



# Multi-criteria decision-making leveraged by text analytics and interviews with strategists

Jorge Iván Pérez Rave<sup>1,2</sup> · Gloria Patricia Jaramillo Álvarez<sup>1</sup> · Juan Carlos Correa Morales<sup>1</sup>

Revised: 12 February 2021 / Accepted: 13 July 2021 / Published online: 28 July 2021  
© The Author(s), under exclusive licence to Springer Nature Limited 2021

## Abstract

Strategic decision-making in organisations is a complex process affected by preferences, experiences, perspectives, and knowledge, which, in most cases, are ambiguous, contradictory, and represented in unstructured data. This paper develops a methodological framework to address strategic decision-making processes from a multi-criteria perspective, assisted by text analytics and interviews. The framework comprises five stages and 12 steps, and is empirically tested in a decision scenario involving a strategic focus for future analytics initiatives in order to stimulate value generation from analytics. The proposed framework enables the discovery, validation, and prioritisation of strategic patterns from relevant interview data. Among six decision alternatives discovered in the validation scenario, customer analytics was the strategic focus most relevant to future analytics initiatives. This article contributes to understanding and addressing complex decision-making processes and mixed research in organisations, through a multi-criteria perspective leveraged by a text-driven computational approach.

**Keywords** Multi-criteria analysis · Analytics · Strategic decisions · Text analytics · Interviews · Customer analytics

## Introduction

The question of how to make decisions in strategic organisational scenarios is a complex process that has seen increased interest in the field of decision-making sciences since the late 1960s (Kellermanns and Floyd 2005; Stagner 1969). This process is characterised by the need to identify and prioritise strategic alternatives (Floyd and Wooldridge 1992; Kellermanns and Floyd 2005), taking into account the perspective of organisation strategists. Strategists are understood as those who are responsible for organisational performance by providing a dynamic environment which forces the organisation to innovate and mitigate competition for survival (Heene 1997; Collins and Clark 2003; Holm et al. 2020). Strategists are thus either part of the top management

team or functional managers, or assume both roles, depending on the size/structure of a firm.

Grošelj and Stirn (2017) explain that strategic decision making involves a diversity of opinions among decision-makers. Each strategist has substantial knowledge and a focus on their specific organisational context (e.g., area, process) but also a particular understanding of the whole organisation (e.g., personal bias of the strategist; Stagner 1969; value conflicts and mental models not shared among and across leaders, Kruse 2008). Each strategist thus has their own idea of the “best” route via which to contribute to organisational performance and competitive advantage, which is consistent with the recognised causal ambiguity between strategic resources and performance (Barney 1991; Kellermanns and Floyd 2005).

To illustrate a strategic decision-making issue in the current era, consider, for example, an organisation in which a set of strategists are debating about which would be the best management style to adopt in all organisational projects to potentiate project success, recognising and taking advantage of the current organisational capabilities/resources. This is a strategic issue considering its implications for the current and future projects’ management/performance/value and organisational performance (Dutton et al. 1983; Stacey 2009). Some essential questions in this regard include: how

✉ Jorge Iván Pérez Rave  
investigacion@idinnov.com

Gloria Patricia Jaramillo Álvarez  
gpjarami@unal.edu.co

Juan Carlos Correa Morales  
jccorrea@unal.edu.co

<sup>1</sup> Universidad Nacional de Colombia, Medellín, Colombia

<sup>2</sup> IDINNOV S.A.S, Medellín, Colombia



should such a goal be operationalised? How should decision alternatives be defined? From which criteria should alternatives be evaluated? How should the relevance of the criteria and alternatives be measured? To answer these questions, among others, implies considering the named strategists. However, answers for each strategic issue will be influenced by subjective and heterogeneous aspects (e.g., mental models, interests, experience, decision power, area leading in the organisation, and available information), making consensus difficult and affecting quality decision making (Burnett 1998; Dutton et al 1983; Stacey 2009). Besides, there is not a universal answer to these questions, because each organisation has a particular latent culture, capabilities, knowledge, resources, and style in general. Likewise, there is no comprehensive objective structured data that allows us to simply analyse it and obtain a magic response to these questions. Therefore, decision-making approaches recognise that at the early stage, a strategic issue as the exposed example should be adequately formulated and specified (Ansoff 1980; Dutton et al 1983; Stacey 2009).

The decision-making process is even more complex in the current era (and future projections) (Grošelj, and Stirn 2017; Zhang et al. 2019) due to the extensive use of virtual work environments (Danilova et al. 2020), which have been accelerated by issues such as the Covid-19 pandemic. For example, decision making during online sessions is highly dependent on the operating conditions of the technology, and may reduce the potential for (a) interactions between decision-makers (Deshpande 2016; Tiell 2017), (b) sensory experiences (Snyder 2021), and (c) feedback during the sessions (Pham and Hoai 2021).

In summary, strategic decision-making in the organisational context is a complex process with consequences for organisational performance, and is affected by preferences, experiences, perspectives, mental models, and knowledge, which, in most cases, are ambiguous (Musaji et al. 2020), contradictory (Dooley and Fryxell 1999), and represented in unstructured data (Markham et al. 2015).

One recent concern requiring consensus among organisational strategists is the reactive perspective that is frequently used to define/deploy analytics initiatives. Although there are several successful cases using analytics methods and their combinations, studies report a disconnect between such initiatives and organisational strategic priorities (Becker 2017; Park et al. 2017). According to Nacarelli and Gefen (2021), Curtis and Perryman (2021), and Wang (2019), between 60 and 85% of analytics initiatives fail to achieve their objectives and generate value for the organisation.

We aim to contribute to consensual strategic decision-making by developing a multi-criteria framework assisted by text analytics and interviews. This framework allows the discovery, evaluation, and choice of consensual strategic patterns from a computation/evidence-based perspective, and is

applied in a strategic domain of analytics. This domain refers to the priority strategic focus required by an organisation to ensure value generation from analytics. This is important because it is consistent with international concerns about the decision-making process among strategists and recent failures in analytics initiatives (Nacarelli and Gefen 2021; Wang 2019; Curtis and Perryman 2021). Each organisational area (e.g., human resources, marketing, production, logistic) frequently uses isolated analytics initiatives supported from the individual perspectives (e.g., interests, mental models) of each strategist instead of a common strategic focus. This isolated/individual view is causing some projects to result in conflict, repetitive tasks, efforts that are focused merely on methodological goals, and superficial/fragmented/partial/reactive analytics “solutions.” These projects are based on sub-optimal purposes, abandoned before completion, high resource consumption, and do not favour organizational performance. In other words, analytics initiatives are frequently disconnected from strategic priorities.

A consensual strategic focus for future analytics initiatives in harmony with organisational priorities would help with: (a) the better use of limited/core resources, (b) the granularity of the solutions, (c) teamwork, and (d) the development/accumulation of more relevant organisational capabilities in analytics to favour organisational performance and competitive advantage.

We take a computation/evidence-based perspective in order to develop a multi-criteria decision analysis framework assisted by text analytics and interviews, to contribute to consensual strategic decision-making processes.

Multi-criteria decision analysis allows prioritising decision alternatives in the organisational management arena. This challenge is dependent on the multiple preferences, perspectives, and mental models of the decision makers, and has been widely addressed as a multi-criteria decision problem (Marttunen et al. 2017).

Interviews are useful because they are a traditional way to capture valuable data from strategists and promote discussions, analyses and decision-making processes. Interviews are “a way of finding out what others feel and think about their worlds” (Rubin and Rubin 1995, p. 1). Interviews are therefore a valuable technique to stimulate participation and deepen the thinking/feelings of people about various organisational concerns/challenges/needs (Kee and Thompson-Hayes 2012), which is also useful in the current and future virtual work environments (Danilova et al. 2020).

In addition, if through the conscious verbal language of individual (e.g., strategist), from the late 1960s to date the difficulty of making consensual strategic decisions in practice has been reported, the discovery of non-explicit patterns in that language can be a useful mechanism to help generate such consensus based on the underlying (latent) evidence. Text analytics is a powerful resource for this



purpose, because according to Moulin (1992), a writer/speaker chooses relevant information from the world that surrounds them, then builds a conceptual map (concepts and relationships) and expresses it through an oral or written discourse (linguistic level) to describe beliefs, emotions, sensations, knowledge, behaviours, attitudes, and so on, without them being physically present (Kazemzadeh et al. 2016). It is thus possible to identify verbal language about the relevant strategic concerns in an organisation through interviews, to transform these verbal manifestations into textual data (manual or automatic transcriptions), and then to mine it through text analytics. Text analytics is thus useful in the present context because it allows valuable latent patterns to be discovered in textual data (including interview transcriptions derived from virtual or face-to-face environments) concerning processes, people, technology, networks, innovation, and learning, among other organisational facets (Khan and Vorley 2017).

This paper extends frameworks for unstructured data-enabled decision making, such as those by Markham et al. (2015) (in new product development: define questions, identify sources, create dictionaries and rules, collect data, conduct analysis, assess data, scores, and decide); Amado et al. (2018) (in marketing: literature collection, dictionaries, document-term matrix, and topic modelling); and Mathaisel and Comm (2021) (in political marketing: text preprocessing, topic modelling, sentiment analysis, and visualisation), providing the following:

- (1) A mixed methodological framework (five stages and 12 steps) to assist computation/evidence-based managerial decision-making processes, based on valuable written natural language collected from interviews (e.g., to strategists).
- (2) Procedures to examine four properties (reliability, validity, novelty, and robustness) in the latent patterns discovered from texts, using a data-driven approach nourished with psychometrical considerations.
- (3) Original evidence derived from applying the framework to the strategic focus of analytics initiatives. Addressing such a case led to customer analytics being identified and chosen as a priority topic for the organisation studied. We also provide suggestions about possible uses of the proposed framework in future marketing scenarios focused on customer analytics.

Currently, mixed research is required in several organizational fields (Harrison 2013). Krishen and Petrescu (2021, p. 1) state “Perhaps academic research methodologies needed to evolve over the last few decades to augment and interrogate logical positivism and interpretivism.” This paper expands the understanding of mixed research

into strategic decision-making, taking a multi-criteria perspective assisted by text analytics and interviews.

## Literature review

This section explains two essential elements for a better understanding of the paper: the argument that analytics is a strategic issue, and its role in the organisational context.

### The strategic issue

A strategic issue in the organisational arena can be understood from several perspectives: prospective, current opportunity/problem, geographic, and procedural. The prospective perspective assumes a strategic issue as a forthcoming development (inside or outside the organisation) that is likely to have a relevant effect on the organisation’s ability to achieve its objectives (Ansoff 1980). From a current opportunity/problem perspective, it is a substantial/structural gap between the performance objective of the organisation and the achieved performance (Baaij and Reinmoeller 2018). The geographic viewpoint represents this as a kind of problem of interest to many organisations in various countries (Sandul 2015). From a procedural view, a strategic issue involves a complex task that does not have a well-defined route to address it (Charest 2007).

Value generation from analytics (VGA) is an organisational issue which can be defined from these perspectives: (1) it is an opportunity/problem essential for meeting a firm’s current and future strategic objectives; (2) it is a topic of increased global interest for academia and organisations; and (3) it is a multi-criteria decision problem that is categorised as a complex decision problem.

VGA is an opportunity for considering the possibilities to take advantage of the volume, variety, velocity, and value (Sagiroglu and Sinanc 2013) of big data, and several applications in this area have contributed to improving organisational performance. It is worth noting the achievements of organisations such as T-Mobile (reduction of customer abandonment by 50%; van Rijmenam 2013) and McLaren Formula One (failure prevention; Hungerford 2017). Park et al. (2017, p. 76) state that “80% CEOs or executive teams view big data analytics initiatives have the potential to drive business value such as creating new revenue streams, improving operational efficiency or cutting cost.” This is consistent with Larson (2019, p. 95): “business analytics is a core competency critical to organizations to stay competitive.” VGA also represents a practical gap; for example, most big data analytics projects fail to achieve their mission (Park et al. 2017; Kabanda 2019). In fact, according to Becker (2017, p. 2320), “Big Data projects have still been experiencing very high failure rates.”



