Project Leadership and Society 4 (2023) 100093

Contents lists available at ScienceDirect

Project Leadership and Society

journal homepage: www.sciencedirect.com/journal/project-leadership-and-society

Empirical Research Paper

Unsupervised machine learning for project stakeholder classification: Benefits and limitations

Costanza Mariani *, Yuliya Navrotska, Mauro Mancini

Department of Management, Economics and Industrial Engineering, Politecnico di Milano Via Lambruschini, 4/B, 20156, Milano, Italy

A R T I C L E I N F O

Keywords:
Unsupervised machine learning
Stakeholder management
Projects

A B S T R A C T

The literature has shown that an accurate classification of project stakeholders allows for more comprehensive planning of their management strategies. The most used classification methods have limitations stemming from using a small number of stakeholder attributes thus returning high-level and imprecise classification results. This work investigates the potential benefits and limitations of adopting unsupervised machine learning clustering as an alternative method to automatically recognize stakeholder groups. The paper demonstrates the application of a PAM algorithm for project stakeholder classification, employing qualitative and quantitative data collected from a real project in an IT Italian company. The results show that the use of unsupervised clustering leads to a more granular and detailed stakeholder grouping that enables the design of better refined and customized stakeholder management strategies. Furthermore, the results of the paper demonstrate that the use of this methodology, when data is taken from a structured dataset, reduces the degree of subjectivity in classification, promoting a data-driven approach to project stakeholder management.

1. Introduction

1.1. The role of artificial intelligence in supporting new stakeholder landscape challenges

In the last decade, many authors and practitioners in the project management field have highlighted the significant importance of stakeholder involvement in order to successfully achieve project outcomes (Olander and Landin, 2005). Projects impact multiple stakeholders with divergent interests, objectives, and socio-cultural backgrounds. Their willingness to get involved, contribute, and accept project outcomes directly impacts project success (Cleland, 1988; Turner and Zolin, 2012; Eskerod et al., 2015; Aaltonen et al., 2015). Consequently, project organizations are increasingly urged to devote attention to the stakeholder management process which assumes strategic importance in order to ensure an alignment of goals between the project and its stakeholders in the co-creation of a unique project's result (Jepsen and Eskerod, 2009). The importance of a structured stakeholder management process that includes the identification, analysis, and definition of tailored communication strategies has been highlighted in the literature (Lehtinen and Aaltonen, 2020), emphasizing how this can result in a more inclusive decision-making process that enhances stakeholder participation and leads to more sustainable project practices and results (Jepsen and Eskerod, 2009; Di Maddaloni and Davis, 2018; Derakhshan et al., 2019; Di Maddaloni and Sabini, 2022).

Nowadays, project stakeholder landscapes are becoming more challenging: in the last decades, organizations have faced unprecedented waves of changes propelled by globalization and by the spread of disruptive technologies that have impacted project management, accelerating project delivery and product life cycles (Abyad, 2017; Whyte, 2019). In this context, artificial intelligence systems defined as "systems that display intelligent behaviour by analysing their environment and taking actions - with some degree of autonomy - to achieve specific goals"(OECD, 2022), are gaining relevance in companies and project management where these innovative technologies already find some applications in predicting, detecting, and simulating processes in project planning and execution (Holzmann et al., 2022; Fridgeirsson et al., 2021; Magaña Martínez and Fernandez-Rodriguez, 2015). Among the available AI approaches, Machine learning is one of the most widely used applications in business contexts when the objective is to retrieve purposeful knowledge from a set of data (Shrestha et al., 2021; Asad et al., 2020). Machine Learning algorithms can be categorized into supervised and unsupervised; in supervised learning the analyst provides the computer with complete examples to be used as instructions to

* Corresponding author.
E-mail addresses: Costanza.Mariani@polimi.it (C. Mariani), yuliya.navrotska@mail.polimi.it (Y. Navrotska), Mauro.Mancini@polimi.it (M. Mancini).

https://doi.org/10.1016/j.plas.2023.100093
Received 13 March 2023; Received in revised form 1 August 2023; Accepted 2 August 2023
Available online 3 August 2023

perform the required task, while in unsupervised learning the machine works without any labelled example and autonomously recognize patterns in the dataset provided. Thus, supervised machine learning techniques encompass classification and regression analysis in which the machine is instructed to automatically make predictions about output variables of a system based on a set of ideal examples. On the contrary, in unsupervised learning the machine is provided with unlabelled data with the aim of identifying hidden patterns or logical structures, such as clusters, in the dataset provided (Ray, 2019).

1.2. Proposing unsupervised machine learning for stakeholder classification

The current applications of Machine Learning in project management literature concern the "hard" planning and controlling side of a project, allowing predictions to be made about project progress, effort, duration, risk and final outcomes achievement (Holzmann et al., 2022; Bento et al., 2022), (Flyvbjerg, Bent; Budzier, Alexander; Chun-kit, Ricky Lau; Agard, Karlene; Agard, Karlene; Leed, 2022). Specifically, AI methods such as probabilistic models, fuzzy theory, machine learning, and neural networks are reported to be employed in project risk management for risks' assessment and identification (Taroun & Yang, 2013; Wang & Jin, 2019) and mitigation action selection (Mariani & Mancini, 2023); at the same time also project scheduling (Faghihi et al., 2015), effort estimation (Elish et al., 2013; Bisi & Goyal, 2016; Moradbeiky, 2023) and cost budgeting (Afzal et al., 2019; Cheng et al., 2015; Kamoona & Budayan, 2019) seem to be fertile ground for the application of AI. While all these applications concern quantitative project management areas, this paper aims to spot the light on the possible benefits and limitations that can results from applying unsupervised machine learning to one of the most qualitative areas of project management, namely stakeholder engagement, typically lacking quantitative assessments. Indeed, unsupervised machine learning algorithms enable to perform a "pattern spotting" among unlabelled data, that can be particularly beneficial during stakeholder analysis when stakeholders need to be mapped and classified considering certain pre-defined attributes.

To date, several theoretical stakeholder classifications models exist in the literature and are employed in practice, like the salience model (Mitchell et al., 1997) and the power interest matrix (Mendelow, 1981), however, some authors have highlighted relevant limitations. Many of the models rely dominantly on project team members' brainstorming and checklists thus, they suffer from being biased by the subjectivity of the evaluator (Pérez Vera, 2018); in addition, the currently available frameworks often present simplified stakeholder classifications that consider few descriptive attributes to perform the classification (Mainardes et al., 2012).

The aim of this work is to fill these gaps by investigating the benefits and limitations of adopting unsupervised machine learning clustering as an alternative method to identify stakeholder groups; Therefore, coherently with the objectives stated, the following research question has been formulated:

RQ. How can the application of unsupervised machine learning support project stakeholder classification?

The paper is structured in the following way: the literature review section revises stakeholder theory, highlighting the limitations of the existing classification methods and the potential of employing unsupervised machine learning in the classification process. The following section presents a case of unsupervised clustering application in an IT Italian project stakeholder classification. Finally, the last part shows the case results and uncovers contributions to the theory and to practitioners, discussing the generalizability of the results, the limitations, and the possible future research advancements.

2. Literature review

2.1. Project stakeholder classification methods: insights from general stakeholder theory

Stakeholder management has received relevant interest in both the strategic management literature and in the project management literature (Eskerod et al., 2015; Huemann et al., 2016). A project is considered as a temporary organization (Lundin and Söderholm, 1995) established to create benefits, goods or services; to this end, the project needs a consistent number of resources that may come from individuals, groups, or entities – namely stakeholder - which may affect or be affected by the project (Andersen, 2008; Freeman, 1984). Consequently, many authors have focused on how to develop an inclusive involvement of project stakeholder in both research and practice, often referring to general stakeholder theory ((Aaltonen and Kujala, 2010; Jepsen and Eskerod, 2009; Vuorinen and Martinsuo, 2018; Lehtinen and Aaltonen, 2020; Machiels et al., 2023)) and highlighting both the benefits and possible drawbacks of inclusive stakeholder management strategies (Eskerod et al., 2016). Several authors have pointed out the importance of project stakeholder classification, one of the main component of stakeholder management, highlighting the need for further knowledge on the topic (Elias et al., 2002; Jepsen and Eskerod, 2009). In fact, despite the importance of classifying project stakeholders, the literature on the subject is rather limited and the existing classifications frameworks mostly refer to general stakeholder theory (Jepsen and Eskerod, 2009; Aaltonen and Kujala, 2016; Freeman, 1984) in his fundamental work entitled *Stategic Management: A Stakeholder approach*, considered as one of the foundations of stakeholder theory, highlights that stakeholders having similar interests or rights should be classified into different groups and managed through diversified relational approaches. This idea culminated in the development of the stakeholder theory that focuses on how organizations engage in relationships with diverse stakeholder groups and on how the managerial decision-making process can be influenced by stakeholder interests and claims (Jones and Wicks, 1999; Phillips et al., 2010; Eskerod et al., 2015).

