forall \{q_1^a, \cdots, q_p^a\} \subseteq C$.

Proof. We provide the proof for $j(\cdots)$. The proof for $j'(\cdots)$ is equivalent.

(6.1) Similar to the proof of property (5.1) where q is replaced by $(q_1^a, \cdots, q_p^a$ and q' by $q_1^b, \cdots, q_p^b$.

(6.2) Similar to the proof of property (5.2), where q_1 and q_2 are replaced by $q_1^a, \cdots, q_p^a$ and $q_1^b, \cdots, q_p^b$, respectively.

(6.3) Similar to the proof of property (5.3) where q and q' are replaced by $q_1^a, \cdots, q_p^a$ and $q_1^b, \cdots, q_p^b$, respectively.

□

Appendix B.5. Properties of Equation (26) and Equation (27)

Properties (7.1—7.2)**.** *Let $q \in C$. Then, the following properties hold:*

(7.1) $0 \leq o(q) \leq 1$.

(7.2) $0 \leq o'(q) \leq 1$.

Proof. Properties (7.1) and (7.2) are trivial since $\beta \in [0, 1]$. □

Appendix B.6. Properties of Equation (27) and Equation (28)

Properties (8.1—8.2)**.** *Let $q_i^a, q_i^b \in C$, $1 \leq i \leq p$ with $p \leq m$ and m is the total number of condition attributes. Then, the following properties hold:*

(8.1) $0 \leq o(q_1, \cdots, q_p) \leq 1$.

(8.2) $0 \leq o'(q_1, \cdots, q_p) \leq 1$.

Proof. Properties (8.1) and (8.2) are trivial since $\beta \in [0, 1]$. □

Appendix B.7. Properties of Equation (30) and Equation (31)

Properties and the corresponding proofs of Equation (30) and Equation (31) are similar to those of Equation (16) and Equation (17).

Appendix B.8. Properties of Equation (32) and Equation (33)

Properties and the corresponding proofs of Equation (32) and Equation (33) are similar to those of Equation (18) and Equation (19).

Appendix B.9. Properties of Equation (37) and Equation (38)

Properties and the corresponding proofs of Equation (30) and Equation (31) are similar to those of Equation (16) and Equation (17).

Appendix B.10. Properties of Equation (39) and Equation (40)

Properties and the corresponding proofs of Equation (32) and Equation (33) are similar to those of Equation (18) and Equation (19).

Appendix C. Characteristics of symbolic and nominal condition attributes

Table C.27: Values of condition attribute Gender

Gender Code	Number	%
f	11824	63.27
m	6864	36.73
Total	18688	100

Table C.28: Values of condition attribute EntreType

Type Code	Number	%
i	17699	94.71
g	989	5.29
Total	18688	100

Table C.29: Values of condition attribute ActivityType

Activity Code	Activity Type	Number in db	%
1	Animal Rearing	473	2.53
2	Artist	15	0.08
3	Arts and Crafts	178	0.95
4	Baker	89	0.48
5	Beauticians	121	0.65
6	Butcher	82	0.44
7	Carpentry	136	0.73
8	Clothes Shop	565	3.02
9	Computing Classes	17	0.09
10	Construction	227	1.21
11	Cosmetics	83	0.44
12	Education	33	0.18
13	Electrician	145	0.78
14	Farming	3054	16.34
15	Fishing	103	0.55
16	Fishmonger	159	0.85
17	Food Production	409	2.19
18	Food Shop	304	1.63
19	Furniture Making	22	0.12
20	General Store	1642	8.79
21	Green LoanGreen Loans	123	0.66
22	Greengrocer	346	1.85
23	HairdressersSalo	207	1.11
24	Hardware Shop	68	0.36
25	Health Shop	16	0.09
26	Home Improvements	135	0.72
27	Improving Sanitatio	179	0.96

Activity Code	Activity Type	Number in db	%
28	Internet Cafe	58	0.31
29	Itinerant Trader	415	2.22
30	Jeweller	84	0.45
31	Laundry	24	0.13
32	Market (Clothing)	264	1.41
33	Market Stall	918	4.91
34	Market Stall (Food)	1158	6.20
35	Mobile Phones	140	0.75
36	Pharmacy	59	0.32
37	Photographer	24	0.13
38	Plumbing	21	0.11
39	Production	860	4.60
40	Raising livestock	424	2.27
41	Raising poultry	354	1.89
42	Recycling	286	1.53
43	Repaying MoneyLende	44	0.24
44	Restaurant/Cafe	362	1.94
45	Sewing/Tailoring	1960	10.49
46	Shop	823	4.40
47	Solar Energy	5	0.03
48	Trader	177	0.95
49	Transport	794	4.25
50	Vehicle Repairs	169	0.90
51	Vehicle Spares	106	0.57
52	Welder	82	0.44
53	Workshop	146	0.78
-	Total	18688	100

Table C.30: Values of condition attribute Country

Country Code	Country Name	Number	%
1	Benin	1091	5.84
2	Bosni	228	1.22
3	Cambo	2310	12.36
4	Ecuad	1951	10.44
5	Indon	16	0.09
6	Malaw	851	4.55
7	Pakis	6801	36.39
8	Phili	1845	9.87
9	Rwand	58	0.31
10	Togo	931	4.98
11	Vietn	1501	8.03
12	Zambi	1010	5.40
13	Zimba	95	0.51
-	Total	18688	100