The diversity of references and cases cited for the first consideration also partially covers the second (global interest in the subject). This interest is also supported by a bibliometric exploration of Scopus (April 24, 2020), performing analytics in the context of big data, excluding 2020: *TITLE (“analytics”) AND TITLE-ABS-KEY (“big data”) AND (EXCLUDE (PUBYEAR, 2020))*. The search provided 4,643 documents from 2010 to 2019, and showed a notable increase by year (2010: 1 document; 2015: 523; and 2019: 1074). Several countries are interested in this subject, such as the United States (1426 documents), India (762 documents), China (434 documents), and the United Kingdom (309 documents). In summary, it is possible to state that VGA is a topic of global interest for helping to generate value for organisations.

With respect to the third consideration (multi-criteria decisions), there is no universal idea of how to generate value from analytics in organisations. The complexity of this concern is not merely methodological, but also ontological. For example, how can researchers assume and interpret analytics in the organisational context? Most efforts to respond to this question are merely procedural, perhaps because of the popularity of dealing with large volumes of data and the status that advertising portrays. However, one of the main causes of failure in big data initiatives are “wrong/inadequate skills and incorrect business objectives” (Becker 2017, p. 2321). Larson (2019, p. 98), based on several works, states that “the success and value of a business analytics project will be defined at the start of the project.” Similarly, Strohmeier and Piazza (2013), in a review of 100 papers on data mining (a resource for analytics) in the field of human resource management, conclude that solutions rarely consider the needs and requirements of the specific domains (and subdomains). It should therefore come as no surprise that most analytics initiatives fail to deliver valuable results for the organisation (Becker 2017; Campion et al. 2016;

Shah et al. 2019). Campion et al. (2016, p. 958), in a personnel selection context, state, “little is known regarding how organizations can leverage these advancements [big data] to develop more effective personnel selection procedures, especially when the data are unstructured (text-based).” Larson (2019, p. 98) points out that “analytics is still evolving; thus, best practices are just starting to emerge.” In other words, the problem (or opportunity) stemming from analytics is a confusing issue; that is, it is not precise or structured, and there is no clear goal relating to the classical proposition “to generate value for organizations from big data.”

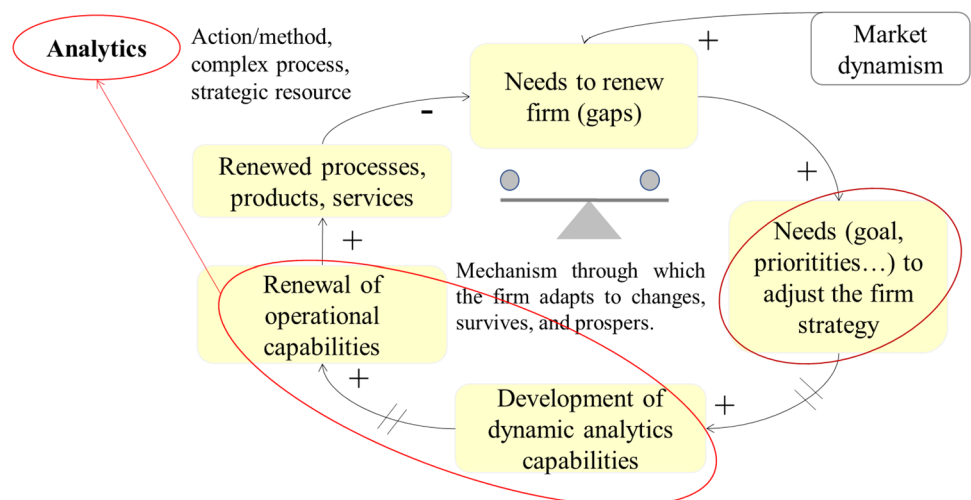
The described considerations (current practical opportunity/problem, prospective solution, globalised topic, and procedure) thus reinforce the assumption that VGA is a strategic issue for organisations.

### Role of analytics in the organisational context

If VGA is a strategic issue, it is reasonable to understand the role of analytics from a systems thinking viewpoint (Senge 1996) that allows feedback relationships, temporal delays, and the balance achieved through adaptation mechanisms to renew firms and generate competitive advantages to be considered. Figure 1 shows a proposed archetype in this regard.

Considering Fig. 1, it is worth noting that market changes and pressures require firms to generate and deploy strategic initiatives to respond successfully to the needs of their environment. A fundamental mechanism for the adaptation and responses of firms is the development of dynamic capabilities. Teece et al. (1997, p. 516) describe dynamic capabilities as “the firm’s ability to integrate, build, and reconfigure internal and external competences to address rapidly changing environments.” Collis (1994) asserts that dynamic capabilities govern changes in operational capabilities (day-to-day activities), which are responsible for the processes, products, and services of firms (Winter 2003). According to

**Fig. 1** Archetype for the role of analytics in organisations



Pavlou and Sawy (2011), among others, dynamic capabilities are an essential driver of a firm's competitive advantage through the renewal of operational capabilities.

From the dynamic capability approach, analytics can be understood as actions/methods, processes, and strategic resources, and their final mission is to contribute to the generation of a firm's competitive advantage through the renewal of their day-to-day operations, through the entire data life cycle. This is consistent with the provider-user relationship recognised between analytics and functional organisation areas (Hagen 2021); analytics delivering insights from the data provided by the internal/external operations of functional areas; likewise, these areas use such insights for better decision-making processes in the organisational arena.

This renewing role of analytics is also consistent with the methodological, process, and strategic perspectives. From the methodological perspective, analytics is "a way of extracting value from these huge volumes of information, and it drives new market opportunities and maximizes customer retention" (Zakir et al. 2015, p. 81) or "the extraction of hidden insight about consumer behavior from Big Data applying that insight in business decisions" (Ertemel 2015, p. 45). From a process perspective, analytics is "the process of inspecting, cleaning, transforming, and modeling data with the goal of highlighting useful information, suggesting conclusions, and supporting decision making" (DeKroon and Karp 2013, p. 4) or "the process of developing actionable insights through problem definition and the application of statistical models and analysis against existing and/or simulated future data" (Cooper 2012, p. 3). From a strategic perspective, Fagerjord and Kueng (2019, p. 7) considered the successful case of Netflix, and saw analytics as "a strategic capability that creates competitive advantage and allows it to achieve superior returns". Similarly, Grinko et al. (2017, p. 5) refer to analytics "as a strategic differentiator for management."

It is therefore vital for an organisation to identify and prioritise a strategic focus for future analytics initiatives (e.g., ideas, projects, investments, approaches, technologies, and applications) that guides and stimulates a proper renovation of operational organisational capabilities, using correct and relevant processes, actions, and methods of analytics.

Without correctly identifying and understanding a strategic focus of analytics that is consistent with strategic organisational priorities, other forms of analytics (action/method and process) will tend to fail. However, in the real organisational arena, it is not easy to generate consensus among strategists about the strategic focus of analytics; this issue has no universal answer, it is multi-criteria decision-making depending on the internal and external factors surrounding the strategists. It is also not a precise problem; for example, it is not easy to see a strategic concern and to immediately understand it. How does each strategist interpret it? What

is the most complete and parsimonious representation of it? What are the underlying preferences, alternatives, and criteria of the strategists?

The real organisational arena in the strategic scenario therefore requires soft and hard skills, and a combination of analytics frameworks with other complementary resources. We take this opportunity to propose a multi-criteria decision framework, assisted by text analytics and interviews with a set of strategists.

The importance of using text analytics and interviews is supported in that one of the main weaknesses of multi-criteria decisions is subjectivity in establishing objectives, alternatives, and criteria (and their weights); that is, the stage of problem structuration (Marttunen et al. 2017).

The observation scenario provides the following:

- (a) Relevant and non-common textual data: The strategists are very busy because considering the organisational structure, they are part of the top management team. Obtaining primary data from interviews lasting for 45–60 min with a set of strategists is thus a practical challenge in the natural environment. The data obtained was fundamentally related to long-term situations (e.g., the current situation of the firm, the organisation vision in five years, the main practical challenges of achieving this vision in the future, etc.). That is, the underlying thinking in this data is not something that changes from one day to the next; unlike a traditional corpus of tweets, the hidden patterns in the data are not very changeable in the short term.
- (b) It is hard to identify consensual patterns among organisational strategists, because each one tends to observe, interpret, and act based mainly on a fragment of the reality generally represented by the area or process that the strategists manage (e.g., finance, human resources, quality, etc.).
- (c) The field work was performed during the Covid-19 pandemic, forcing data collection to be carried out when the organisation had implemented working at home (or teleworking). Online sessions had already been implemented to deploy strategic, tactical, and operational decision-making processes in the organisation. This adds value to the scenario studied, because teleworking and virtual environments are world megatrends (Hajkowicz and Roy 2013; Naotunna and Priyankara 2020).

These considerations reinforce the relevance of the observation scenario and the importance of adopting a multi-criteria perspective for the problem of VGA from a strategist's perspective. Unlike most multi-criteria decision-making works which are assisted by interpretative tools (e.g., SWOT, brainstorming, and the Delphi technique) for structuring the problem (e.g., alternatives,



criteria, preferences) and evaluating alternatives within criteria; however, our framework uses a text-driven computational approach, employing data from interviews with strategists.

From a practical perspective, the empirical application of the proposed framework in the observation scenario provides insights about consensual strategic patterns related to analytics, considering the explicit/implicit thinking of functional managers (in this context, they are also members of the management team). These insights provide valuable input for the CEO to consider during the institutionalisation of the strategic focus for future analytics initiatives in the organisation.

## Methodology

Figure 2 summarises the proposed methodological framework to systematically assist multi-criteria decision analysis with text analytics and interviews. The framework (Fig. 2) is based on a mixed paradigm (qualitative—quantitative) from a text-driven computational approach that leverages interview data for organisational strategists.

Note in Fig. 2 that the proposed framework assumes a general multi-criteria approach nourished by discovery patterns/resources/capabilities derived from text analytics (text pre-processing, topic modelling, distributed