2.2. Overview of the extant classification frameworks

Within this broad theoretical framework, many authors have focused on highlighting how diverse stakeholder group interact with the company, providing for this purpose different classification frameworks. The literature on stakeholder classification can be conceptually divided into (i) theoretical categorization of stakeholders based on the possession of some attributes or certain variables and (ii) frameworks or models that seek to quantify the intensity of an attribute or characteristics on a scale that enables a prioritized stakeholder assessment and management. The first category includes papers that theoretically group stakeholders based some attributes or variables such as (Clarkson, 1995) who groups stakeholders into primary and secondary considering the formal or informal contractual relationship established with the company or (Sirgy, 2002) who classifies stakeholders into internal or external considering as a criterion the amount of resources that are exchanged between the company and the stakeholders. Table 1 summarizes and shortly describe the stakeholder classification frameworks identified in literature, outlining whether they allow to perform a stakeholder prioritization.

Although these classifications are relevant in defining whom stakeholders are, and highlighting their interests and possible claims, they do not provide any information regarding how to deal with all of them simultaneously. According to (Fassin, 2009) this is not possible in the absence of an evaluation of some utilization criteria that define a prioritization among different stakeholder groups. Hence, to fulfil this theoretical requirement, the literature proposes (ii) frameworks and models that seek to quantify the possession of one or more attributes to perform a stakeholder mapping and classification that enables to

Table 1
Stakeholder characteristics.

Author	Classification criteria	Stakeholder Prioritization
Goodpaster (1991)	The strategic and the moral stakeholder	No
(Savage et al., 1991)	Stakeholder's potential powers to threaten or cooperate with the organization	No
Clarkson (1995)	The primary and the secondary	No
(Mitchell et al., 1997)	Power, legitimacy and urgency	Yes
Rowley (1997)	Network density and the centrality of the organization focus	No
Luoma and Goodstein (1999)	Private and public stakeholders	No
Sirgy (2002)	Internal, external, distal (those who can indirectly influence the survival and growth of the business firm through influence exerted on the firm's external groups)	No
(Post et al., 2002)	Resource providers, industry participants, socio-political stakeholders	No
(Mendelow, 1981; Johnson et al., 2008)	Matrix representing the level of stakeholder power and interest	Yes
(Van Der Laan et al., 2008)	Primary and secondary stakeholders based on the entity of the relationship with the corporation	No
Fassin (2009)	Classical stakeholders, stakewatchers, stakekeepers	No

prioritize individuals and manage them according to the prioritization result. Within this perspective, two of the most adopted model, both at the company and project level (Project Management Institute, 2021; Aaltonen et al., 2008), are the Salience Model (Mitchell et al., 1997) that incorporates the evaluation of three stakeholder attributes - power, urgency, and legitimacy and the Power/Interest matrix which consists of a grid where power and interest are fixed on two axes is (Mendelow, 1981).

2.3. Limitations of current classification models

These traditional classification models are widespread in practice, however for most of them the literature has highlighted some limitations, outlined in some very limited empirical papers that tested the models. The first limitation is that these traditional classification schemes are static, meaning that they do not allow for monitoring the dynamism and changes that can occur in stakeholder attributes during a project's lifecycle (Winch and Bonke, 2002). This aspect has been highlighted by (Aaltonen et al., 2008) who propose a method for incentivizing and tracing the changes in the salience attribute in stakeholders during a project course and by (Beringer et al., 2013) who underlines that stakeholder attributes and attitude toward the project tend to change in different stages of the project lifecycle. The second limitation is that models such as power interest matrix or salience model consider a small number of stakeholder attributes and characteristics, providing high-level classification results in the form of a two-dimensional stakeholder matrix. This approach has been found restrictive in some empirical papers such as (O'Higgins and Morgan, 2006) who point out that their research would have benefited from being able to consider more stakeholder ideology and values; to overcome this issue, both (Winch & Bonke, 2002) and Olander and Landin (2005) suggest using a combination of multiple methods, in order to derive a more granular perspective of stakeholder classification which considers more attributes. Finally, a third limitation relates to the way the management usually identifies and map the intensity of stakeholder attributes, both in theoretical classifications and in prioritized stakeholder matrices. This process is dominantly performed through brainstorming and checklists, relying mainly on individual managers' judgement. This involves a significant degree of subjectivity in the evaluation, which can lead to vagueness and imprecision in the classification process (Pérez Vera, 2018). Although some authors have tried to limit the subjectivity of experts by using specific techniques such as fuzzy logic (Bendjenna et al., 2012), classifications are still largely based on individual judgments that risk to introduce biases in the overall stakeholder evaluation. In addition, the entire mapping process is generally carried out manually, drawing and updating charts and tables, thereby increasing the risk of introducing errors into the classification model.

The aim of this paper is to propose an alternative and data–driven method to classify project stakeholders into different groups. To this purpose we suggest the use of unsupervised machine learning as a possible method to overcome the limitation highlighted in the literature;

2.4. Unsupervised machine learning

Machine learning involves the development of computational approaches to automatically analyse patterns, learn from data, and make decisions with minimal or nonexplicit human assistance (Samuel, 1969). Machine learning algorithms are generally categorized in the two classes of supervised and unsupervised learning. Supervised machine learning algorithms, learn from past labelled data to predict the membership of unseen new data points to certain classification categories. Indeed, starting from the analysis of a known training dataset, supervised machine learning algorithms produce an inferred function to make predictions about the output values. In this way, the system is able to provide a classification for any new input after a sufficient training (Kotsiantis, 2007). The logic and purpose of supervised machine learning algorithms do not match with stakeholder classification because in these analyses the purpose is not to predict the single stakeholder membership in a pre-determined category as much as to figure out and describe what categories or groupings may lie within an unlabelled dataset (Pérez Vera, 2018; Reyad et al., 2021). This aim can be pursued adopting unsupervised machine learning algorithms that are used when the input data is neither classified nor labelled. Unsupervised learning algorithms infer a function to describe a hidden structure from unlabelled data. The system doesn't figure out the right output, but it explores the data and can draw inferences from datasets to describe hidden structures from them (Hahne et al., 2008). Unsupervised machine learning includes several classes of algorithms, which can be employed in an alternative way depending on the purpose of the analysis to be performed. For example, when the aim is to simplify the data structure without losing information, dimensionality reduction algorithms can be used (Huang et al., 2018) or alternatively, associative algorithms can be used when the aim is to associate variables in large databases for understanding what is the impact that associations among some variables can have on others (Cohen et al., 2001). When, like in stakeholder classification, the purpose of the analysis is to uncover similarities among data points grouping similar elements together, clustering techniques can be effectively employed (Reyad et al., 2021). In fact, clustering consists of a set of methods for grouping objects into homogeneous classes. A cluster is a set of objects that have similarities with each other but, conversely, have dissimilarities with objects in other clusters. The input of a clustering algorithm is a sample of unlabelled items, while the output is given by a number of clusters into which the items in the sample are divided according to a measure of similarity (Charu and Chandan, 2013). These algorithms prove to be highly effective in several fields of application, including social sciences, where they can be used to group samples of populations or geographical areas with similar demographic or social characteristics (Marbouti et al., 2021; Pavone et al., 2021) and marketing segmentation where clustering algorithms have been used in literature to group customers or products with similar characteristics or attributes in a variety of field such as customers in the electricity market (Motlagh et al., 2019), food shoppers (Su et al., 2019; McCarthy et al., 2020) cars buyers (Morton et al., 2017) and tourists (Alén et al., 2017; Cavagnaro et al., 2018). The advantage that comes from applying clustering algorithms is that they provide as output a description of the characteristics of each cluster, which enables the decision maker to make strategic and tailored decisions about actions to be taken for different groups (Ray, 2019). Hence, like in market segmentation, also in project stakeholder classification these algorithms can enable both an effective "pattern spotting" that groups stakeholders into clusters and offer a valuable starting point for framing stakeholder management strategies customized for each different group (Veerappa and Leitier, 2011; Pérez Vera, 2018; Reyad et al., 2021).

2.5. The four categories of unsupervised clustering methods

The literature classifies clustering methods into four main categories: probabilistic, density-based, hierarchical and partitional clustering. Probabilistic models have been studied by many scholars (Wolfe, 1963; Cavalli-Sforza and Edwards, 1967; Bock, 1996; Ajiboye et al., 2014) since the beginning of the research interest in cluster analysis. The probabilistic model-based approach assumes that each of the group present in the dataset is generated by an underlying probability distribution. Thus, each component probability distribution corresponds to a cluster. Density-based clustering is a nonparametric approach where the clusters are identified considering the high-density areas observed in the dataset (Charu and Chandan, 2013). In density-based clustering each cluster is a region of high density of objects, and it is separated from other clusters by sparser areas.

Hierarchical clustering algorithms, approach the problem of clustering by developing a binary tree-based data structure called dendrogram. Once the dendrogram is constructed, the analyst can automatically choose the right number of clusters by splitting the tree at different levels to obtain different clustering solutions for the same dataset without rerunning the clustering algorithm again. Among the different categories of clustering methods, partitional clustering algorithms are the most used because of their simplicity, competitive computational capability, scalability, and effectiveness. Partitional clustering methods assign data to k clusters (where k is an input parameter), by optimizing an objective function that captures the local and global structure of grouping. After the initial assignment, an iterative relocation procedure moves data points from one cluster to another so that data objects within the same cluster are similar or close (Celebi, 2015).