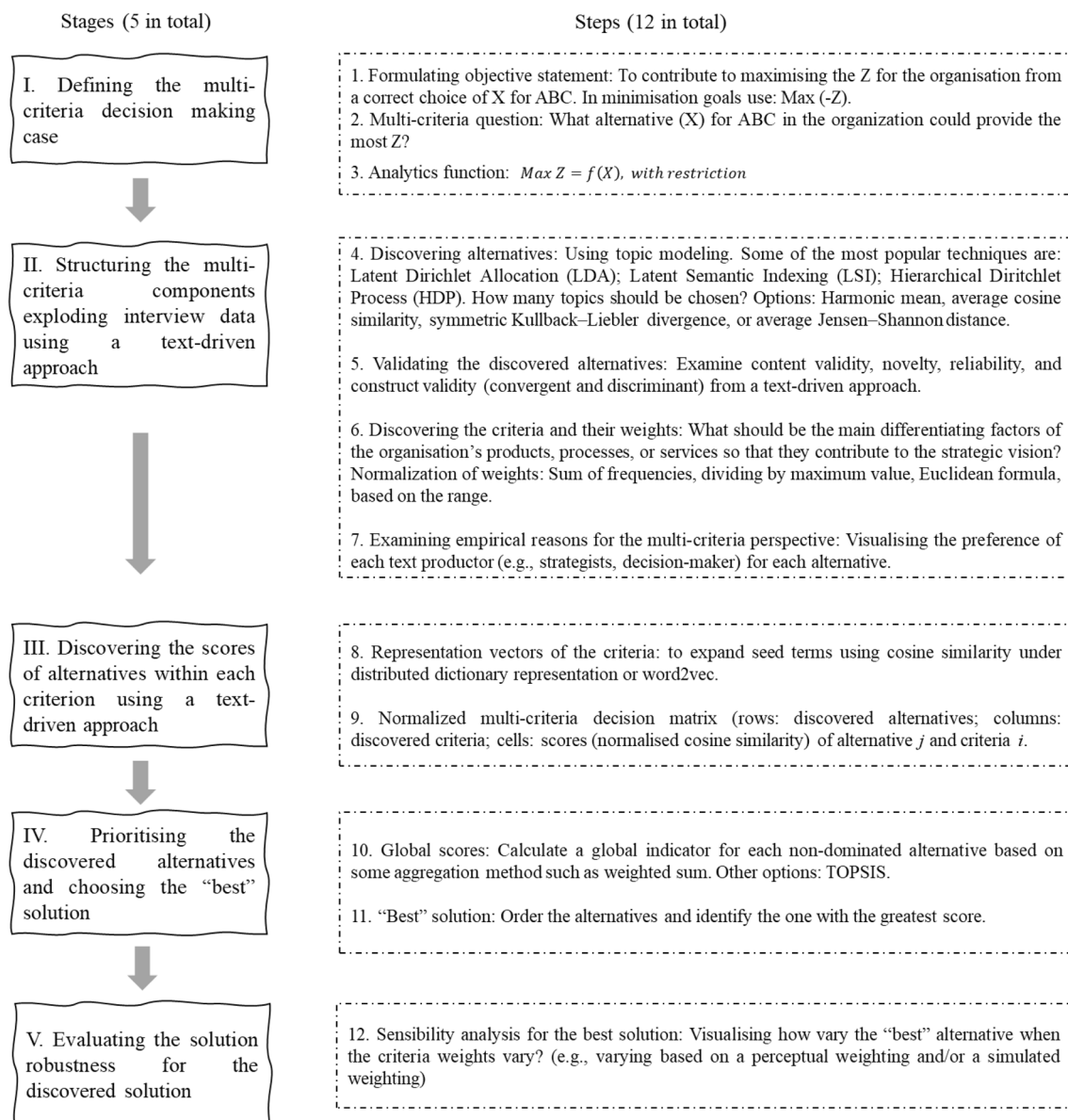


Fig. 2 Framework for multi-criteria decision analysis assisted by text analytics and interviews



dictionary representation, cosine similarity) which increase the value of interview data from text producers (in this case, the strategists). Such resources/capabilities also enable more automated operations, which increase efficiency, transparency, and the reproducibility of the decision-making process.

In the next subsections, we show and explain the deployment of the framework (five stages and 12 steps) and its empirical results, considering the studied decision-making scenario on VGA.

## Results and discussion

This section is structured in the stages of the proposed framework (see Fig. 2). In this study, the decision-makers were all strategists who direct functional areas and at the same time form part of the top management team. These strategists play an essential role in the context of data analytics (providers of data and customers of data products) considering the day-to-day operations in the organisation (Hagen 2021).

### Defining the multi-criteria decision making case

- *Objective statement*: To contribute to maximising the VGA for the organisation using the correct choice of strategic focus for future analytics initiatives.
- *Multi-criteria question*: What alternative (X) for the future development of analytics initiatives in the organisation could provide the most VGA?
- *Analytics function*: see Eq. 1:

$$\text{Max VGA} = \sum_j \sum_i W_i S_{ij} X_j \quad \text{with} \quad \sum_j X_j = 1 \quad (1)$$

Let  $X_j$  be a binary variable that represents the event of chosen (1) or not (0) for topic  $j$  (alternatives) and  $W_i$  is the weight (relative importance) of criterion  $i$ . Let  $S_{ij}$  be the scores for the alternative  $j$  within the criterion  $i$ . Traditionally, in multi-criteria decision scenarios,  $S_{ij}$  is calculated based on perceptual rating (Kreković et al. 2016; Marttunen et al. 2017; Munier et al. 2019; Pérez-Rave et al. 2015). However, we use a text-driven approach based on cosine similarity (Eq. 2) between terms' vectors ( $tv$ ) representing alternative  $j$  and criteria  $i$ . This procedure is explained in the section “Discovering the scores of alternatives within each criterion using a text-driven approach.

$$\text{CoSim}(\theta)_{ij} = \frac{tv_i \cdot tv_j}{\|tv_i\| \|tv_j\|} \quad (2)$$

With  $\|tv\|$  the vector length. In addition, in the observation scenario, for reasons of focus, specialisation, and better use

of resources, the company restricted the number of alternatives to choose one strategic topic to focus on and develop future analytics initiatives.

### Structuring the multi-criteria components exploding interview data using a text-driven approach

This structuration implies defining the alternatives, decision criteria, and criteria weights; however, instead of structuring the multi-criteria case from the traditional interpretivist perspective (e.g., FODA, Delphy method, brainstorm) (Marttunen et al. 2017), we incorporate text analytics resources that allow a reduction of subjectivity and increase of transparency, efficiency, and reproducibility.

### Discovering the alternatives

From a qualitative perspective, we administered eight questions (online sessions) to put to a set of organisation strategists (eight managers and directors of a Colombian organisation). Two of these questions were “Considering the strategic role of the organisation, how do you see the company in 5 years?” and “What are the main concerns or needs that the organisation should begin to address to come as close as possible to that privileged position?” This stage produced text data derived from interviews with strategists, which constitute the corpus under study. From a text analytics perspective, this stage involved the following steps: (1) consolidating and reading the interview transcripts (corpus); (2) preparing the data using normalisation (lower case, remove punctuation, corrections), and stopwords removal; and (3) making a basic summary of the data considering the number of pages, words, and characters to answer the following question: What numerical synthesis reasonably describes the basic content of the strategists' thinking?

The interview transcriptions range from 3 to 5 pages, with a mean of 4.75 and a standard deviation of 1.49. Similarly, the number of document characters demonstrates the long duration of the interviews. For example, while a tweet has no more than 278 characters (mean of 34 characters) and text analytics applications largely use that type of data, the present corpus of interview transcriptions ranges from 3691 to 10,416 characters, with a mean of 6375.5 characters and a standard deviation of 2611.98 characters. In other words, this mean value is 187.5 times greater than the common size of a tweet. Gong and Poellabauer (2017) called their sample (in the context of detecting depression) a “long interview” sample, with a duration between 7 and 33 min. In the present study, however, the duration of interviews ranged from 45 to 60 min. The 5919 words that we obtained from interviews with strategists thus constitute a valuable corpus



with complex and challenging patterns for discovery and conversion into actionable information.

We defined a coherent page as the unit of analysis in order to explore the common underlying context for the strategists during the interviews and not merely the particular context of each strategist (document), in the present study. This segmentation is a practical approach that allows the application of topic modelling to documents or parts of documents. The integrative framework proposes diverse existing techniques to deploy topic modelling to discover latent patterns representing the decision alternatives. We chose Latent Dirichlet Allocation (LDA, Hou-Liu 2018; under Gibbs' method) to show the first application of topic modelling in the framework, to discover latent topics from text data (interviews), because it is efficient and widely recognised in the literature, and makes no previous assumptions about the content of documents. For example, Mathaisel and Comm (2021, p. 3) use LDA in political marketing and state that "it is robust, common to open-source software..., and is the most widely used technique." LDA requires the specification of two parameters; one is the corpus, and the other is the number of topics. The latter is still an open question (Hou-Liu 2018); however, the proposed framework suggests considering at least four metrics (incorporated into the *ldatuning* library, in R; Nikita 2020) to establish this parameter: harmonic mean, average cosine similarity, symmetric Kullback–Liebler divergence, and average Jensen–Shannon distance. Figure 3

suggests two decisions for the number of topics; one is near to 13 and the other is 6 or 7.

We obtained the five most likely terms for each suggestion (6, 7, and 13 topics), and, considering interpretability, parsimoniousness, and the number of strategists (8), there were six chosen topics. Figure 4 summarises the content/label of each topic using the five most probable terms.

Note, in Fig. 4, the reasonable coherence of the composition of the latent topics that the study discovered. From a practical–strategic perspective of data analytics, therefore, a proposed interpretation of the topics (see Fig. 2), from an inductive process, is the following: *F1. Operations analytics* (customer, technological, productive, process, believe); *F2. Product analytics* (knowledge, product(b), important, see, model); *F3. Market analytics* (market, leader, privileged, world, need); *F4. Strategic analytics* (area, strategic, product, manage, service); *F5. Human resource analytics* (human, see, person, go, management); and *F6. Customer analytics* (customer, new, form, see, enterprise). That is, from the authors' viewpoint, which is to realise the inductive process (unsupervised learning), it is obvious that the most-used terms represent the discovered patterns (topics) reasonably; however, is this also perceivable from the perspective of other people related to organisational management?

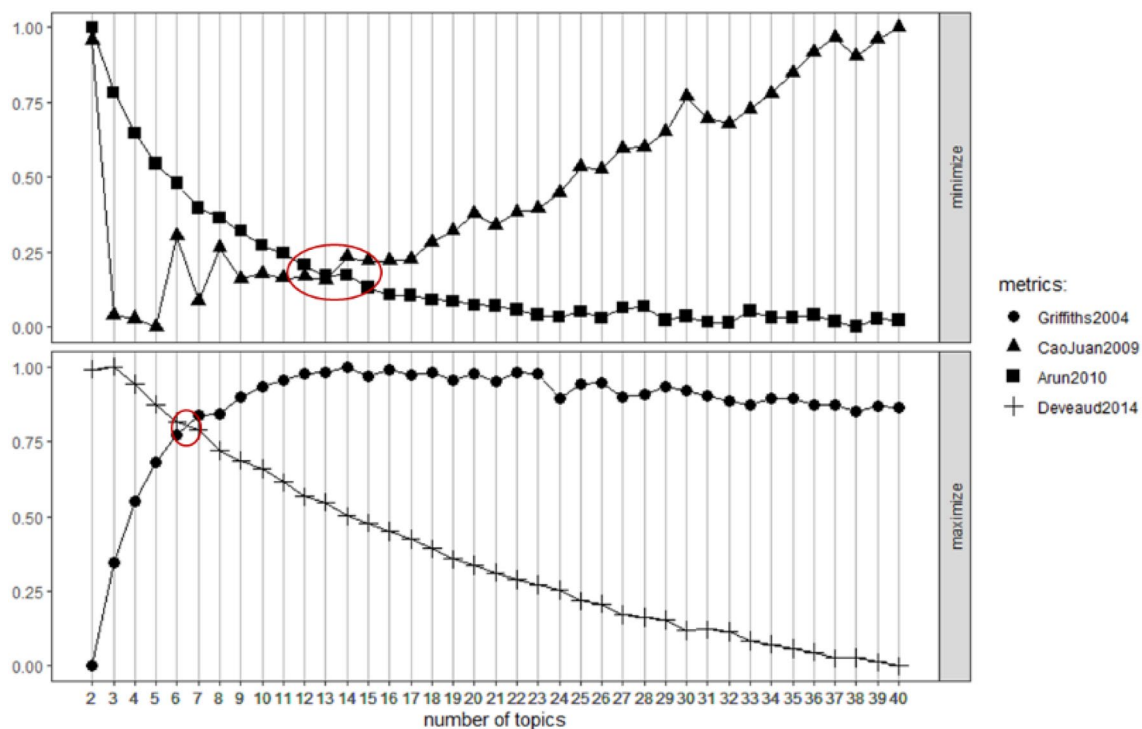


Fig. 3 Results of exploring the number of predefined topics with referenced metrics







Fig. 4 Top five terms in each latent topic. Note the exact product name was replaced by “product(b)”

### Validating the discovered alternatives

This stage considers four central properties: the content validity, novelty, reliability, and construct validity (convergent and discriminant) of discovered patterns (alternatives) aiming to incorporate psychometrical considerations.

**Is it reasonable to assume content validity for the discovered alternatives?** A fundamental property of useful topics in computation fields is coherence, which from a qualitative perspective implies that humans can understand and interpret the topic (Chang et al. 2009). In management fields, this property can be similar to content validity. According to Liman et al. (2020), and Sekaran and Bougie (2016), content validity involves consulting expert opinions about the items, wordings, and phrases contained in the measure. In the present context of organisational management, we thus consider that the meaning and interpretation of a set of top terms for each topic is the first exploratory step to infer its content validity. However, to explore such a property from other perspectives, we administered a survey to a sample of 34 professionals in areas related to organisational management. Twenty four of them responded to the instruction in the period May 10–13, 2020. The instruction was as follows: *Please assign each word list (row) to the concept (column) with which that list has the most affinity.* The sample involved several areas related to organisational management. With respect to undergraduate education, this sample included: industrial engineering (11/24), business management (5/24), production engineering (2/24), systems engineering (2/24), and others (4/24 each) from economy, psychology, accountancy, and social communication. It is worth noting that 19/24 respondents had finished postgraduate education —masters (10) and specialisation (7)—and two people are currently studying —one masters and one specialisation. Postgraduate education was in the areas of business (8), finance (3), and analytics, education, management systems, systems, logistics, economy, engineering, and mental health (1 each).

Table 1 provides the results of the proxy strategy for exploring the content validity of the patterns (topics), considering the participants’ perceptions of the affinity between the terms and the proposed concepts.

Table 1 shows that, in all cases, the participants assigned each set of terms more frequently to the proposed label (concept): F1—operations analytics (54.17%), F2—product analytics (66.67%), F3—market analytics (62.5%), F4—strategic analytics (50%); F5—human resource analytics (75%), and F6—customer analytics (41.67%). In five or six cases (excepting F6), the frequencies were greater than 50%. In fact, when exploring validity from a statistical perspective, researchers usually expect that different measures of the same construct share a minimum of 50% of the total information.

**Is it reasonable to assume novelty for the discovered alternatives?** This property is important because the definition of analytics specifically includes “non-obvious patterns”. The firm under observation recognised the novelty and value of the discovered patterns, however, although this recognition in the study domain is necessary, it is not enough to reach conclusions about the novelty of the discovered patterns. In order to provide more generalist evidence, we therefore performed a search on Scopus, using proposed labels (or their synonyms) for the discovered topics (e.g., “customer analytics”) in the title and restringing for articles or conference papers (generally more empirical applications; hence, excluding books, book chapters, reviews, etc.). We considered the total numbers of documents reported, the percentage of this value with respect to the total documents recovered for the global theme (e.g., “data analytics”), the year of the first document that Scopus provided, and the ratio (2019) with respect to the number of documents in the first year named.

Table 2 presents the results of the search strategy exploring the novelty of topics. It provides exploratory evidence for five aspects of interest; for example, the search (r2) for the concept “operations analytics” provided 84 documents,



**Table 1** Results of a proxy strategy for exploring the content validity of the patterns

Terms of discovered patterns (topics)	Labels derived from the interpretative–inductive process					
	Strategic analytics	Human resource analytics	Market analytics	Operations analytics	Product analytics	Customer analytics
F4. Area, strategic, product, management, service	<b>50%</b> <b>(12)</b>	0% (0)	4.17% (1)	16.67% (4)	16.67% (4)	12.5% (3)
F5. Human, see, person, go, management	8.33% (2)	<b>75%</b> <b>(18)</b>	4.17% (1)	4.17% (1)	0% (0)	8.33% (2)
F3. Market, leader, privileged, world, need	12.5% (3)	4.17% (1)	<b>62.5%</b> <b>(15)</b>	0% (0)	8.33% (2)	12.5% (3)
F1. Customer, technological, productive, process, believe	8.33% (2)	4.17% (1)	0% (0)	<b>54.17%</b> <b>(13)</b>	4.17% (1)	29.17% (7)
F2. Knowledge, product, important, see, model	4.17% (1)	0% (0)	25% (6)	4.17% (1)	<b>66.67%</b> <b>(16)</b>	0% (0)
F6. Customer, new, form, see, enterprise	25% (6)	4.17% (1)	8.33% (2)	12.5% (3)	8.33% (2)	<b>41.67%</b> <b>(10)</b>

Percentages are expressed by file ( $\sum \text{rows} = 100\%$ ); in parenthesis: absolute frequencies (number of respondents). In diagonal (bold): frequencies of the terms preassigned (proposed composition) to the discovered latent topics

**Table 2** Exploring the novelty of the patterns outside the domain (May 11, 2020)

No	Search expression	(A) Total docs	(A)/(r1) (in %)	(B) Year of the first doc. <sup>a</sup>
(r1)	TITLE (“data analytics” OR “big data analytics” OR “business analytics”) AND <sup>b</sup>	4381	NA	2002
(r2)	TITLE (“operation analytics” OR “analytics of operations” OR “process analytics” OR “analytics of processes”) AND <sup>b</sup>	84	1.92	2005
(r3)	TITLE (“product analytics” OR “analytics of products” OR “service analytics” OR “analytics of service”) AND <sup>b</sup>	15	0.34	2004
(r4)	TITLE (“market analytics” OR “analytics of market”) AND <sup>b</sup>	6	0.37	2010
(r5)	TITLE (“strategic analytics” OR “strategy analytics”) AND <sup>b</sup>	7	0.16	2008
(r6)	TITLE (“human resource analytics” OR “analytics of human resource” OR “people analytics” OR “personnel analytics” OR “employee analytics”) AND <sup>b</sup>	22	0.50	2012
(r7)	TITLE (“customer analytics” OR “analytics of customer”) AND <sup>b</sup>	18	0.41	2006

<sup>a</sup>The first document that the database reported; it does not refer to the first document in the world

<sup>b</sup>(LIMIT-TO(DOCTYPE, “ar”) OR [LIMIT-TO(DOCTYPE, “cp”)])

representing 1.92% of the total documents that the database reported for (r1) analytics in general.

The results in Table 2 reinforce the novelty of the proposed underlying patterns. Note that in all cases, the participation of each type of analytics in the general field is less than 2% (second column, Table 2), with five of six types of analytics accounting for less than 1%. Similarly, the first documents that Scopus reported for the search expression area are relatively recent (2004–2012), and the low degree of maturity in the study of the underlying types of analytics to date is remarkable. These types of analytics, underlying the list of identified terms, are therefore important research fields for future studies and for the practice of organisational management.

**Is it reasonable to assume reliability for the discovered alternatives?** Reliability refers to consistency among successive measures of the same construct to assure reproducibility. By analogy with these properties and considering the present scenario of text analytics in organisational management, the average cosine similarity between term vectors may therefore be a useful measure to infer internal consistency, that is, a form of reliability. In fact, in a study scenario using a Likert scale, which management research frequently uses, Chakrabarty (2018) proposes, among other things, cosine similarity (using only the frequencies of the item) as a non-parametric measure of reliability, and compares this metric with the traditional Cronbach’s alpha. Similarly, whether the consistency (using the average cosine similarity) among



terms of a discovered pattern (topic) is not only reasonable but also greater than the consistency among such terms and the terms of the other patterns could favour convergent and discriminant validity. However, two problems in the present text analytics scenario are defining “a reasonable value” for the average internal consistency and determining the number of terms to use to represent the patterns (concepts) under study.

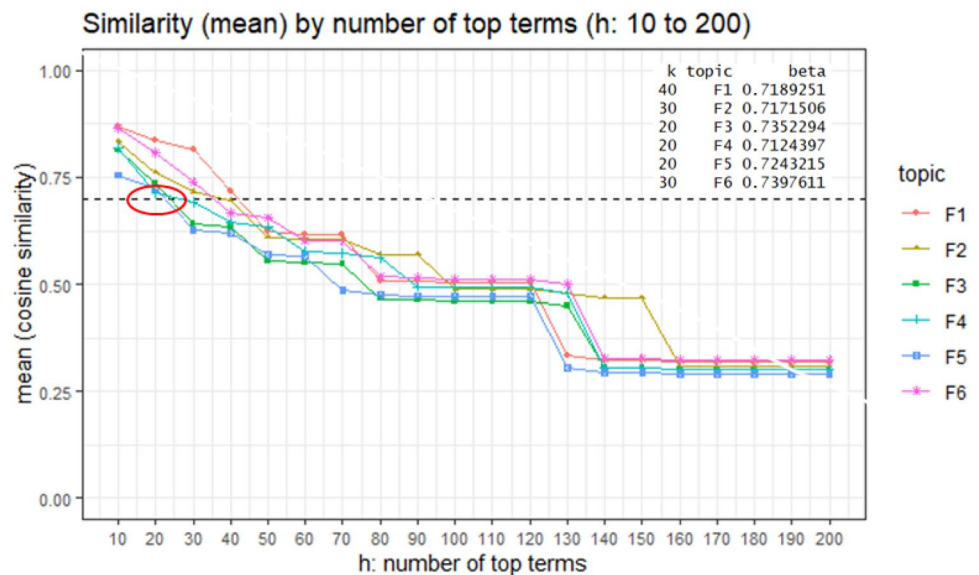
Several references based on associations between observation measures in organisational management, such as item validity and reliability (Cronbach’s alpha or composite), use 0.7 as the minimum threshold for acceptance. In fact, it is also common to use that value in studies in computation fields (Soni et al. 2016). In the present study, we fixed the minimum threshold to accept reliability based on the average cosine similarity at 0.7.

The framework suggests two aspects regarding the number of terms to use to represent the concepts under study: (1) understanding the behaviour of the cosine similarity average by several numbers of top terms; that is, the terms that are most likely to occur in each topic; and (2) identifying the lowest number that satisfies the defined threshold (0.7).

Figure 5 presents the results of the simulations, performed with LDA (under Gibbs’ method).

Figure 5 shows that the reliability (based on the mean cosine similarity) of the discovered patterns decreases when the number of top terms increases. That is, only a certain number of terms (the most likely terms) share a strong similarity. This is consistent with the practice of scale development and validation in management research because it is common to start with many items (or questions), and then, using qualitative and qualitative procedures, finish with a reduced scale. This is because only a small number of items satisfy several psychometric properties, such as reliability.

**Fig. 5** Behaviour of the cosine similarity mean by number of terms



In the present scenario, the situation is similar, because Fig. 5 shows that using a reduced number of terms, at least the top 20, ensures the minimal threshold (0.7) for the average cosine similarity for all patterns (topics). It is worth noting that topics F1 (operations analytics), F6 (customer analytics), and F2 (product analytics) present the most consistent values and, would thus allow the inclusion of more than 20 terms, however, based on the premises of parsimony and the Pareto principle (few—vital and many—trivial), this study used the same number of terms for all the patterns: the 20 most likely terms.

**Is it reasonable to assume construct validity for the discovered alternatives?** Construct validity means really measuring the theoretical concept that the study intends to measure, and no other concept. Construct validity is commonly explored through two manifestations: convergent and discriminant validity. Based on Rial et al. (2006), it is therefore possible to infer whether the observable measures share more information (convergent) than with the measures of other constructs (discriminant).

Considering the top terms chosen, the study explored discriminant validity through Eq. (3), which adopts the criterion of Fornell and Larker (1981) for the analysis of questionnaire data considering the covariance matrix. In this paper, however, we used the cosine similarity for each topic (alternatives), and expressed the response through the validity ratio (rv),

$$r_v = \frac{\text{Av.CoSim}(\text{Terms}_i)^2}{\text{Max}[\text{Av.CoSim}(\text{Terms}_i, \text{Terms}_j)^2]} \quad \forall i \neq j \quad (3)$$



where  $\text{Av.CoSim}(\text{Terms}_i)$  is the average cosine similarity among the representation vectors of top terms within topic  $i$  and  $\text{Av.CoSim}(\text{Terms}_i, \text{Terms}_j)$  is the average between topics ( $i \neq j$ ).