K-means clustering is the most widely used partitional clustering algorithm. This method requires as input parameters the number of groups (k) and a distance metric. After choosing k representative points as the initial centroids, the algorithm assigns data points to the closest centroid based on a proximity measure. Although K-means is currently the best-known and most used clustering algorithm, it has the key limitation of working with numerical values only; consequently, when dealing with mixed data sets, a K- medoid clustering algorithm is preferred. K-medoids is similar to K-means, but it is proved to be more robust and less sensitive to outliers (Maione et al., 2019). The k-medoids algorithm is based on finding k representative samples within the data set, called medoid, whose average dissimilarity with respect to all the other objects in the same cluster is minimal.

The k clusters are constructed through the association of each sample to its nearest representative object (medoid): each point is then assigned to the closest cluster based on its proximity to the medoid; The algorithm proceeds iteratively until each representative object is the medoid of the cluster.

The drawback of the partitioning methods is that a predefined K parameter-namely the number of clusters – is required in order to perform the cluster analysis. Thus, a method named Average Silhouette Analysis, which consists in estimating the optimal number of clusters k by optimizing an objective function that is an average of silhouette coefficients, is usually employed in order to define the most proper K parameter for a specific dataset. The silhouette coefficient is an internal measure for cluster validation that considers both the intra-cluster and inter-cluster distances. The optimal number of clusters k is determined by maximizing the average of all the silhouettes in the dataset (Gentleman and Carey, 2008). Table 2 summarizes the different types of clustering with their advantages and disadvantages.

The literature shows only few papers that proposes the use of machine learning for stakeholder classification (Pérez Vera, 2018; Pérez Vera and Bermudez Peña, 2022); however, these works employ supervised machine learning and thus classify stakeholders into predefined classifications through labelled input data. Instead in this paper, the authors propose to adopt an unsupervised clustering approach allowing the automatic recognition of groupings among stakeholders. Specifically, given the simplicity of the application and the ability to effectively handle datasets with mixed variables, the authors suggest adopting the k-medoid clustering algorithm previously described to perform an effective stakeholder classification.

3. Research approach

The aim of this paper is to demonstrate the potential benefits and limitations of employing unsupervised machine learning to support project stakeholder classification. To answer this research question, we employed a single exploratory case study. The adoption of this method finds justification in the fact that case studies, in addition to answering research questions regarding the "how" and "why" of a specific phenomenon, can also be used to evaluate the impacts and outcomes of specific interventions applied in a specific context (Dul and Hak, 2007). In this particular case, the intervention consists of applying an innovative method for classifying stakeholders, and since the literature on the subject is limited and does not allow for building on existing theories and empirical results, the approach adopted is exploratory in nature (Yin, 2012). To frame the case study, the authors purposefully selected a company, collected data regarding the stakeholders involved in one of the company's major projects, and applied an unsupervised clustering procedure, sharing the results of the analysis with the company management. This paper provides an empirical demonstration of the model implementation which aims to show what are the benefits of adopting a data-driven stakeholder classification process; the following paragraphs illustrate the steps followed to involve the company's management and collect the data necessary to perform the analysis; the steps are also reported in Fig. 1; a short overview of the clustering procedure is then provided, before the discussion of the results obtained.

3.1. Data collection

The process followed to collect the data aimed to improve the validity and reliability of the results, employing a combination of primary and secondary sources of information. A medium-large Italian IT company, specialized in the design and delivery of hardware and software solutions, was purposefully selected to demonstrate the application of unsupervised clustering to support project stakeholder classification. The company has different sites in Italy, the US, and New Zealand and it owns a strong network of partnerships with the most important worldwide IT providers, and a consistent number of relationships with different and multicultural stakeholders. In fact, the company project's stakeholders operate in many fields - Finance, Manufacturing, Healthcare, Public Administration, and Retail – allowing the collection of stakeholder data coming from diverse and variegate industries. The authors decided to concentrate on an IT company to show the potential benefit arising from adopting unsupervised machine learning techniques for stakeholder classification in companies where projects are managed in a complex and fast-changing environment (Brynjolfsson and Hitt, 2000). The procedure followed to collect the data consisted of the following steps:

Table 2
The four clustering methods with their advantages and disadvantages.

	Advantages	Disadvantages	Examples	Applicability to Stakeholder Classification
Probabilistic models (Wolfe, 1963; Cavalli-Sforza and Edwards, 1967; Ajiboye et al., 2014)	The estimation of the number of clusters as the choice of clustering method are statistical problems.	Difficulty in applying it on real data due to lack of a distribution.	Gaussian mixture model	Difficult to apply because data concerning different individuals rarely take on a specific statistical distribution. In addition, stakeholder datasets often contain outliers (e.g., individuals who have particularly high/low values of a specific attribute) that these methods cannot handle.
Density based models (Charu and Chandan, 2013; Campello et al., 2020)	No initial assumptions are required. Allows to discover arbitrary shapes.	Decrease of efficiency when dimensionality of data increases	DBSCAN, OPTICS	Difficult to apply because they become highly computationally intensive as the dataset increases. They struggle in handling clusters of varying densities and shapes – while stakeholder groups often have irregular density.
Hierarchical models (Everitt et al., 2011)	Intuitive and easy to explain	Slow, sensitive to outliers, not able to manage mixed data	Agglomerative, Divisive	Difficult to apply because they need large dataset to provide reliable results (while often project stakeholder database counts around 100 records). They do not handle mixed data (continuous and categorical) while stakeholder data are often expressed in both measures.
Partitional Clustering (Maione et al., 2019)	Simplicity, computational capability, scalability, effectiveness, ability to manage mixed data.	Requires computation of number of clusters K	K-means, K-medoid	It is the most suitable class of algorithms for stakeholder classification as it handles mixed data, it is simple to apply and less sensitive to outliers. It has good computational capabilities, also when increasing the database and can handle several stakeholder attributes simultaneously.

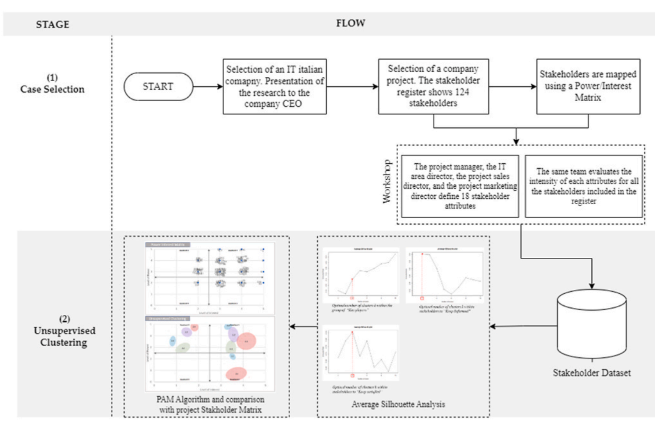

Fig. 1. Research steps followed and clustering procedure.

- The research project was presented to the CEO of the company who expressed his interest in the potentiality of performing a data-driven stakeholder classification process. Therefore, the research team was allowed to collect stakeholder data coming from a past company project. The project, a software delivery for a client about to go public, involved many listed companies and organizations as stakeholders, and it is the largest project that has been managed by the company.
- The first step consisted of involving the project manager to retrieve data from the stakeholder register used to enlist project stakeholders. The register reported a list of 124 stakeholders, operating in different industries (Infrastructure, Health sector, Manufacturing, Finance) and having different roles (Project Manager, Sales Specialist, Delivery Technicians, Purchase Specialists), which were included in the dataset to be used for the analysis.
- The second step consisted of a workshop during which the project manager, the project manager, the IT area director, the project sales director, and the project marketing director, all senior figures in the company with more than fifteen years of experience, were asked to identify, based on their experience, the relevant stakeholder attributes that could be crucial to be evaluated both for the design of the IT service and for its final, successful delivery. The workshop was necessary in the absence of a structured corporate dataset collecting data and information regarding project stakeholders. To ensure the accuracy and reliability of the primary data collected during the workshop, the authors recorded and transcribed the workshop, analyzing, through a coding procedure the information provided by the participants. The final list of attributes, which included individuals' demographic characteristics as well as psychographic attitude toward the project, is reported in Table 3 in the format validated by the company's CEO.
- In the third step, the same group of respondents was asked to evaluate the extant company stakeholder documentation in order to express for each of the 124 stakeholders the grade of intensity of the 13 variables expressing the psychographic attributes, based on a Likert scale (from 1 to 5 defined as "Very Low", "Low". "Medium", "High", "Very High"). To ensure a proper accuracy and reliability of the data collection phase the entire workshop (which lasted 4 h) was recorded and transcribed.

The choice of focusing on respondents with specific roles was driven by the internal organizational guidelines, according to which the project manager, the IT manager, the project sales manager and the project marketing manager are the roles responsible for stakeholder management and engagement. Indeed, each project is followed by a dedicated team that can provide relevant insights about the stakeholders involved in the client's project. The validity and reliability of this methodological approach is aligned with some authors, according to whom professional roles like the Service Managers and the Project Managers are the best informants since they continuously interact with customers and use customer information for making project decisions (Olson et al., 1995; Song and Parry, 1997). The output of the data collection process was thus a dataset of 124 stakeholders described through a set of 18 attributes (of which 5 are demographic and 13 psychographic), with magnitude expressed in terms of linguistic variables ranging from 1 to 5. The data collected were recorded in an Excel worksheet and subsequently imported into the analytical software R Studio to be explored, cleaned, and used to train the proposed model.