The Eq. (3) results are reasonable because Fornell and Larcker (1981) state that, in psychometric contexts, if the average variance explained for a latent factor is greater than the squared inter-construct correlations, there is reasonable evidence of discriminant validity (Aldás and Uriel 2017). In a context of linguistic computation, Washtell (2010, p. 49) also notes, “as with Pearson’s  $R^2$ , it [squared cosine similarity] represents the degree or proportion of similarity (consider that the square of an angle’s cosine and that of its sine total 1).” This is consistent with Wang et al.’s (2015) use of such a measure because they apply the squared cosine similarity to measure feature redundancy. In other words, Eq. (3) is consistent with the need to find more shared associations (similarity in this case) among the observable variables belonging to the same latent variable (topic in this case) than between those and other latent variables (Aldás and Uriel 2017). In the present non-parametric context of text analytics, when  $r_v$  is farthest (to the right) from 1 more exploratory evidence in favour of discriminant validity is found.

Table 3 provides the results of exploring the discriminant validity for the topics (discovered alternatives); we can interpret the diagonal [squared  $\text{Av.CoSim}(\text{term}_i)$ ] as the proportion of similarity among the terms (observable variables) of the same topic, and the other cells have such a proportion but apply to inter-patterns (topics) [squared  $\text{Av.CoSim}(\text{term}_i, \text{term}_j)$ ].

In Table 3, note that all the topics (discovered alternatives) present a ratio ( $r_v$ , see Eq. 3) greater than 1; for example, the proportion of similarity for F1 is 1.18 times greater than the maximum proportion of its competitors (0.593). In fact, all the exposed ratios ranged from 1.06 to 1.18 with a mean of 1.12. In other words, Table 3 shows that, in all cases, the observable variables (terms) for each topic share more similarity than exists between those and the terms of other topics.

Aldás and Uriel (2017, p. 543) state that there is convergent validity when different measures (e.g., indicators) of

the same construct share a large proportion of variance or are widely correlated. In fact, a traditional criterion for convergent validity in management research is that the proportion of variance that a latent factor explains is greater than 0.5 (Aldás and Uriel 2017). From this perspective, squared cosine similarity (the proportion of similarity that the terms of the same topic share) can be an exploratory measure of convergent validity in non-parametric scenarios of text analytics. Note that, in Table 3, the diagonals are greater than 0.5; that is, in all cases, the similarity among the observable measures of the same construct is greater than 50%.

In summary, in analogy with traditional concepts and metrics in management research (imported from psychometry), the patterns discovered using text analytics satisfy the fundamental properties of latent factors: content validity (reasonable affinity between terms and expected concepts), reliability ( $\text{Av.CosSim} > 0.7$ ), discriminant validity ( $r_v > 1$ ), and convergent validity ( $\text{Ave.CosSim}^2 > 0.5$ ). In the next section, we thus use the six topics as decision alternatives to help the strategists to make a more rational decision about the priority focus of future analytics initiatives.

### Discovering the criteria and their weights

The interviews conducted with the strategists considered another question: What should the three main differentiating factors of the organisation’s products, processes, or services be, so that they contribute to the strategic vision? To facilitate the finding of answers and assume a comprehensive perspective, we listed nine possible factors (with their meaning) for each strategist (e.g., efficiency, opportunity, customer focus, security). We chose an initial list of factors from previous analyses of the literature about total quality management (EFQM and Malcolm Baldrige, e.g., Winn and Cameron 1998; Maderuelo 2002; Rao et al. 1999; Tamimi 1995), and value addition (e.g., Gidey et al. 2014). The strategists, however, could also express and choose other factors that they considered to be relevant (two additional factors). At least 2 strategists chose 5/11 factors. We then used four normalisation methods (described in Munier et al. 2019) to estimate the criteria weights: M1 (sum of frequencies), M2 (dividing by maximum value), M3 (Euclidean formula),

**Table 3** Square of the cosine similarity mean for the top 20 between terms of the same topic (diagonal) and between different topics

Topics	F1	F2	F3	F4	F5	F6
F1	<b>0.701</b>	0.511	0.503	0.433	0.407	0.593
F2	0.511	<b>0.579</b>	0.389	0.372	0.347	0.493
F3	0.503	0.389	<b>0.540</b>	0.368	0.388	0.510
F4	0.433	0.372	0.368	<b>0.507</b>	0.393	0.462
F5	0.407	0.347	0.388	0.393	<b>0.524</b>	0.457
F6	0.593	0.493	0.510	0.462	0.457	<b>0.653</b>
Ratio $r_v$	1.18	1.13	1.06	1.10	1.15	1.10



and M4 (based on the range). We considered the number of strategists (Freq) who chose each criterion. Table 4 shows the criteria weights normalised by the four named methods.

Table 4 shows that, independently of the normalisation method used, C1 and C2 are the most important criteria, considering the critical factors that the strategists chose during the interviews. However, the advantage of probing several normalisation methods is that the distances between the criteria vary, and this can affect the final decision when including the alternatives. For example, from M1, based on the sum of frequencies, the distance between C1 and C2 is 0.047, but from M2 (dividing into the maximum value) and M4 (based on range), it is 0.143 and 1.0 respectively.

### Examining empirical reasons for the multi-criteria perspective

To empirically confirm the need to assume a multi-criteria viewpoint, we explore an individual perspective (a “better” alternative for the strategist) of the present decision-making case. Figure 6 provides the underlying preferences that each

strategist presented for the alternatives, which we calculated from text analytics using the sum of the frequencies of the top 20 terms on each page, then aggregating (sum) by strategists and normalising (0–1) based on the range. A score of 1 thus represents the best alternative from the strategists’ perspective, and 0 is the worst alternative.

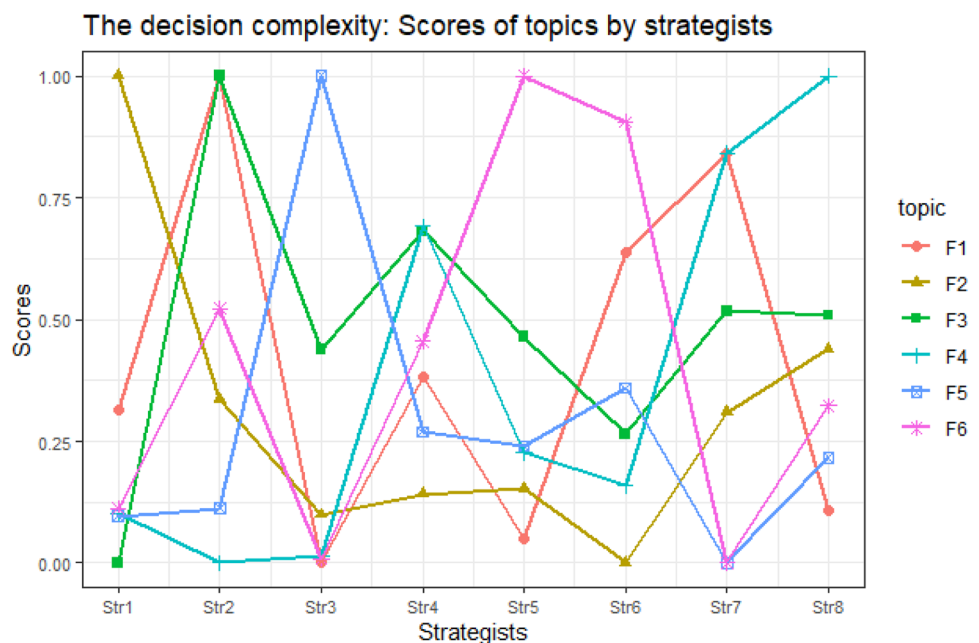
Figure 6 shows that the strategists have diverse underlying preferences regarding the alternative (topics). This shows notable complexity when a decision is necessary. Strategist 1 (Str1) prefers F2, Strategist 2 prefers F3 and F1, Strategist 3 tends towards F5, Strategist 4 prefers F3 and F4, Strategists 5 and 6 prefer alternative F6, Strategist 7 tends towards F1 and F4, and Strategist 8 prefers F3; the discrepancy is huge.

Figure 6 is consistent with the interpretative aspects of the observation scenario, which highlighted the complexity involved in achieving consensus among strategists about the priority analytics focus on the organisation. In other words, Fig. 6 symbolises an important meeting between eight strategists (with great decision-making power) who have different viewpoints regarding the organisation’s direction concerning analytics: that is, a complex meeting in which reaching

**Table 4** Relative importance of criteria using four normalisation methods

Criteria	Freq ( $f$ )	M1 (Sum) $f_i/\text{sum}(f_i)$	M2 (Largest val.) $f_i/\text{max}(f_i)$	M3 (Euclidean) $f_i/[\text{sum}(f_i^2)]^{1/2}$	M4 (Range) $[f_i - \text{min}(f_i)]/[\text{max}(f_i) - \text{min}(f_i)]$
C1	7	0.333	1.000	0.670	1.000
C2	6	0.286	0.857	0.575	0.800
C3	4	0.190	0.571	0.383	0.400
C4	2	0.095	0.286	0.192	0.000
C5	2	0.095	0.286	0.192	0.000

**Fig. 6** Underlying preferences of strategists (Str) for alternatives (topics: F1–F6)



a consensus may be a utopic challenge. The multi-criteria perspective employed by the proposed framework is therefore useful in order to contribute to the necessary consensus in an argumentative way.

### Discovering the scores of alternatives within each criterion using a text-driven approach

To contribute to the consensus between the strategists in a transparent and justified manner, we followed the proposed framework and assumed the five factors (words) chosen by the strategists and their synonyms as a set of seed terms (13 terms in total). Based on the instructions of Garten et al. (2018, distributed dictionary representation), and considering the corpus vocabulary (1938 terms in total), we created a representation vector for each criterion using the average of the cosine similarity between the seed terms within each criterion. We then calculated the cosine similarity between these criteria vectors and the other terms (non-stopword) in the entire corpus vocabulary. In total, 25 terms were chosen to represent the criteria (five for each criteria).

We built a normalised multi-criteria decision matrix for each strategist, in which the rows are the alternatives, the columns the criteria, and the cells the normalised scores (based on a range) that we derived from computing the alternatives for each criterion using a text-driven computational approach. We calculated each cell $_{ij}$  (it is  $S_{ij}$  in Eq. 1) using the cosine similarity (see Eq. 2) between vector representations of alternative  $j$  (considering terms representing topic  $j$ ) and criteria  $i$  (terms representing criteria  $i$ ). Note that in this text-driven computational approach, the strategists were not asked to obtain the alternative scores; they did not have to adapt their evaluation to predefined numerical/linguistic categories.

Figure 7 presents the results of evaluating the alternatives (topics) in each decision criterion.

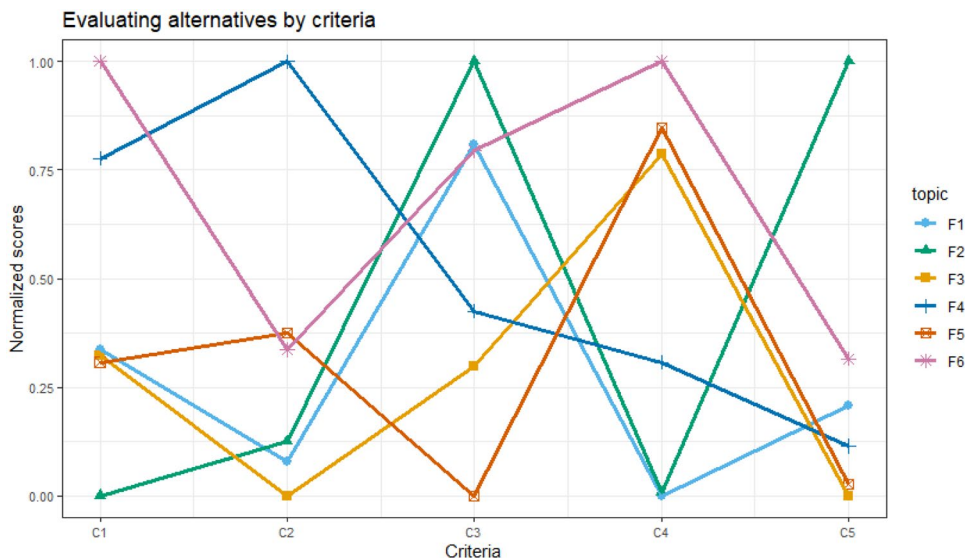
Note, in Fig. 7, that there are three dominated alternatives (F1: strategic analytics, F3: market analytics, and F5: product analytics) and three non-dominated alternatives (F2: human resource analytics, F4: operations analytics, and F6: customer analytics). It is therefore possible to reduce the problem of six alternatives and five criteria to a more practical problem with three alternatives (non-dominated) and five criteria.

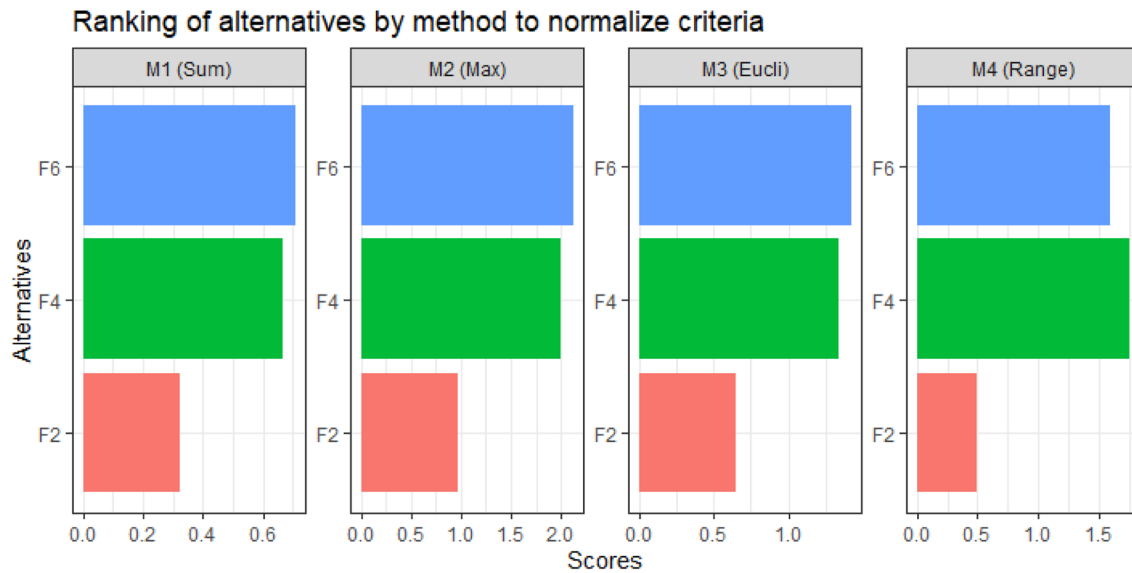
### Prioritising the discovered alternatives and choosing the “best” solution

We focused only on non-dominated alternatives “which are better in at least one criterion when compared to another alternative and there is no alternative better than them in all criteria” (Fontana and Cavalcante 2014, p. 1621). The scores of the alternatives are then calculated by means of a global indicator (weighted sum; see Eq. 1), considering the weights of the criteria (according to the normalisation method used; see Table 4). Next, the alternatives are ranked, and the “best” solution is identified (highest score obtained). Figure 8 provides a global weighted evaluation of the alternatives by the method used to normalise the criteria weights.

Figure 8 shows that it is possible to exclude the F2 alternative (human resource analytics) because F4 (operations analytics) and F6 (customer analytics) are greatly superior to it. It is worth noting that there are no notable differences between the last alternatives; and when using the M1, M2, and M3 methods, F6 is slightly superior to F4, but, when considering M4, F4 has a higher score than F6.

Fig. 7 Evaluation of the alternatives for the criteria





**Fig. 8** Ranking of the most relevant decision alternatives

In summary, at this point of the study, it is possible to note that, among the six initial alternatives that emerged from the documents (interviews with strategists), two are the most relevant (F6 and F4) for adoption in the current scenario of the organisation. However, remember that one fundamental restriction highlighted by Eq. (1) is that only one alternative must be chosen. Based only on Fig. 8, alternative F6 is therefore the most relevant. In order to consider other analytical results, however, the next section reports the sensitivity analyses using these three most relevant alternatives.

### Evaluating the solution robustness for the discovered solution

This stage examines the “best alternative” when using two additional criteria weightings (perceptual and simulated).

#### Considering a perceptual weighting of criteria

For this analysis, each strategist chose three criteria based on their perception of their relevance to achieving the strategic vision. They then assigned ratings: 5 (first most relevant), 3 (second most relevant), and 1 (third most relevant). We aggregated the scores and explored the sensitivity of the current solution to this new perceptual weighting scenario.

Table 5 compares the results for the M1 method (using the number of strategists who chose each criterion; see Table 4) and the M1' method (based on the perceptual rating from strategists).

The last column of Table 5 provides the difference between these two results, which range from -0.045 to 0.056. Note that these differences are not particularly great,

however, from a relative perspective, criterion C4 increased by 60% (0.152 vs 0.095), but the other criteria varied from -13.5% (C1) to 11.6% (C3). Figure 9 thus provides the ranking of alternatives in order to explore the sensitivity of the current solution to these new scenarios (perceptual rating from the strategists).

Figure 9 shows that the solution is still the same: F6 and F4 are the most relevant alternatives for the organisation, however, in this scenario, F4 is not superior to F6 in any case.

#### Considering a simulated weighting of criteria

This section explores the behaviour of the non-dominated alternatives, considering eight scenarios. Each scenario involves varying (simulations) the number of strategists in favour of each criterion, ranging from one to eight strategists. This procedure thus provides several weightings for the criteria, for example 1/8 (12.5%) to 8/8 (100%) for each criterion. Once again, we aggregated (see Eq. 1) the

**Table 5** Scores by criterion from the perceptions of strategists (M1') and the frequentist perspective (M1)

Criteria	Rating	M1'	M1 (see Table 4)	Difference M1'-M1
C1	19	0.288	0.333	- 0.045
C2	17	0.258	0.286	- 0.028
C3	14	0.212	0.190	0.022
C4	10	0.152	0.095	0.056
C5	6	0.091	0.095	- 0.004



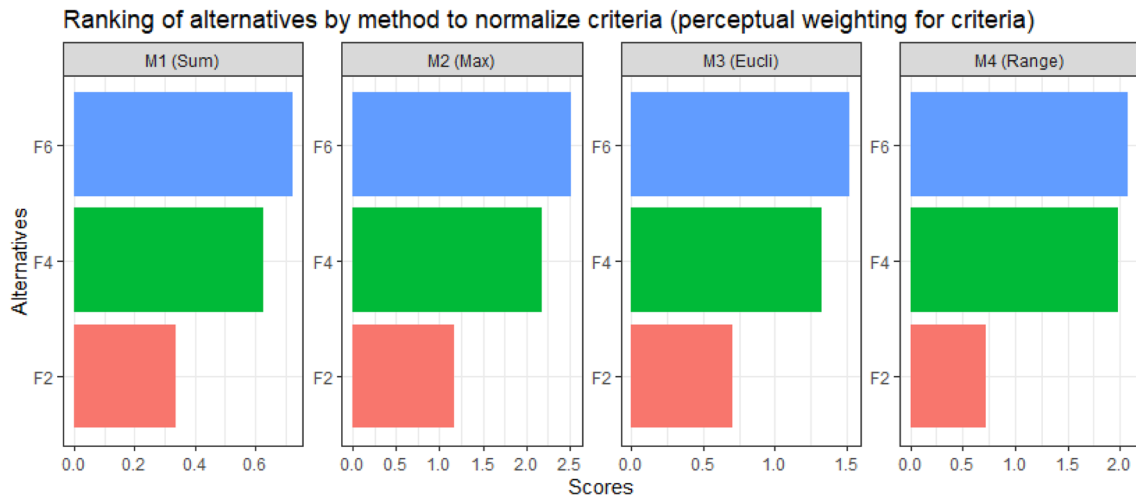


Fig. 9 Ranking of the most relevant alternatives in the scenario of perceptual punctuation of criteria

scores and then explored whether the current solution (F6) remained superior and stable compared to the other alternatives. Figure 10 shows the results of exploring the behaviour of F2, F4, and F6 considering the eight weighting scenarios.

Figure 10 shows that, in practically all the cases, F6 (customer analytics) remains superior and stable in comparison with the other alternatives. Similarly, F4 (operations analytics) is superior to F2 (human resource analytics). In other words, the conclusion that we obtained in previous sections regarding the “best” alternative remains consistent.

In summary, in the specific observation context, the sensitivity analysis supports the recommendation to focus the main organisational resources/efforts on developing future analytics initiatives in the domain of customer analytics (F6), in order to improve VGA.

### Conclusions and future work

This paper developed a novel mixed methodological framework comprising five stages and 12 steps that combine text analytics, interviews, and multi-criteria decision analysis. The framework includes procedures to explore four properties (reliability, content validity, novelty, and construct validity) based on traditional management research resources (e.g., surveys and psychometric considerations) and a text-driven computational approach (e.g., distributed dictionary representation, cosine similarity).

Despite the benefits of multi-criteria decision analysis for the rational decision-making process, this assumes that there is already a structured problem (e.g., objective, alternatives,

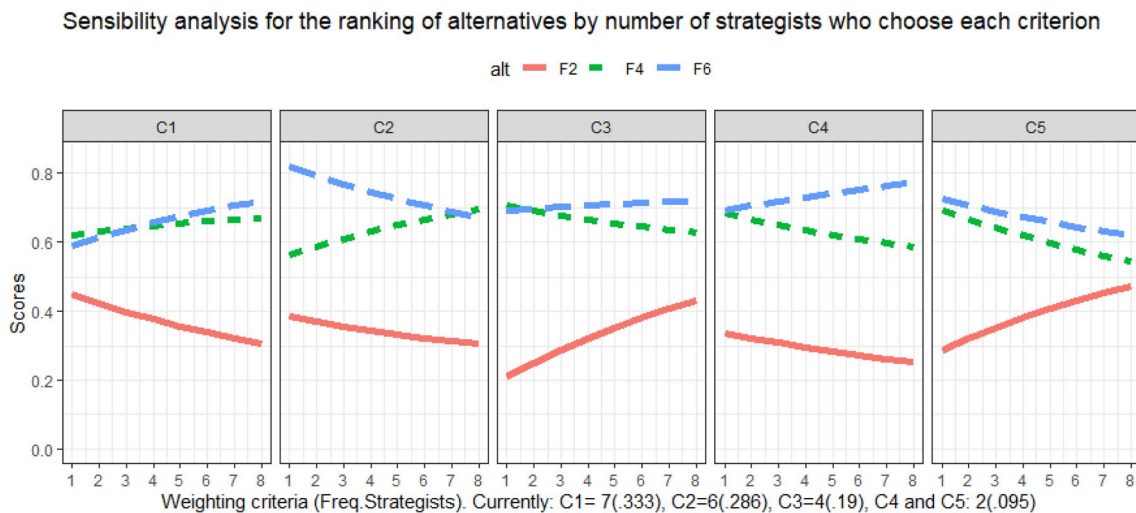


Fig. 10 Sensibility analysis for the most relevant alternatives using simulated changes in criteria weights





criteria, and criteria weights), which, according to Martunen et al. (2017), is not necessarily the case, as found in the observation scenario. Multi-criteria decision analysis is therefore traditionally combined with problem structuring methods, however, most of these (structuring) methods are supported only by interpretivism, so they are intensive in subjective appraisals, meanings, and so on (e.g., SWOT, Delphi, Stakeholder's analysis, brainstorming). However, the proposed framework helps to overcome these challenges using a text-driven computational approach by discovering valuable latent patterns from interview data (transcriptions), such as criteria, alternatives, criteria weights, and scores of the alternatives within the criteria. This favours efficiency, reproducibility, and transparency in the decision-making process.