3.2. Unsupervised clustering: a proposed approach for stakeholder classification

Unsupervised clustering includes algorithms that can detect similarities within a set of different input data. Thus, in unsupervised

Table 3
The stakeholder attribute list defined together with the project manager, the IT area director, the project sales director, and the project marketing director.

ATTRIBUTE (D= DEMOGRAPHIC; P= PSYCHOGRAPHIC)	ATTRIBUTE DESCRIPTION	JUSTIFICATION
NATIONALITY (D)	Stakeholder's nationality	The company operates in a global context, and most of the stakeholders identified in the register are international. Managers included nationality as cultural differences can impact on stakeholder grouping.
AGE (D)	Stakeholder's age	Stakeholders' age was included as it can influence both on group formation and on subsequent communication strategies.
GENDER (D)	Gender of individual stakeholder (male, female)	Gender was included as a demographic attribute useful in supporting differentiated communication strategies
EDUCATIONAL BACKGROUND (D)	Educational background: Secondary School (SS), High School (HS), bachelor's degree (D), master's degree (MD), Doctor of Philosophy (PhD)	The educational background was included as a socio-demographic attribute to frame clusters and related management strategies.
ROLE (D)	Professional role held by the stakeholder in the organization to which he or she belongs	The stakeholder role was added as an information useful to address targeted managerial actions.
EXPERTISE IN OWN ROLE REFLECTS (P)	Degree of skills and knowledge owned by the stakeholder in the specific job position	Managers included this attribute to be able to evaluate the seniority and the general expertise in different clusters.
EXPERTISE IN IT (P)	Stakeholders' degree of knowledge of IT infrastructures, including knowledge of existing IT solutions, ability to identify its requirements and to assess risks.	Being IT the core business of the project, this attribute was included to understand how to deal with stakeholders having different IT competences and skills.
BUSINESS SKILLS (P)	Refers to the ability to successfully manage business issues, achieve favourable outcomes, design well defined plans that accounts all the strengths and potential issues of the company	Included to assess the how to deal with stakeholders that have different attitude toward achieve outcomes within the expected deadline, following a plan or managing issues and risks.
ATTITUDE TOWARDS THE PROJECT (P)	Stakeholders' attitude toward the project defined in the following possible behaviours: Positive (P), supportive (S), indifferent (I), negative (N), resistant (R) – partially retrieved from (Boschetti et al., 2012)	Included for understanding how to deal with stakeholders who have very diverse attitudes toward the project.
COMMUNICATION SKILLS (P)	Concern the ability to communicate clearly and effectively with the company and with other stakeholders, creating the understanding and trust necessary to work on a project (Barrett, 2006)	Included in order to assess stakeholder communication ability and how to deal – also on an operational level – with diversified communication skills.
NEGOTIATION SKILLS (P)	The ability to reach a common agreement with the company and with other stakeholders, as outlined in (Roloff et al., 2003)	Included in order to be able to differentiate between stakeholders with high negotiation skill that enable smoother and faster processes and stakeholder with low negotiation skill who require a more careful management.
HARD SKILLS (P)	Technical' skills in the exercise of the activity carried out by the company to which the individual belongs, as defined in (Lyu and Liu, 2021)	In addition to soft skills such as communication and negotiation, hard skills were added to differentiate especially between the more technically savvy customers and suppliers and those with less expertise.
RESULT-ORIENTATION (P)	The ability to focus on the output and move efficiently toward the goals, meanwhile measuring the effects of the work and analysing the results.	Included to enable managers to differentiate between stakeholder with high or low result orientation.
FLEXIBILITY (P)	Tolerance for projects' changes and readiness to adapt.	Included in order to differentiate the approach to stakeholder engagement strategy. More "Agile" and informal in case of flexibility to change or very formal in case of traditional stakeholders
INTEREST (P)	Degree of interest of the stakeholders in the project activities and results, as defined in (Mendelow, 1981; Johnson et al., 2008)	This attribute was kept as already part of the project stakeholder classification method.
POWER (P)	Expresses whether a stakeholder is able to impress and actually achieve its expectations on the project, as defined (Mendelow, 1981; Johnson et al., 2008; Mitchell et al., 1997)	This attribute was kept as already part of the project stakeholder classification method.
LEGITIMACY (P)	Perception or assumption that the actions of the stakeholder are desirable, proper, or appropriate as defined in (Mitchell et al., 1997)	Included in order to diversify the management of stakeholder according to their legitimacy.
URGENCY (P)	Degree of time sensitivity and criticality owned by the stakeholders toward the project, as defined in (Mitchell et al., 1997)	Included to diversify the response times toward stakeholders considering their urgency in making requests and claims.

learning, the machine independently recognizes patterns and structures within the input values. As pointed out in the theoretical background this automatic recognition of similarity in the data can be particularly useful for identifying, with a data-driven approach, categories of stakeholders showing similar characteristics. Among the different unsupervised clustering algorithms presented in the literature review, partitional methods represents a preferrable option to perform project stakeholder classification: these algorithms are in fact characterized by greater speed and effectiveness (Ben Salem et al., 2018). Furthermore, stakeholder attributes were expressed through continuous (e.g. "Age"), categorical (e.g. "Gender", "Role") and ordinal (e.g. "Communication skills") variables, therefore the Partitional Medoid Clustering (PAM) was identified by the authors as an appropriate alternative since it is able to handle datasets consisting of mixed categories of data. This algorithm is less sensitive to outliers than alternatives such as k-means, making it possible to correctly classify stakeholders with a high variability in the attribute's intensity. The high interpretability of the clusters favoured by the fact that the medoids are represented by real data points and the possibility of supporting incremental updates, make it particularly suitable for a stakeholder classification process that may undergo modifications during the project lifecycle (Arunachalam and Kumar, 2018). To enable an effective comparison between the method currently employed by the company for classifying stakeholders and the PAM algorithm, the project power interest matrix was employed as an input for the cluster analysis. In fact, we decided to run the algorithm separately for each quadrant of the power interest matrix, so that the different grouping logic could emerge clearly and enable an easier comparison between the two methods. As most clustering techniques, PAM requires the number of clusters as an input parameter (k), thus the first step performed in each quadrant of the power interest matrix was to select the optimum number of clusters through the Silhouette plot: the optimal number of clusters corresponds to the point at which the average silhouette width is maximized. Subsequently the PAM algorithm was launched on the dataset containing the individual stakeholders and the variables defined in Annex 1. Specifically, the algorithm received the data in input as a $n \, x \, m$ matrix where:

- n represents the number of individuals (stakeholders) that the algorithm must categorize.
- m is the number of variables (stakeholder attributes) based on which the stakeholders must be categorized by the algorithm.
- each element a_{ij} of the matrix represents the evaluation expressed by the management of the stakeholder i related to the characteristic j.

$$\begin{pmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nm} \end{pmatrix}$$

The PAM algorithm is intended to find a set of objects, called *medoids,* among the points of the input dataset. The term medoid refers to a data point within a cluster for which the average dissimilarity between it and all the other members of the cluster is minimal; it corresponds to the most centrally located point in the cluster. Once the medoids were identified, the algorithm performed a swap procedure in order to improve the quality of the clustering solution by iteratively swapping data points between clusters. Through this process, the PAM algorithm refines the clustering solution by optimizing the assignment of data points to clusters, with the aim of minimizing the total dissimilarity and creating more compact and well-separated clusters (Botyarov and Miller, 2022). This process is iteratively performed until defining k representative medoids able to minimize the sum of the dissimilarities of the observations to their closest representative object. The results of the procedure applied to the company's stakeholder classification are reported in the following paragraph.

4. Results

4.1. Clusters description

To classify project stakeholders, the company employed a power interest matrix, with each stakeholder's level of power and interest expressed on a Likert scale ranging from 1 to 5. As mentioned in the previous paragraph, the algorithm was run on each quadrant of the matrix instead of considering the entire dataset, enabling to emphasize the enhanced granularity of the clustering in respect to the current method employed by the company. The project power interest matrix included the 124 stakeholders identified in the stakeholder register and reported a significant number of individuals in the quadrants including stakeholders classified as “Key players” (*Quadrant 1- high power and high interest*), and in the other two including stakeholders to “Keep informed” (*Quadrant 2- low power and high interest*) and to “Keep satisfied” (*Quadrant 3- high power and low interest*); on the contrary only 3 stakeholders were included in the “Minimal effort” quadrant, not allowing to perform a clustering analysis in that specific area of the matrix. On the three quadrants under consideration, the optimal number of clusters was first estimated using the Average Silhouette Analysis - whose outputs are graphically represented in the Silhouette plots reported in Fig. 2. The optimal number of clusters that maximize the Silhouette coefficient is four in Quadrant 1, two in Quadrant 2 and four in Quadrant 3.

Subsequently the PAM algorithm was employed to define the optimal number of medoids, enabling to identify, for each quadrant, clusters of stakeholders having different characteristics. Fig. 3 shows a graphical comparison between the power interest matrix currently adopted in the company and the clusters identified running the PAM algorithm. The main difference that emerges is in the level of granularity in the grouping of stakeholders, made possible by considering 18 attributes simultaneously instead of the two-dimensional attributes of the power and interest matrix.

In the first quadrant of stakeholders, the “Key players” (*Quadrant 1*), four subclusters were identified based on the stakeholders' attributes reported in *Annex 1*; stakeholders belonging to subcluster 1.1 showed a high level of expertise in their own field (Accounting, IT support, Purchase Management), a good knowledge of IT infrastructures, requirements and of IT risk management. Besides the technical competences, stakeholders in this group had excellent managerial, negotiation and communication abilities. They worked with a particular focus on the achievement of the established goal showing a high level of commitment toward the project. They were willing to participate in the project's decision-making process and in problem identification and resolutions. Stakeholders belonging to subcluster 1.2 were characterized by lower experience in their own field and weaker expertise in terms of the IT business and communication skills. Furthermore, they showed a neutral attitude and a medium degree of commitment toward the project. Subcluster 1.3 encompasses stakeholders who do not have either good communication or negotiation skills as well experience in the IT sector. Furthermore, they are perceived to be just sufficiently interested and committed to the project, showing, in some cases, a resistant behaviour which implies a high risk of difficulties in structuring an effective relationship with them during the management of the project. Subcluster 1.4 includes stakeholders that are experts in their field and demonstrate to have good knowledge of IT solutions. Similarly, to the previous group, they show a resistant attitude toward the project, however they demonstrate good negotiation, business, and communication skills.

The application of the PAM algorithm on the group of stakeholders to “Keep informed” (*Quadrant 2*), led to the identification of two subclusters; Both clusters represented stakeholders that proved expertise in their own role. However, individuals belonging to subcluster 2.2 showed stronger business knowledge and competences in the IT industry. Nevertheless, the main difference between the two identified subclusters can be traced in their attitude toward the project: individuals belonging to subcluster 2.2 had a neutral attitude and were weakly committed toward the project, while stakeholders in subcluster 2.1 were strongly interested in the project but with a resistant attitude. Finally, the clustering procedure applied to the “Keep satisfied” stakeholders (*Quadrant 3*) led to the identification of four subclusters: the first one (subcluster 3.1) includes individuals that showed the lowest degree of all the skills with respect to the other clusters identified in the same quadrant. Clusters 3.2 and 3.3 both included stakeholders with excellent communication and negotiation skills; The two groups differed only in terms of the attitude, since cluster 3.2 included stakeholders having a positive attitude toward the project, while the distinguishing characteristics of cluster 3.3 was the neutrality towards project activities and decisions. To conclude, subcluster 3.4 grouped individuals with a medium/high level of skills but lacking the ability to be result oriented with

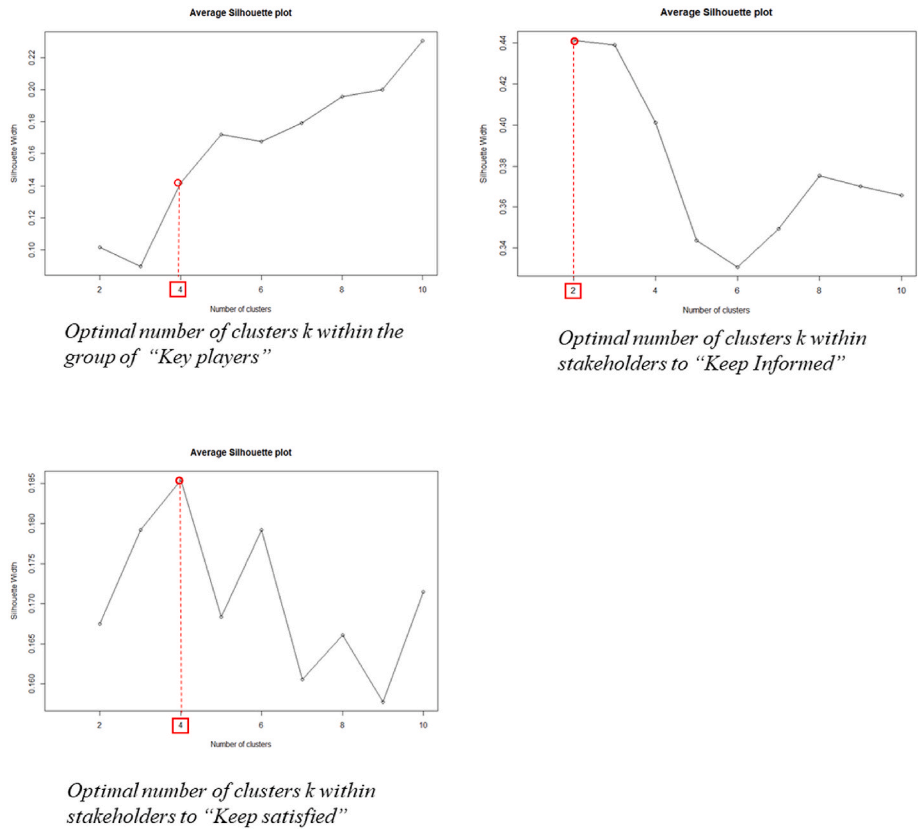

Fig. 2. Average Silhouette Plot: output of the analysis.

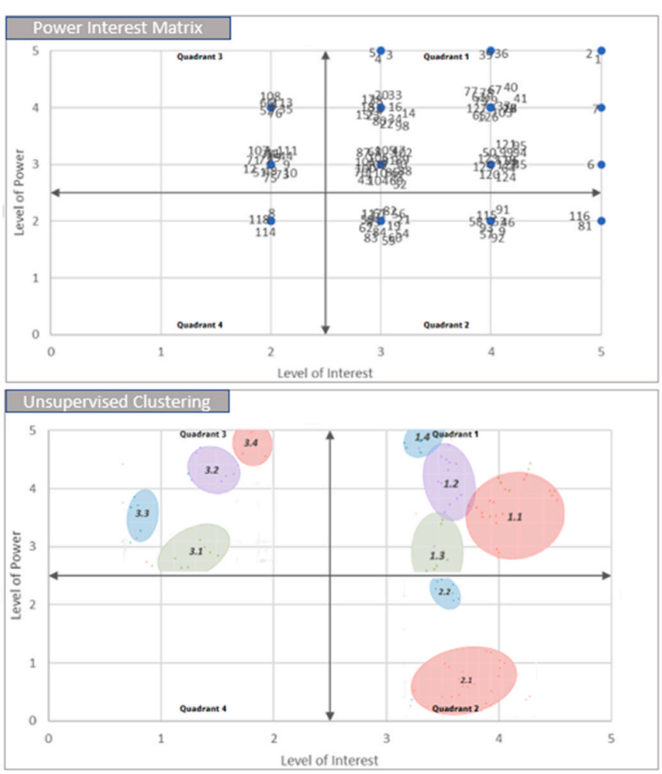

Fig. 3. Comparison between the project power interest matrix which is reported above and the T-SNE plot of the unsupervised clustering, reported below.

respect to project objectives. Aside from enabling a deeper description of the clusters, the advantaged that derives from employing clustering algorithms can be effectively highlighted by reporting a single classification as example. Stakeholders numbered as 40 and 41 in the original power interest matrix are both mapped with a power and interest value of 4. Thus, when adopting the power interest matrix logic, the actions and strategy developed for their engagement to the project should be identical; however, when adding some attributes for consideration and running the cluster analysis, stakeholder 40 is included in cluster 1.1 while stakeholder 41 in cluster 1.2. This does not imply a dramatic shift in perspective, as both stakeholders have high power and interest toward the project. However, the two different clusters group individuals with a notable difference in the level of IT experience. This aspect, which does not emerge in the power interest matrix, is of great relevance to the company, which can differentiate engagement strategies between a stakeholder who is perfectly aligned with the company's technical know-how and another who is less experienced, for whom a higher-level type of communication should be defined.

4.2. Accuracy of clustering procedure results

The quality of the clustering obtained through the PAM algorithm was validated using the Silhouette index, that assess the compactness and separation of clusters based on the distances between data points within and between clusters by calculating for each data points (a) the average distance between the data point and all other points in the same cluster and (b) the average distance between the data point and all points in the nearest neighboring cluster (i.e., the cluster with the minimum average distance). The Silhouette coefficient for a data point is given by (b - a)/max(a, b) and it ranges from -1 to 1, where a value close to 1 indicates that the data point is well-clustered, a value close to 0 indicates that the data point is on or near the decision boundary between two clusters, and a value close to -1 indicates that the data point may be assigned to the wrong cluster (Hahne et al., 2008). In our study, the classification inaccuracy (negative index results) was found in the 15.1% of the records in Q2, in the 13.3% in Q1 and in the 7.8% in Q3; however, as can be seen from the table reported in Annex 1, the negative values are never too close to -1, meaning that the classification shows an overall satisfactory accuracy.