This paper provides insights into a real critical problem in a strategic domain of analytics concerning how to discover, prioritise, and validate patterns involving several organisation strategists in order to improve competitive advantage. The deployment of the framework discovered six underlying topics (latent alternatives) from textual data (interview transcriptions). It validated these using content validity from the viewpoint of 24 professionals in management areas; similarly, it validated them through reliability and convergent/discriminant validity using measures/procedures based on semantic similarity and distributed dictionary representation. Furthermore, it analysed the novelty of the discovered topics through a bibliometric strategy supported by the number of scientific papers related to each topic. It also confirmed the complexity of making decisions in a strategic scenario of analytics, and identified, ranked, and argued the best decision alternative in the study scenario. We also tested the "best" alternative using sensitivity analysis to confirm its robustness when the weighting of criteria varies. This article thus contributes to understanding and addressing complex decision-making processes in organisational environments, from a multi-criteria perspective assisted by the leveraging of interview data using a text-driven computational approach.

This paper also found that customer analytics is an essential topic for future analytics initiatives in the observation context. It thus demonstrates that the managerial decision-making process can be improved (e.g., reliable, valid, robust, reproducible) by combining qualitative (e.g., interviews with organisational strategists) and quantitative data/techniques (e.g., text analytics, and multi-criteria analysis) using a text-driven computational approach. This application case can help managers/practitioners recognise underlying decision-making alternatives for future analytics initiatives and contribute to better VGA in the organisational arena. The discovered solution is consistent with Kitchens et al. (Kitchens et al. 2018, p. 545), who state, "By building on one of the firm's most important resources

(its customers) and one of its least imitable (its data), customer analytics represents an important strategic initiative with the potential to create significant and sustainable competitive advantage." Customer analytics is assumed beyond a "simple" action, method, or process; this is an integral domain that uses/influences several functional areas and can contribute to the generation of sustainable competitive advantage in the organisational context.

Based on this scope of the customer analytics concept and considering studies such as Louro et al. (2019), Sun et al. (2014), and Tandon (2021), there is a potential multi-criteria concern to be addressed with the proposed framework, regarding how to reconfigure the concepts, techniques, resources (including customer data), and capabilities of marketing and analytics to achieve an integrative framework that improves customer insight/satisfaction. To address this question using the proposed framework, a set of relevant information sources in an organisation can initially be interviewed (online or in person), textual data (interview transcriptions) can then be transformed in discovered patterns (alternatives, criteria and their weights, and scores -see Eq. (1)- of alternatives within criteria), and finally a global evaluation deployed for ranking alternatives and deciding. Another use of the proposed framework in the domain of customer analytics, based on Hallikainen et al. (2020), Patel et al. (2014), Xu et al. (2021), and Yerpude and Singhal (2019), is to stimulate interactions in real time between the customer and organisation, through the dynamic use of customer data and the organisation's communication channels (e.g., social media), so that additional sales of products/services are stimulated. Similarly, according to Jayaram et al. (2015) and Nauck et al. (2006), another multi-criteria challenge that can feasibly be addressed using the proposed framework, and taking into account company particularities, is determining which technologies/products for the use of customer data are the most appropriate to stimulate acceptance and use by the marketing team, in order to achieve a more timely and effective decision-making in matters related to the clients.

The five stages and 12 steps comprising the proposed methodological framework (see Fig. 1) are intensive in computational tasks for text processing/analysis, which contributes to overcoming the challenges attributed to mixed research which require high operations/time/resources and quantitative/qualitative skills (Krishen and Petrescu 2021; Molina-Azorin 2011).

The proposed framework can be used to improve the quality of decision-making processes, including those in the current/future virtual work environments (e.g., teleworking). For example, it is feasible and efficient to apply the framework under such an environment to capture and analyse textual data about other relevant organizational concerns. In an automated or semi-automated manner, the framework



offers guidelines to transform such text data in consensual latent patterns that reasonably represent alternatives, criteria, criteria weights, and scores of alternatives within criteria. According to the framework, these components are considered in a systematic manner to make relevant decisions in a more transparent, justified, efficient, and reproducible manner.

This paper also facilitates future dialogue and brings together work by companies, academics, consultants, and governments to potentiate better decision-making processes in person-based environments or virtual workplaces (e.g., teleworking).

The proposed methodological framework is also applicable to other types of data originated by natural language (e.g., open questions administered in questionnaires by mail, customer complaints and claims, employee suggestions), which contribute towards formulating and specifying strategic issues for organisations. For example, textual data from operational level employees (e.g., suggestions, expectations, needs), customers (e.g., complaints and claims), suppliers (e.g., supplier performance reports), and competing organisations (e.g., mission statements, descriptions of products and services, comments on social networks) can reveal a variety of decision alternatives unknown to strategists and, thus, nurture the decision-making processes around the generation of new competitive advantages. Future studies can reproduce/improve the proposed framework using these other textual data for other relevant organisational concerns, and other companies/domains and methods. Future studies can reproduce/improve the proposed framework these other textual data other relevant organisational concerns, and other companies/domains and methods.

**Acknowledgements** Minciencias: Ministry of Science, Technology and Innovation, Colombia.

## Declarations

**Conflict of interest** The authors have no conflict of interest to declare.

## References

- Aldás, J., and E. Uriel. 2017. Análisis multivariante aplicado con R. Ediciones Paraninfo, SA.
- Amado, A., P. Cortez, P. Rita, and S. Moro. 2018. Research trends on big data in marketing: A text mining and topic modeling based literature analysis. *European Research on Management and Business Economics* 24 (1): 1–7.
- Ansoff, H. 1980. Strategic issue management. *Strategic Management Journal* 1 (2): 131–148.
- Baaij, M., and P. Reinmoeller. 2018. Spotting the real strategic issues and developing superior insights. In *Mapping a winning strategy: developing and executing a successful strategy in turbulent markets*, 61–91. Bingley: Emerald Publishing Limited. <https://doi.org/10.1108/978-1-78756-129-820181009>.
- Barney, J. 1991. Firm resources and sustained competitive advantage. *Journal of Management* 17 (1): 99–120.
- Becker, D. 2017. Predicting outcomes for big data projects: Big data project dynamics (BDPD): Research in progress. In *2017 IEEE international conference on big data (big data)*, pp. 2320–2330. IEEE.
- Burnett, J. 1998. A strategic approach to managing crises. *Public Relations Review* 24 (4): 475–488.
- Campion, M., M. Campion, E. Campion, and M. Reider. 2016. Initial investigation into computer scoring of candidate essays for personnel selection. *Journal of Applied Psychology* 101 (7): 958–975.
- Chakrabarty, S. 2018. Cosine similarity approaches to reliability of Likert scale and items. *Romanian Journal of Psychological Studies (RJPS)* 6 (1): 3–16.
- Chang, J., S. Gerrish, C. Wang, J. Boyd-Graber, and D. Blei. 2009. Reading tea leaves: How humans interpret topic models. In *Advances in neural information processing systems*, ed. M. Jordan, Y. LeCun, and S. Solla. vol. 22, pp. 288–296.
- Charest, M. 2007. Intelligent data mining assistance via case-based reasoning and a formal ontology (Doctoral dissertation, Université du Québec à Trois-Rivières). <http://depot-e.uqtr.ca/id/eprint/1537/>.
- Collins, C. J., and K. D. Clark. 2003. Strategic human resource practices, top management team social networks, and firm performance: The role of human resource practices in creating organizational competitive advantage. *Academy of Management Journal* 46 (6): 740–751.
- Cooper, A. 2012. What is analytics? Definition and essential characteristics. *CETIS Analytics Series* 1 (5): 1–10.
- Curtis, M., and B. Perryman. 2021. Model lifetime value; assessing the enterprise-wide inputs to data science model effectiveness. Assessing the enterprise-wide inputs to data science model effectiveness (April 20).
- Daniilova, K., J. Iden, D. Knudsen, C. Costabile, T. Johannessen, B. Johanson, and A. Ulfsten. 2020, November. Virtual work during covid-19 lockdown—lessons learned. In *Norsk konferanse for organisasjoners bruk av IT* (Vol. 28, No. 1).
- DeKroon, N., and B. Karp. 2013. An auditor's guide to data analytics, 2013. [https://chapters.theiia.org/raleigh-durham/Events/Training%20Presentations/2013\\_May\\_Raleigh%20IIA%20Presentation\\_Data%20Analysis.pdf](https://chapters.theiia.org/raleigh-durham/Events/Training%20Presentations/2013_May_Raleigh%20IIA%20Presentation_Data%20Analysis.pdf). Accessed 23 June 2020
- Deshpande, A. 2016. A qualitative examination of challenges influencing doctoral students in an online doctoral program. *International Education Studies* 9 (6): 139–149.
- Dooley, R., and G. Fryxell. 1999. Attaining decision quality and commitment from dissent: The moderating effects of loyalty and competence in strategic decision-making teams. *Academy of Management Journal* 42 (4): 389–402.
- Dutton, J., L. Fahey, and V. Narayanan. 1983. Toward understanding strategic issue diagnosis. *Strategic Management Journal* 4 (4): 307–323.
- Ertemel, A. V. 2015. Consumer insight as competitive advantage using big data and analytics. *International Journal of Commerce and Finance* 1 (1): 45–51.
- Fagerjord, A., and L. Kueng. 2019. Mapping the core actors and flows in streaming video services: What Netflix can tell us about these new media networks. *Journal of Media Business Studies* 16 (3): 166–181.
- Floyd, S. W., and B. Wooldridge. 1992. Managing strategic consensus: the foundation of effective implementation. *Academy of Management Perspectives* 6 (4): 27–39.
- Fontana, M., and C. Cavalcante. 2014. Use of Promethee method to determine the best alternative for warehouse storage location assignment. *The International Journal of Advanced Manufacturing Technology* 70 (9–12): 1615–1624.



- Fornell, C., and D. Larcker. 1981. Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research* 18 (1): 39–50.
- Garten, J., J. Hoover, K. Johnson, R. Boghrati, C. Iskiwitch, and M. Dehghani. 2018. Dictionaries and distributions: Combining expert knowledge and large scale textual data content analysis. *Behavior Research Methods* 50 (1): 344–361.
- Gidey, E., K. Jilcha, B. Beshah, and D. Kitaw. 2014. The plan-do-check-act cycle of value addition. *Industrial Engineering & Management* 3 (124): 2169–2316.
- Gong, Y., and C. Poellabauer. 2017. Topic modeling based multimodal depression detection. In *AVEC 2017—Proceedings of the 7th Annual Workshop on Audio/Visual Emotion Challenge, co-located with MM 2017*, pp. 69–76.
- Grinko, A., T. Bochulia, P. Hrynko, I. Yasinetska, and I. Levchenko. 2017. Forming concept of intellectualization information provision of managing an enterprise. *Eastern-European Journal of Enterprise Technologies* 5 (3): 4–14.
- Grošelj, P., and L. Stirn. 2017. Soft consensus model for the group fuzzy AHP decision making. *Croatian Operational Research Review*, 207–220.
- Hagen, J. (2021, January). Collaboration for big data analytics: Investigating the (troubled) relationship between data science experts and functional managers. In *Proceedings of the 54th Hawaii international conference on system sciences*, p. 254.
- Hajkowicz, S., and C. Roy. 2013. Forecasting the future. *Superfunds Magazine* 385: 18–20.
- Hallikainen, H., E. Savimäki, and T. Laukkanen. 2020. Fostering B2B sales with customer big data analytics. *Industrial Marketing Management* 86: 90–98.
- Harrison, R.L. 2013. Using mixed methods designs in the Journal of Business Research, 1990–2010. *Journal of Business Research* 66 (11): 2153–2162.
- Heene, A. 1997. The nature of strategic management. *Long Range Planning* 30 (6): 933–938.
- Holm, D. B., R. Drogendijk, and H. ul Haq. 2020. An attention-based view on managing information processing channels in organizations. *Scandinavian Journal of Management* 36 (2): 101106.
- Hou-Liu, J. 2018. Benchmarking and improving recovery of number of topics in latent Dirichlet allocation models, Corpus ID: 54092738, <https://www.semanticscholar.org>
- Hungerford, N. 2017. The biggest science project on the planet' was on a racetrack on Sunday. Published sun, Sep 17, 7:53 pm. <https://www.cnbc.com/2017/09/16/big-data-in-formula-one-mclaren-applied-technologies-at-singapore-summit.html>. Accessed 20 June 2020
- Jayaram, D., A. Manrai, and L.A. Manrai. 2015. Effective use of marketing technology in Eastern Europe: Web analytics, social media, customer analytics, digital campaigns, and mobile applications. *Journal of Economics, Finance and Administrative Science* 20 (39): 118–132.
- Kabanda, G. 2019. An evaluation of big data analytics projects and the project predictive analytics approach. *Oriental Journal of Computer Science and Technology* 12 (4): 132–146.
- Kazemzadeh, A., J. Gibson, P. Georgiou, S. Lee, and S. Narayanan. 2016. A Socratic epistemology for verbal emotional intelligence. *PeerJ Computer Science* 2: e40.
- Kee, K., M. Thompson-Hayes. 2012. Conducting effective interviews about virtual work: Gathering and analyzing data using a grounded theory approach. In *Virtual work and human interaction research*, pp. 192–212. IGI Global.
- Kellermanns, F., S. Floyd. 2005. Strategic consensus and constructive confrontation: Unifying forces in the resource accumulation process. *Innovating Strategy Process*, 149–162.
- Khan, Z., and T. Vorley. 2017. Big data text analytics: An enabler of knowledge management. *Journal of Knowledge Management* 21 (1): 18–34.
- Kitchens, B., D. Dobolyi, and A. Abbasi. 2018. Advanced customer analytics: Strategic value through integration of relationship-oriented big data. *Journal of Management Information Systems* 35 (2): 540–574.
- Kreković, G., A. Pošćić, and D. Petrinović. 2016. An algorithm for controlling arbitrary sound synthesizers using adjectives. *Journal of New Music Research* 45 (4): 375–390.
- Krishen, A.S., and M. Petrescu. 2021. Interdisciplinary research as methodologically and substantively creative. *Journal of Marketing Analytics* 9: 1–2.
- Kruse, S. 2008. Reflections on leading and learning for change: An introduction. *Education and Urban Society* 40 (6): 655–669.
- Larson, D. 2019. A review and future direction of business analytics project delivery. In *Aligning Business Strategies and Analytics*, pp. 95–114. Cham: Springer.
- Liman, A., M. Aliyu, and M. Halliru. 2020. Moderating role of job satisfaction on the relationship between multi-level marketing and distributors retention. *Ilorin Journal of Human Resource Management* 4 (1): 160–170.
- Louro, A., M. Brandão, J. Jaklič, and A. Sarcinelli. 2019. How can customer analytics capabilities influence organizational performance? A moderated mediation analysis. *BBR: Brazilian Business Review* 16 (4): 369–382.
- Maderuelo, J. 2002. Gestión de la calidad total: El modelo EFQM de excelencia. *Medifam* 12 (10): 41–54.
- Markham, S.K., M. Kowolenko, and T.L. Michaelis. 2015. Unstructured text analytics to support new product development decisions. *Research-Technology Management* 58 (2): 30–39.
- Marttunen, M., J. Lienert, and V. Belton. 2017. Structuring problems for multicriteria decision analysis in practice: A literature review of method combinations. *European Journal of Operational Research* 263 (1): 1–17.
- Mathaisel, D.F., and C.L. Comm. 2021. Political marketing with data analytics. *Journal of Marketing Analytics* 9 (1): 56–64. <https://doi.org/10.1057/s41270-020-00097-1>.
- Molina-Azorin, J.F. 2011. The use and added value of mixed methods in management research. *Journal of Mixed Methods Research* 5 (1): 7–24.
- Moulin, B. 1992. Conceptual-graph approach for the representation of temporal information in discourse. *Knowledge-Based Systems* 5 (3): 183–192.
- Munier, N., E. Hontoria, and F. Jiménez-Sáez. 2019. Strategic approach in multi-criteria decision making. Berlin: Springer
- Musaji, S., W. Schulze, and J. De Castro. 2020. How long does it take to get to the learning curve? *Academy of Management Journal* 63 (1): 205–223.
- Nacarelli, V., and D. Gefen. 2021. Trustworthiness and the Adoption of Business Analytics. *Information Systems Management*, 1–15.
- Naotunna, N., and H. Priyankara. 2020. The impact of telework on creativity of professional employees in Sri Lanka: Componential and social cognitive theoretical views. *International Journal of Mobile Learning and Organisation* 14 (3): 357–369.
- Nauck, D., D. Ruta, M. Spott, and B. Azvine. 2006, September. A tool for intelligent customer analytics. In *2006 3rd International IEEE Conference Intelligent Systems*, pp. 518–521. IEEE.
- Nikita, M. 2020. ldatuning: Tuning of the latent Dirichlet allocation models parameters. R package version 1.0.2. <https://CRAN.R-project.org/package=ldatuning>
- Park, G., S. Park, L. Khan, and L. Chung. 2017. IRIS: A goal-oriented big data analytics framework on Spark for better Business decisions. In *2017 IEEE International Conference on Big Data and Smart Computing (BigComp)*, pp. 76–83. IEEE.



- Patel, A., H. Gheewala, L. Nagla. 2014, March. Using social big media for customer analytics. In *2014 Conference on IT in Business, Industry and Government (CSIBIG)*, pp. 1–6. IEEE.
- Pavlou, P., and O. El Sawy. 2011. Understanding the elusive black box of dynamic capabilities. *Decision Sciences* 42 (1): 239–273.
- Pérez Rave, J., M. Trujillo, G. Castro, and G. Gómez. 2015. Multi-criteria modelling of mercury pollution prevention level in dental institutions. *Ingeniare. Revista Chilena De Ingeniería* 23 (1): 128–144.
- Pham, V., and A. Hoai. 2021, March. A study of facebook-based peer comments on L2 writing. In *17th international conference of the asia association of computer-assisted language learning (AsiaCALL 2021)*, pp. 114–120. Atlantis Press.
- Rao, S. S., L. E. Solis, T. S. Rangunathan. 1999. A framework for International Quality Management research: Development & Validation of a measurement instrument. *Total Quality Management*, 10 (7): 1047–1075.
- Rial, A., J. Varela, J. Avalo, and J. Lévy. 2006. El Análisis factorial. Modelización con estructuras de covarianzas en ciencias sociales. Temas esenciales, avanzados y aportaciones especiales, Netbiblo.
- Rubin, H., and I. Rubin. 1995. *Qualitative interviewing: The art of hearing data*. Sage Publications.
- Sandul, V. 2015. Some aspects of organization and work of scientific and industrial staff of nuclear power enterprises. *Nuclear Energy and the Environment*, 2(6), UDC 621.039. [http://dnic.com.ua/\\_ld/0/81\\_No\\_6\\_ENG.pdf#page=20](http://dnic.com.ua/_ld/0/81_No_6_ENG.pdf#page=20).
- Sagiroglu, S., and D. Sinanc. (2013, May). Big data: A review. In *2013 international conference on collaboration technologies and systems (CTS)*, pp. 42–47. IEEE.
- Sekaran, U., and R. Bougie. 2016. *Research methods for business: A skill building approach*. Wiley.
- Senge, P. 1996. Systems thinking. *Executive Excellence* 13 (1): 15–16.
- Shah, S., A. Gochtovt, and G. Baldini. 2019. Importance of project management in business analytics: academia and real world. In *Aligning business strategies and analytics*, pp. 81–94. Cham: Springer.
- Snyder, K. 2021. The digital art therapy frame: creating a ‘magic circle’ in teletherapy. *International Journal of Art Therapy*, 1–7.
- Soni, A., D. Viswanathan, J. Shavlik, S. Natarajan. 2016. Learning relational dependency networks for relation extraction. In *International Conference on Inductive Logic Programming*, pp. 81–93. Cham: Springer.
- Stacey, T. 2009. Design-a strategic issue. *Journal of the Southern African Institute of Mining and Metallurgy* 109 (3): 157–162.
- Stagner, R. 1969. Corporate decision making: An empirical study. *Journal of Applied Psychology* 53 (1): 1.
- Strohmeier, S., and F. Piazza. 2013. Domain driven data mining in human resource management: A review of current research. *Expert Systems with Applications* 40 (7): 2410–2420.
- Sun, N., J.G. Morris, J. Xu, X. Zhu, and M. Xie. 2014. iCARE: A framework for big data-based banking customer analytics. *IBM Journal of Research and Development* 58 (5/6): 4–1.
- Tamimi, N. 1995. An empirical investigation of critical TQM factors using exploratory factor analysis. *International Journal of Production Research*, 33 (11): 3041–3051.
- Tandon, U. 2021. Predictors of online shopping in India: An empirical investigation. *Journal of Marketing Analytics* 9 (1): 65–79.
- Teece, D., G. Pisano, and A. Shuen. 1997. Dynamic capabilities and strategic management. *Strategic Management Journal* 18: 509–534.
- Tiell, L. 2017. An ethnographic case study on the phenomena of blended learning teachers (Doctoral dissertation, University of Findlay).
- van Rijmenam, M. 2013. T-mobile USA cuts down churn rate by 50% with big data. PubliSehd 7:00 PM, February 14. <https://datafloq.com/read/t-mobile-usa-cuts-downs-churn-rate-with-big-data/512>. Accessed 20 Jun 2020
- Wang, C. 2019. Beyond Technology: Design a Value-Driven Integrative Process Model for Data Analytics (Doctoral dissertation, Robert Morris University).
- Wang, D., F. Nie, and H. Huang. 2015. Feature selection via global redundancy minimization. *IEEE Transactions on Knowledge and Data Engineering* 27 (10): 2743–2755.
- Washtell, J. 2010. Expectation vectors: A semiotics inspired approach to geometric lexical-semantic representation. In *Proceedings of the 2010 workshop on geometrical models of natural language semantics*, pp. 45–50. Association for Computational Linguistics
- Winn, B., and K. Cameron. 1998. Organizational quality: An examination of the Malcolm Baldrige national quality framework. *Research in Higher Education*, 39 (5): 491–512.
- Winter, S. G. 2003. Understanding dynamic capabilities. *Strategic Management Journal* 24: 991–995.
- Xu, Z., C. Vail, A.S. Kohli, and S. Tajdini. 2021. Understanding changes in a brand’s core positioning and customer engagement: A sentiment analysis of a brand-owned Facebook site. *Journal of Marketing Analytics* 9 (1): 3–16.
- Yerpude, S., and T. Singhal. 2019. “Custolytics”: Internet of Things based customer analytics aiding customer engagement strategy in emerging markets—an empirical research. *International Journal of Emerging Markets*.
- Zakir, J., T. Seymour, and K. Berg. 2015. Big data analytics. *Issues in Information Systems* 16 (2): 81–90.
- Zhang, H., G. Kou, and Y. Peng. 2019. Soft consensus cost models for group decision making and economic interpretations. *European Journal of Operational Research* 277 (3): 964–980.

**Publisher’s Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Jorge Iván Pérez Rave, PhD(c)**. Industrial Engineer (Universidad de Antioquia, Colombia). Specializations: (1) Statistics, and (2) Systems Engineering (Universidad Nacional de Colombia). Masters: (1) Systems Engineering (Universidad Nacional de Colombia), and (2) Visual Analytics and Big Data (UNIR, Spain). Ph.D candidate (Systems Engineering, Universidad Nacional de Colombia), and Ph.D candidate (Business Management; Universidad de Valencia, Spain). Director of the IDINNOV research group, IDINNOV S.A.S.

**Gloria Patricia Jaramillo Álvarez, PhD**. Civil Engineer (Universidad Nacional de Colombia), PhD. in Planning and Management of Hydraulic Resources (Universidad Politécnica de Valencia). Professor of Decision Analysis. Universidad Nacional de Colombia, Department of Computer Science, Medellín.

**Juan Carlos Correa Morales, PhD**. Statistician (Universidad de Medellín), Master of Statistics (University of Kentucky), PhD. in Statistics (University of Kentucky). Professor of School of Statistics, Universidad Nacional de Colombia, Medellín, Colombia.