5. Discussion: the potential contributions of unsupervised clustering to project stakeholder management

5.1. Contribution to project stakeholder management research

The main contribution of this work to stakeholder management research is to propose an alternative method to perform a reliable stakeholder classification. Over the past 20 years of research, alongside traditional classification models such as those proposed by (Mitchell et al., 1997; Johnson et al., 2008), few authors have proposed alternative methods for classifying stakeholders, although empirical studies have reported difficulties for project managers in properly managing this process during the life cycle of a project (Jepsen and Eskerod, 2009). This paper contributes to stakeholder theory by proposing an alternative and innovative stakeholder classification method that overcomes some of the limitations that the current methods reported in literature hold. The salience method by (Mitchell et al., 1997) consider only three attributes (Power, Legitimacy and Urgency) and treat them as “present or absent,” when it is clear that each operates on a continuum or series of continua. The method results in a classification into three main groups (Dormant, Discretionary and Demanding stakeholders) and four secondary groups (Dominant, Dangerous, Dependent and Definitive stakeholders) that is based on the preliminary assumption that only those three attributes are relevant to classify all the stakeholders. In comparison to the salience model, the adoption of clustering algorithms leads to a different and more customized perspective of analysis: while traditional classifications consider few attributes as the absolute fundamentals for evaluating stakeholders, cluster analysis seeks to understand among many possible attributes which ones best characterize and define them. This implies that while for some stakeholders the degree of power, legitimacy, and urgency is decisive, for others, different attributes may better characterize them. This logic, which is in line with the goal of unsupervised clustering to perform a “pattern spotting” among data, enables to overcome also the limitations of the two-dimensionality of power interest matrices (Mendelow, 1981; Johnson et al., 2008). Indeed, although they consider the degree of ownership of the two attributes, allowing for a prioritized stakeholder management, the two-dimensionality results in high level classifications, that are often too rough, especially when the stakeholders identified are many in number (Mainardes et al., 2012). In contrast, unsupervised clustering works effectively on large datasets by more granularly identifying groups of stakeholders with similar characteristics. For example, in our case study which considered the power interest matrix as an input, the application of the PAM algorithm allowed the identification of groups of stakeholders showing very different attitudes, ranging from neutrality to a strong resistance toward the project. While highlighting the limitations of the current classification methods, this shade the lights on the benefit of adopting clustering algorithms, as a more detailed and tailored grouping allows for more customized engagement strategies.

5.2. Contribution to stakeholder management in practice

The main purpose of project stakeholder management is to enable the project manager to take adequate action in relation to the stakeholder characteristics and their interest toward the project (Jepsen and Eskerod, 2009). Both the process of identifying and classifying stakeholders are part of the stakeholder analysis and aims to identify stakeholders, as well as to characterize them on several dimensions or attributes. Despite the richness of the literature about stakeholder classification, some authors have highlighted relevant limitations; according to (Jepsen and Eskerod, 2009) the guidelines provided in current literature are far too general to be useful for project managers who may find them too unspecific for the application in a project context;

Considering these limitations, our study contributes to practice demonstrating that supervised machine learning can be employed as an adaptive and data driven classification method, that identify, and group stakeholder based on their similarities and relationships without relying on predefined categories. Second, as some empirical papers have demonstrated, the intensity of the attributes as well as the attitude towards the project can change over the project’s duration (Jepsen and Eskerod, 2009) and across project types (Eskerod and Huemann, 2013). The presented classification method is flexible and thus adaptable to the initial planning phase, during which project stakeholders are mapped for the first time, and to any classification adjustments that may occur during the project lifecycle. In fact, when changing the intensity of an attribute owned by a stakeholder in the dataset the algorithms is fast and precise in updating the classification result. Third, unsupervised clustering allows project managers to detect outliners - that is, stakeholders who do not belong to any cluster identified by the algorithm. These stakeholders may represent individuals with unique perspectives toward the project and their prior identification can support the project manager in developing ad hoc strategies for their engagement.

While in the case analyzed stakeholders were individuals, unsupervised clustering can also be employed to group larger entities and organizations; this type of analysis can be carried out before setting up an analysis for individuals, to understand at a higher level how organizations are grouped according to certain characteristics. This process can be employed in the early stages of the project to get an overall overview of the stakeholder network represented by other organizations involved in or impacted by the project. Although the use of unsupervised clustering for stakeholder prioritization offers considerable benefits, some limitations must be pointed out. In the case we presented, data were collected in the form of management judgements of stakeholder attributes expressed on a scale ranging from 1 to 5, normalizing in this way the values assumed by some of the variables under consideration. However, in many corporate settings, project stakeholder datasets are built from information taken from a variety of sources which often do not allow to obtain standardized and consistent values. This means that the adoption of unsupervised clustering can be facilitated to the extent that the company already has a structured data management system (for example a CRM) for storing information about its stakeholders. In addition, the analyses were performed using the R programming software; although more user-friendly tools also exist, the development and interpretation of the stakeholder clustering results require data science skills, which are not always widespread in companies (Davenport and Patil, 2012). Thus, although the use of clustering algorithms enables the development of more detailed stakeholder management strategies, their ultimate application in business settings may be limited by still partially scarce data science skills.

6. Limitation and conclusion

The application of a PAM algorithm on real data coming from a case study company, proved the effectiveness of the method, highlighting the existence of subclusters of stakeholders having different characteristics within the categories identified adopting as a method a power interest matrix. We will now conclude outlining the limitations of this work, which may also represent venues for future research. The first limitation concerns the dataset used for the analysis: the number of stakeholders considered, 124 in total, is rather limited and distributed in only three quadrants of the power interest matrix. The availability of a larger database, extracted from a structured CRM, could lead to more accurate and detailed clusters identification. The lack of a structured enterprise database led also to the need to collect some data - the additional attributes and the level of their intensity for each stakeholder - in a workshop, relying on the experience of five company's managers. This introduces subjectivity in the analysis that in future research can be mitigated through techniques for reducing the uncertainty associated with individual judgments, such as Fuzzy Logic. In addition, the fact that certain attributes were selected by the company rather than others and that data were collected in a single company, limit the generalizability of the results obtained. Finally, in order to better highlight the comparison between the two methods, it was decided to use the project power interest matrix as input of the analysis, so that the degree of granularity and the emergence of subclusters could be examined quadrant by quadrant and compared with the current classification and management approach. Future research could explore the outcome of the application of the clustering algorithm on the entire dataset, testing its practicality of use by managers and the degree of efficiency in ensuring a prioritized stakeholder management.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Annexes

Annex 1
Stakeholders showing negative values of the Silhouette index ("Silhouette width")

	Individual's ID	Cluster	Neighbor	Silhouette width
KEY PLAYERS	15	1	4	-0.03368859
	11	1	3	-0.12676933
	13	3	2	-0.02092149
	20	4	1	-0.05587866
KEEP INFORMED	2	1	2	-0.03259065
	5	1	2	-0.08019896
	3	1	2	-0.08727233
	28	1	2	-0.11177614
	31	1	2	-0.14645273
KEEP SATISFIED	40	1	2	-0.02762439
	35	1	2	-0.09966809
	54	2	3	-0.01463907
	3	2	3	-0.03266248
	46	4	3	-0.14293994

References

Aaltonen, K., Kujala, J., 2010. A project lifecycle perspective on stakeholder influence strategies in global projects. Scand. J. Manag. 26 (4), 381–397. https://doi.org/10.1016/j.scaman.2010.09.001.

Aaltonen, K., Kujala, J., 2016. Towards an improved understanding of project stakeholder landscapes. Int. J. Proj. Manag. 34 (8), 1537–1552. https://doi.org/10.1016/j.ijproman.2016.08.009.

Aaltonen, K., Kujala, J., Oijala, T., 2008. Stakeholder salience in global projects. Int. J. Proj. Manag. 26 (5), 509–516. https://doi.org/10.1016/j.ijproman.2008.05.004.

Aaltonen, K., et al., 2015. Stakeholder dynamics during the project front-end: the case of nuclear waste repository projects. Proj. Manag. J. 46 (6), 15–41. https://doi.org/10.1002/pmj.

Abyad, A., 2017. Globalization challenges in project management. Middle east Journal of Business 12 (4), 10–15.

Afzal, F., et al., 2019. A review of artificial intelligence based risk assessment methods for capturing complexity-risk interdependencies. Int. J. Manag. Proj. Bus. 14 (2), 300–328. https://doi.org/10.1108/IJMPB-02-2019-0047.

Ajiboye, A., Isah-Kebbe, H., Oladele, T.O., 2014. Cluster analysis of data points using partitioning and probabilistic model-based algorithms. International Journal of Applied Information Systems (IJAIS) 7 (7), 21–26.

Alén, E., Losada, N., de Carlos, P., 2017. Profiling the segments of senior tourists throughout motivation and travel characteristics. Curr. Issues Tourism 20 (14), 1454–1469. https://doi.org/10.1080/13683500.2015.1007927.

Andersen, E.S., 2008. Rethinking Project Management : an Organisational Perspective. FT Prentice Hall.

Arunachalam, D., Kumar, N., 2018. Benefit-based consumer segmentation and performance evaluation of clustering approaches : an evidence of data-driven decision-making. Expert Syst. Appl. 111, 11–34. https://doi.org/10.1016/j.eswa.2018.03.007.

Asad, M., et al., 2020. Transforming business decision making with internet of things (IoT) and machine learning (ML). In: 2020 International Conference on Decision Aid Sciences and Application (DASA), pp. 674–679 (Ml).

Barrett, D.J., 2006. Strong communication skills a must for today's leaders. Handb. Bus. Strat. 7 (1), 385–390. https://doi.org/10.1108/10775730610619124.

Ben Salem, S., Naouali, S., Chtourou, Z., 2018. A fast and effective partitional clustering algorithm for large categorical datasets using a k-means based approach. Comput. Electr. Eng. 68, 463–483. https://doi.org/10.1016/J.COMPELECENG.2018.04.023.

Bendjenna, H., Charre, P., Nacer, E., 2012. Using multi-criteria analysis to prioritize stakeholders. J. Syst. Inf. Technol. 14 (3), 264–280.

Bento, S., et al., 2022. Artificial intelligence in project management: systematic literature review. Int. J. Technol. Intell. Plann. 13 (2), 143. https://doi.org/10.1504/IJTIP.2022.126841.

Beringer, C., Jonas, D., Kock, A., 2013. Behavior of internal stakeholders in project portfolio management and its impact on success. Int. J. Proj. Manag. 31 (6), 830–846. https://doi.org/10.1016/J.IJPROMAN.2012.11.006.

Bisi, M., Goyal, N.K., 2016. Software development efforts prediction using artificial neural network. IET Software 10 (3), 63–71. https://doi.org/10.1049/iet-sen.2015.0061.

Bock, H.H., 1996. Probabilistic models in cluster analysis 23, 5–28.

Boschetti, F., et al., 2012. Assessing attitudes and cognitive styles of stakeholders in environmental projects involving computer modelling. Ecol. Model. 247, 98–111. https://doi.org/10.1016/j.ecolmodel.2012.07.027.

Botyarov, M., Miller, E.E., 2022. Partitioning around medoids as a systematic approach to generative design solution space reduction. Results in Engineering 15 (July), 100544. https://doi.org/10.1016/j.rineng.2022.100544.

Brynjolfsson, E., Hitt, L.M., 2000. Beyond computation: information technology, organizational transformation and business performance. J. Econ. Perspect. 14 (4), 23–48. https://doi.org/10.1257/JEP.14.4.23.

Campello, R.J.G.B., et al., 2020. Density-based clustering. Wiley Interdisciplinary Reviews: Data Min. Knowl. Discov. 10 (2) https://doi.org/10.1002/widm.1343.

Cavagnaro, E., Staffieri, S., Postma, A., 2018. Understanding millennials' tourism experience: values and meaning to travel as a key for identifying target clusters for youth (sustainable) tourism. J. Tourism Futur. 4 (1), 31–42. https://doi.org/10.1108/JTF-12-2017-0058.

Cavalli-Sforza, L.L., Edwards, A.W., 1967. Phylogenetic analysis. Models and estimation procedures. Am. J. Hum. Genet. 19 (3) https://doi.org/10.2307/2406616.

Celebi, M.E., 2015. Partitional Clustering Algorithms.

Charu, A., Chandan, k R., 2013. Data clustering algorithms and applications. An Introduction to Toxicogenomics. https://doi.org/10.1201/9781315373515-1.

Cheng, M.-Y., Hoang, N.-D., Wu, Y.-W., 2015. Cash flow prediction for construction project using a novel adaptive time- dependent least squares support vector machine inference model. J. Civ. Eng. Manag. 21 (6), 679–688. https://doi.org/10.3846/13923730.2014.893906.

Clarkson, M., 1995. A stakeholder framework for analyzing and evaluating corporate social performance. Acad. Manag. Rev. 20 (1), 92–117.

Cleland, D.I., 1988. Projecr stakeholder management. In: Project Management Handbook, second ed.
Cohen, E., et al., 2001. Finding interesting associations without support pruning. IEEE Trans. Knowl. Data Eng. 13 (1), 64–78. https://doi.org/10.1109/69.908981.
Davenport, T.H., Patil, D.J., 2012. Data scientist: the sexiest job of the 21st century. Harv. Bus. Rev. 90 (10), 70–77.
Derakhshan, R., Turner, R., Mancini, M., 2019. Project governance and stakeholders : a literature review. Int. J. Proj. Manag. 37 (1), 98–116. https://doi.org/10.1016/j.ijproman.2018.10.007.
Di Maddaloni, F., Davis, K., 2018. Project manager's perception of the local communities' stakeholder in megaprojects. An empirical investigation in the UK. Int. J. Proj. Manag. 36 (3), 542–565. https://doi.org/10.1016/J.IJPROMAN.2017.11.003.
Di Maddaloni, F., Sabini, L., 2022. Very important , yet very neglected : where do local communities stand when examining social sustainability in major construction projects. Int. J. Proj. Manag. 40 (7), 778–797. https://doi.org/10.1016/j.ijproman.2022.08.007.
Dul, J., Hak, T., 2007. Case Study methodology in business research. Case Study Methodology in Business Research 1–302. https://doi.org/10.4324/9780080552194.
Elias, A.A., Cavana, R.Y., Jackson, L.S., 2002. Stakeholder analysis for R & D project management. R D Manag. 32 (4), 301–310. https://doi.org/10.1111/1467-9310.00262.
Elish, M.O., Helmy, T., Hussain, M.I., 2013. Empirical study of homogeneous and heterogeneous ensemble models for software development effort estimation. Math. Probl. Eng. 1–21.
Eskerod, P., Huemann, M., 2013. Sustainable development and project stakeholder management: what standards say. Int. J. Manag. Proj. Bus. 6 (1), 36–50. https://doi.org/10.1108/17538371311291017.
Eskerod, P., Huemann, M., Savage, G., 2015. Project stakeholder management-past and present. Proj. Manag. J. 46 (6), 6–14. https://doi.org/10.1002/pmj.21555.
Eskerod, P., Huemann, M., Ringhofer, C., 2016. Stakeholder inclusiveness: enriching project management with general stakeholder theory. Proj. Manag. J. 46 (6), 42–53. https://doi.org/10.1002/pmj.21546.
Everitt, B., et al., 2011. Cluster analysis. Wiley series in probability and statistics 330. XII. https://www.wiley.com/en-us/Cluster+Analysis%2C+5th+Edition-p-9780470749913. (Accessed 4 January 2023).
Faghihi, V., et al., 2015. Automation in construction scheduling : a review of the literature. Int. J. Adv. Manuf. Technol. 81 (9–12), 1845–1856. https://doi.org/10.1007/s00170-015-7339-0.
Fassin, Y., 2009. The stakeholder model refined. J. Bus. Ethics 84 (1), 113–135. https://doi.org/10.1007/S10551-008-9677-4/METRICS.
Flyvbjerg, Bent, Budzier, Alexander, Chun-kit, Ricky Lau, Agard, Karlene, Agard, Karlene, Leed, A., 2022. AI in Action: How the hong kong Development Bureau Built the Pss — an Early-Warning-Sign System for Public Works Projects, pp. 1–28. SSRN.
Freeman, E., 1984. Strategic management: a stakeholder approach. Strategic Management: A Stakeholder Approach 1–276. https://doi.org/10.1017/CBO9781139192675.
Fridgeirsson, T.V., et al., 2021. An authoritative study on the near future effect of artificial intelligence on project management knowledge areas. Sustainability 13 (4), 1–20. https://doi.org/10.3390/su13042345.
Gentleman, R., Carey, V.J., 2008. Unsupervised Machine Learning. Bioconductor Case Studies, pp. 137–157. https://doi.org/10.1007/978-0-387-77240-0_10.
Goodpaster, K.E., 1991. Business ethics and stakeholder analysis. Bus. Ethics Q. 1 (1), 53–73. https://doi.org/10.2307/3857592.
Hahne, F., et al., 2008. Unsupervised machine learning. Bioconductor Case Studies. https://doi.org/10.1007/978-0-387-77240-0.
Holzmann, V., Zitter, D., Peshkess, S., 2022. The expectations of project managers from arti fi cial intelligence. A Delphi Study' 53 (5), 438–455. https://doi.org/10.1177/87569728211061779.
Huang, X., Wu, L., Ye, Y., 2018. A review on dimesionality reduction techniques. Int. J. Pattern Recogn. Artif. Intell. 33 (10), 2–23. https://doi.org/10.1142/S0218001419500174.
Huemann, M., Eskerod, P., Ringhofer, C., 2016. Rethink!: Project Stakeholder Management. Project Management Institute. Available at: https://www.projectmanagement.com/pdf/Rethink_Project_Stakeholder_Management.pdf. (Accessed 7 January 2023).
Jepsen, A.L., Eskerod, P., 2009. Stakeholder analysis in projects: challenges in using current guidelines in the real world. Int. J. Proj. Manag. 27 (4), 335–343. https://doi.org/10.1016/j.ijproman.2008.04.002.
Johnson, G., Scholes, K., Whittington, R., 2008. Stakeholder mapping: the power/interest matrix. Exploring Corporate Strategy 156.
Jones, T.M., Wicks, A.C., 1999. Convergent stakeholder theory. Acad. Manag. Rev. 24 (2), 206–221.
Kamoona, K., Budayan, C., 2019. Implementation of genetic algorithm integrated with the deep neural network for estimating at completion simulation. Adv. Civil Eng. 2019, 1–15, 7081073.
Kotsiantis, S.B., 2007. Supervised machine learning: a review of classification techniques. Informatica 31 (3), 249–268.
Lehtinen, J., Aaltonen, K., 2020. Organizing external stakeholder engagement in inter-organizational projects : opening the black box. Int. J. Proj. Manag. 38 (2), 85–98. https://doi.org/10.1016/j.ijproman.2019.12.001.
Lundin, R.A., Söderholm, A., 1995. A theory of the temporary organization. Scand. J. Manag. 11 (4), 437–455. https://doi.org/10.1016/0956-5221(95)00036-U.
Luoma, P., Goodstein, J., 1999. Research Notes. Stakeholders and Corporate Boards: Institutional Influences on Board Composition and Structure 42 (5), 553–563. https://doi.org/10.5465/256976.
Lyu, W., Liu, J., 2021. Soft skills, hard skills: what matters most? Evidence from job postings. Appl. Energy 300 (March), 117307. https://doi.org/10.1016/j.apenergy.2021.117307.
Machiels, T., Compernolle, T., Coppens, T., 2023. Stakeholder perceptions of uncertainty matter in megaprojects : the Flemish A102 infrastructure project. Int. J. Proj. Manag. 41 (1), 102437 https://doi.org/10.1016/j.ijproman.2023.102437.
Magaña Martínez, D., Fernandez-Rodriguez, J.C., 2015. Artificial intelligence applied to project success: a literature review. International Journal of Interactive Multimedia and Artificial Intelligence 3 (5), 77. https://doi.org/10.9781/ijimai.2015.3510.
Mainardes, E.W., Alves, H., Raposo, M., 2012. A model for stakeholder classification and stakeholder relationships. Manag. Decis. 50 (10), 1861–1879. https://doi.org/10.1108/00251741211279648.
Maione, C., Nelson, D.R., Barbosa, R.M., 2019. Research on social data by means of cluster analysis. Appl. Comput. Inform. 15 (2), 153–162. https://doi.org/10.1016/j.aci.2018.02.003.
Marbouti, F., Ulas, J., Wang, C.H., 2021. Academic and demographic cluster analysis of engineering student success. IEEE Trans. Educ. 64 (3), 261–266. https://doi.org/10.1109/TE.2020.3036824.
McCarthy, B., Kapetanaki, A.B., Wang, P., 2020. Completing the food waste management loop: is there market potential for value-added surplus products (VASP)? J. Clean. Prod. 256, 120435 https://doi.org/10.1016/j.jclepro.2020.120435.
Mendelow, A.L., 1981. Environmental scanning–the impact of the stakeholder concept. In: International Conference on Information Systems, ICIS 1981. Available at: http://aisel.aisnet.org/icis1981/20%0AThis.
Mitchell, R.K., Agle, B.R., Wood, D.J., 1997. Toward a theory of stakeholder identification and salience: defining the principle of who and what really counts. Acad. Manag. Rev. 22 (4), 853–886. https://doi.org/10.5465/AMR.1997.9711022105.
Mariani, C., Mancini, M., 2023. Selection of projects ' primary and secondary mitigation actions through optimization methods in nuclear decommissioning projects. Nucl. Eng. Des. 407, 112284. https://doi.org/10.1016/j.nucengdes.2023.112284 (March).
Moradbeiky, A., 2023. FEEM : a flexible model based on artificial intelligence for software effort. Estimation 11 (1), 39–51. https://doi.org/10.22044/jadm.2022.11161.2265.
Morton, C., Anable, J., Nelson, J.D., 2017. Consumer structure in the emerging market for electric vehicles: identifying market segments using cluster analysis. International Journal of Sustainable Transportation 11 (6), 443–459.
Motlagh, O., Berry, A., O'Neil, L., 2019. Clustering of residential electricity customers using load time series. Appl. Energy 237 (December 2018), 11–24. https://doi.org/10.1016/j.apenergy.2018.12.063.
OECD, 2022. Harnessing the power of AI and emerging technologies. Available at: https://www.oecd-ilibrary.org/docserver/f94df8ec-en.pdf?expires=1669115959&id=id&accname=guest&checksum=C91AD0E611EB3604096C525A0860726D.
Olander, S., Landin, A., 2005. Evaluation of stakeholder influence in the implementation of construction projects. Int. J. Proj. Manag. 23, 321–328. https://doi.org/10.1016/j.ijproman.2005.02.002.
Olson, E.M., Walker, O.C., Ruekert, R.W., 1995. Organizing for effective new product development: the moderating role of product innovativeness. J. Market. 59 (1), 48. https://doi.org/10.2307/1252014.
O'Higgins, E.R.E., Morgan, J.W., 2006. Stakeholder salience and engagement in political organisations: who and what really counts? Soc. Bus. Rev. 1 (1), 62–76. https://doi.org/10.1108/17465680610643355.
Pavone, P., et al., 2021. Multidimensional clustering of EU regions: a contribution to orient public policies in reducing regional disparities. Soc. Indicat. Res. 156 (2–3), 739–759. https://doi.org/10.1007/s11205-020-02324-9.
Pérez Vera, Y., 2018. Stakeholders classi fi cation system based on clustering techniques. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics 11238, 241–252. https://doi.org/10.1007/978-3-030-03928-8.
Pérez Vera, Y., Bermudez Peña, A., 2022. A neuro-fuzzy inference system for stakeholder classification Sistema de inferencia neuro-difuso para la clasificación de las partes interesadas. Ingeniare Rev. Chil. Ing. 30, 378–387.
Phillips, R.A., et al., 2010. Strategy, stakeholders and managerial discretion. Strat. Organ. 8 (2), 176–183. https://doi.org/10.1177/1476127010365721.
Post, J.E., Preston, L.E., Sachs, S., 2002. Managing the extended enterprise: the new stakeholder view. Calif. Manag. Rev. 45 (1), 6–28. https://doi.org/10.2307/41166151.
Project Management Institute, 2021. A Guide to the Project Management Body of Knowledge (PMBOK® Guide), seventh ed. (Seventh).
Ray, S., 2019. A quick review of machine learning algorithms. In: Proceedings of the International Conference on Machine Learning, Big Data, Cloud and Parallel Computing: Trends, Prespectives and Prospects, COMITCon 2019, pp. 35–39. https://doi.org/10.1109/COMITCon.2019.8862451.
Reyad, O., et al., 2021. Genetic K-means adaption algorithm for clustering stakeholders in system requirements. In: Advanced Machine Learning Technologies and Applications, pp. 195–204.
Roloff, M., Putnam, L., Anastasiou, L., 2003. Negotiation skills. In: Handbook of Communication and Social Interaction Skills, pp. 801–814.
Rowley, T.J., 1997. Moving beyond dyadic ties: a network theory of stakeholder influences. Acad. Manag. Rev. 22 (4), 887. https://doi.org/10.2307/259248.
Samuel, A.L., 1969. Some studies in machine learning using the game of checkers. II—recent progress. Annu. Rev. Automat. Program. 6 (PART 1), 1–36. https://doi.org/10.1016/0066-4138(69)90004-4.

Savage, G.T., Nix, T.W., Blair, D., 1991. Strategies for assessing and managing organizational stakeholders. Acad. Manag. Exec. 5 (2).

Shrestha, Y.R., Krishna, V., Krogh, G. Von, 2021. Augmenting organizational decision-making with deep learning algorithms : principles , promises , and challenges. J. Bus. Res. 123 (January 2020), 588–603. https://doi.org/10.1016/j.jbusres.2020.09.068.

Sirgy, M.J., 2002. Measuring corporate performance by building on the stakeholders model of business ethics. J. Bus. Ethics 35 (3), 143–162. https://doi.org/10.1023/A:1013856421897/METRICS.

Song, X.M., Parry, M.E., 1997. A cross-national comparative study of new product development processes: Japan and the United States. J. Market. 1–18. https://doi.org/10.1177/002224299706100201.

Su, C.H., et al., 2019. U.S. sustainable food market generation Z consumer segments. Sustainability 11 (13), 1–14. https://doi.org/10.3390/su11133607.

Taroun, A., Yang, J., 2013. A DST-based approach for construction project risk analysis A DST-based approach for construction project risk analysis. J. Oper. Res. Soc. 64 (3), 1221–1230. https://doi.org/10.1057/jors.2013.38.

Turner, R., Zolin, R., 2012. Forecasting success on large projects: developing reliable scales to predict multiple perspectives by multiple stakeholders over multiple time frames. Proj. Manag. J. 43 (5), 87–99. https://doi.org/10.1002/pmj.21289.

Van Der Laan, G., Van Ees, H., Van Witteloostuijn, A., 2008. Corporate social and financial performance: an extended stakeholder theory, and empirical test with accounting measures. J. Bus. Ethics 79 (3), 299–310. https://doi.org/10.1007/s10551-007-9398-0.

Veerappa, V., Leitier, E., 2011. Clustering stakeholders for requirements decision making. In: Requirements Engineering: Foundation for Software Quality, pp. 202–208.

Vuorinen, L., Martinsuo, M., 2018. Value-oriented stakeholder in fl uence on infrastructure projects. Int. J. Proj. Manag. https://doi.org/10.1016/j.ijproman.2018.10.003.

Wang, Y., Jin, X., 2019. Structural risk of diversified project financing of city investment company in China based on the best worst method. https://doi.org/10.1108/ECAM-05-2019-0249 (51608363).

Whyte, J., 2019. How digital information transforms project delivery models. Proj. Manag. J. 50 (2), 177–194. https://doi.org/10.1177/8756972818823304.

Winch, G., Bonke, S., 2002. Project stakeholder mapping: analyzing the interests of project stakeholders. In: The Frontiers of Project Management Research. Project Management Institute, pp. 385–403.

Wolfe, J.H., 1963. Object Cluster Analysis of Social Areas. University of California, Berkeley. Master's thesis.

Yin, R.K., 2012. Case study methods. In: APA Handbook of Research Methods in Psychology, 2. Research designs: Quantitative, qualitative, neuropsychological, and biological, pp. 141–155.